Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-wide association study of four ultrasound-measured carcass quality traits in Suffolk sheep based on three models

Yunxia Hu
Inner Mongolia Agricultural University
Jing Li
Inner Mongolia Agricultural University
Xinle Wang
Inner Mongolia Agricultural University
Chunyan Zhang
Inner Mongolia Agricultural University
Jia Pan
Inner Mongolia Agricultural University
Dongyue Wang
Inner Mongolia Agricultural University
Heng Liu
Inner Mongolia Agricultural University
Lifei Zhang
Inner Mongolia Agricultural University
Youjun Rong
Inner Mongolia Agricultural University
Shaopeng Zhang
Inner Mongolia Agricultural University
Yujie Zhao
Inner Mongolia Agricultural University
Haozhe Zhang
Inner Mongolia Agricultural University
Yanjun Zhang
Inner Mongolia Agricultural University
Yongbin Liu
Inner Mongolia Agricultural University
Ruijun Wang
nmgwrj@126.com

Inner Mongolia Agricultural University

Research Article

Keywords: Suffolk sheep, genome-wide association study, ultrasound-measured carcass quality traits, haplotype block

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11431425/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Genome-wide association study of four ultrasound-measured carcass quality traits in Suffolk sheep based on three models

Yunxia Hu[1†], Jing Li[1†], Xinle Wang[1], Chunyan Zhang[1], Jia Pan[1], Dongyue Wang[1], Heng Liu[1], Lifei Zhang[1], Youjun Rong[1], Shaopeng Zhang[1], Yujie Zhao[1], Haozhe Zhang[1], Yanjun Zhang[1.2.3] ,Yongbin Liu[1.2.3*], and Ruijun Wang[1.2.3*]

[1]College of Animal Science, Inner Mongolia Agricultural University, Hohhot 010018, China

[2]Key Laboratory of Mutton Sheep Genetics and Breeding, Ministry of Agriculture, Hohhot 010018, China

[3]Key Laboratory of Goat and Sheep Genetics, Breeding and Reproduction, Inner Mongolia Autonomous Region, Hohhot 010018, China

*Corresponding authors: Ruijun Wang (email: nmgwrj@126.com).

*Corresponding authors: Yongbin Liu(email:Liuyongbin@imau.edu.cn).

ABSTRACT

Background: Carcass quality traits are economically important in sheep production; however, the genetic variants underlying these traits in Suffolk sheep raised under localized feeding systems in China remain poorly understood. In this study, genome-wide association study (GWAS) was performed to screen Single Nucleotide Polymorphism (SNP) loci affecting four ultrasound-measured carcass quality traits of Suffolk sheep, construct haplotype blocks, and identify candidate genes associated with these traits, so as to provide theoretical support for marker-assisted selection of Suffolk sheep under localized feeding systems in China.

Results: In total, 43 SNPs associated with ultrasound-measured carcass quality traits were detected in Suffolk sheep, including markers reaching either the genome-wide significance threshold or the suggestive significance threshold, consisting of 19 SNPs for backfat thickness (BFT), one SNP for longissimus dorsi muscle thickness (LMT), 12 SNPs for longissimus dorsi muscle area (LMA), and 14 SNPs for backfat to muscle thickness ratio (BMR). Among them, three shared SNP loci were detected for BFT and BMR. Functional annotation identified six key candidate genes, including BTC and DIRAS2 for BFT, RAB30 for LMA, pleiotropic PCDH18 affecting both LMA and LMT, and BRINP3 and KIF23 for BMR. Five haplotype blocks associated with carcass quality traits were constructed using these significant SNPs.

Conclusions: GWAS for four ultrasound-measured carcass quality traits were implemented with three statistical models on 300 Suffolk sheep, yielding 43 trait-associated SNPs. Functional annotation identified six key candidate genes, where PCDH18 exhibited pleiotropic effects on LMA and LMT. Five haplotype blocks were further established using these candidate SNPs. The present findings supply theoretical evidence and promising molecular markers for marker-assisted selection

of ultrasound-measured carcass quality in Suffolk sheep.

Key words: Suffolk sheep; genome-wide association study; ultrasound-measured carcass quality traits; haplotype block

Background

China is one of the world's major mutton producing and consuming countries. Mutton output and quality directly determine the income of sheep farmers[1]. Carcass quality traits represent critical indicators to evaluate production performance and economic value of mutton sheep. These traits directly affect carcass weight and carcass grade, and in turn determine the economic benefits of mutton sheep production[2]. As an excellent mutton sheep breed, Suffolk sheep's carcass quality traits have attracted considerable attention in mutton sheep breeding and production [3]. Unlike traditional carcass quality assessment methods, in-vivo carcass quality evaluation uses ultrasound or X-ray technology to measure specific anatomical sites on live animals[4]. This approach not only enables effective assessment of carcass quality in live animals but also offers potential economic benefits and is more environmentally friendly.

Genome-wide association studies (GWAS) are widely employed to elucidate the genetic basis of complex quantitative traits and identify trait-associated genetic variants[5]. Complex quantitative traits are influenced by multiple genetic loci and environmental factors, while population stratification and relatedness among individuals in the study population can bias association test statistics, thereby reducing the reliability of association signals[6]. The mixed linear model (MLM) effectively controls confounding effects arising from population structure and genetic relatedness by incorporating a genetic relationship matrix to model random genetic effects[7]. MLM-based methods vary in their variance component estimation strategies, computational algorithms, and statistical testing procedures, thereby affecting the computational efficiency of GWAS and the accuracy of associated locus detection[8, 9]. Therefore, applying multiple association analysis methods with different computational and statistical inference procedures under consistent data quality control criteria allows the robustness of association findings to be assessed across different statistical frameworks.

Based on this foundation, this study employed three methods—GCTA-fastGWA-MLM, GEMMA-MLM, and BLINK—to conduct genome-wide association analyses. GCTA-fastGWA-MLM is based on a linear mixed model framework that uses a sparse genetic relationship matrix (GRM) and

an efficient variance component estimation strategy to account for genetic relatedness among individuals while reducing the computational burden of large-scale association analyses[10], thereby facilitating the efficient identification of trait-associated loci[11]. GEMMA-MLM employs a standard linear mixed model based on a GRM and uses efficient numerical algorithms for parameter estimation and association testing, thereby providing an alternative approach to evaluating associations between genetic variants and traits of interest within a mixed model framework[12]. BLINK employs an iterative multi-locus association strategy that uses the Bayesian information criterion (BIC) for model selection and incorporates linkage disequilibrium (LD) information to select candidate markers for inclusion in subsequent association analyses, thereby helping to reduce confounding effects arising from other associated loci[13]. By comprehensively comparing the associated genomic regions identified by the three methods and the consistency of their association signals, this study provides a statistical basis for subsequent candidate gene identification and investigation of the underlying genetic mechanisms.

Carcass quality traits are typically complex quantitative traits regulated by multiple genes, and their genetic basis is relatively complex[14]. Haplotype analysis, by integrating linkage information of adjacent genetic variants across chromosomes, can elucidate the associations between genetic variants and complex traits at the level of localized genomic regions, thereby providing critical evidence for identifying candidate genomic regions and genes that influence carcass quality traits in Suffolk sheep[15, 16].

Based on whole-genome resequencing data from Suffolk sheep, this study employed three GWAS methods—GCTA-fastGWA-MLM, GEMMA-MLM, and BLINK—to investigate BFT, LMT, LMA, and BMR. The objectives were to identify genetic variants significantly associated with these traits and to screen for potential candidate genes. Furthermore, haplotype analysis was performed to assess associations between different haplotype combinations and the corresponding phenotypes, thereby providing insights into the genetic basis of meat production traits in Suffolk sheep and a basis for future marker-assisted selection.

Results

Descriptive Statistics of Phenotypic Data

In the Suffolk sheep population, the mean values of all measured traits were within the expected ranges, and all traits exhibited sufficient phenotypic variation to support subsequent analyses (Table 1). The phenotypic distributions of all traits were approximately normal, supporting their use in subsequent association analyses (Figure 1). Pearson correlation analysis revealed a significant positive correlation between BFT and BMR ($r = 0.795$, $P < 0.001$), a significant negative correlation between BMR and LMT ($r = -0.532$, $P < 0.001$), and significant positive correlations between LMA and LMT ($r = 0.401$, $P < 0.001$) and between BFT and LMA ($r = 0.253$, $P < 0.001$). In contrast, no significant correlations were observed between BFT and LMT ($r = 0.046$, $P > 0.05$) or between BMR and LMA ($r = -0.041$, $P > 0.05$).

Table 1. Descriptive Statistics of Phenotypic Values for Various Traits in Suffolk Sheep

Traits	Sample Size	Maximum	Minimum	Mean	Standard Deviation	Coefficient of Variation
BFT	300	2.12	0.24	1.14	0.31	27.09%
LMT	300	5.31	1.76	3.58	0.63	17.48%
LMA	300	33.77	11.49	23.09	3.82	16.56%
BMR	300	0.71	0.08	0.33	0.11	32.56%

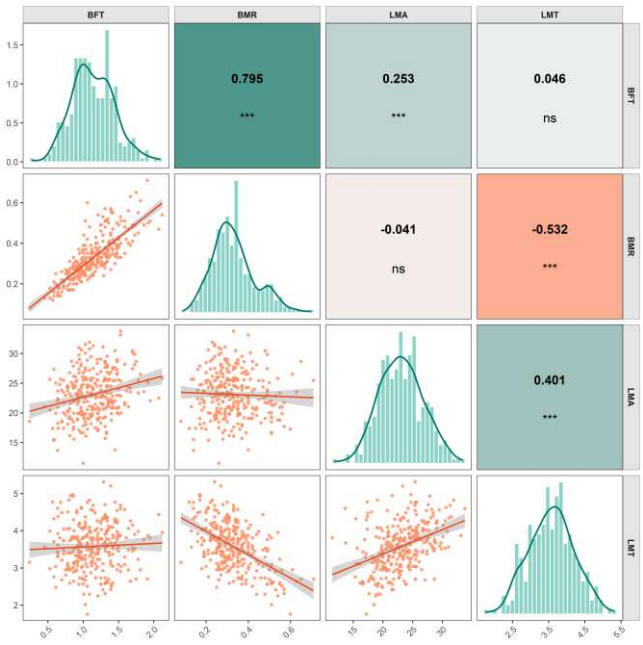

Fig. 1 Pearson correlation analysis among the measured traits. The diagonal panels show histograms of the phenotypic distributions of BFT, BMR, LMA, and LMT, together with their fitted density curves. The lower-left panels display pairwise scatterplots and fitted linear regression lines for the traits, whereas the upper-right panels present Pearson correlation coefficients and their corresponding significance levels. *** indicates a highly significant correlation ($P < 0.001$), whereas ns indicates a non-significant correlation ($P \geq 0.05$).

Genotype Data Quality Control and Population Structure Analysis

Whole-genome resequencing was performed on 300 Suffolk sheep, generating 17,243.32 Gb of raw sequencing data. Quality control was applied to 47,506,993 SNPs, resulting in the retention of 20,180,639 high-quality SNPs. By employing a 1 Mb sliding window, the distribution density of SNP markers across the entire genome was systematically evaluated; these SNPs were distributed across all chromosomal regions (Figure. S1). Based on the quality-controlled SNP dataset, the first three principal component scores for the Suffolk sheep were calculated (Figure. 2). Principal Component Analysis (PCA) revealed significant population stratification among the samples in this study.

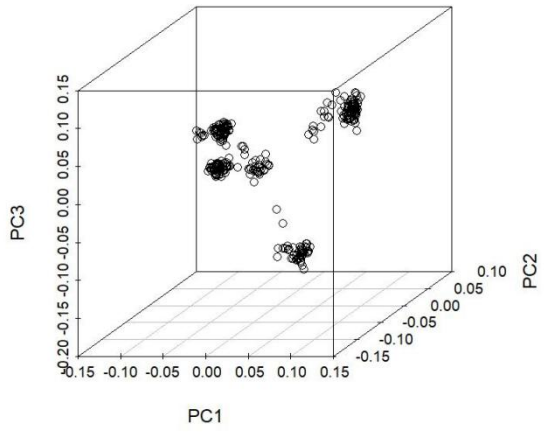

Fig. 2 Principal Component Analysis Plot. The PCA plot displays the first three principal components and the proportion of variance they explain. The x-axis represents the proportion of variance explained by PC1, the y-axis represents the proportion of variance explained by PC2, and the z-axis represents the proportion of variance explained by PC3.

To examine kinship among individuals in the Suffolk sheep population, PLINK software was used to construct both an Identity by State (IBS) matrix and a Genomic Relationship Matrix (GRM) to

assess pairwise genetic relationships (Figure. S2, Figure. S3). The results showed that IBS values within the population ranged from 0.130 to 0.297, indicating that most individuals were distantly related. To further verify individual kinship within the population, the population genomic kinship matrix was visualised. The results were consistent with the results from the IBS matrix: most individuals were distantly related, whereas some were closely related. This phenomenon was associated with the co-occurrence of parent-offspring pairs or sibling pairs within the population.

Genome-Wide Association Analysis

To balance false-positive control with the discovery of potential association signals, this study adopted a genome-wide significance threshold (0.05/1,091,290) and a suggestive significance threshold (1/1,091,290). To balance the reliability of GWAS results with the comprehensiveness of candidate locus selection, this study included SNP loci satisfying either the genome-wide significance threshold or the suggestive significance threshold. This approach reduced the risk of overlooking potential association loci due to overly stringent significance thresholds, thereby enabling a more comprehensive identification of genetic variants associated with the target traits. These loci could then be utilized for further screening of candidate genomic regions and genes.

For BFT, the three methods collectively identified 6 significant SNPs and 71 suggestive SNPs (Fig. 3); for LMT, the three methods collectively identified 24 suggestive SNPs (Fig. 4); for LMA, the three methods collectively identified 15 significant SNPs and 236 suggestive SNPs (Fig. 5); and for BMR, the three methods collectively identified 12 significant SNPs and 122 suggestive SNPs (Fig. 6).

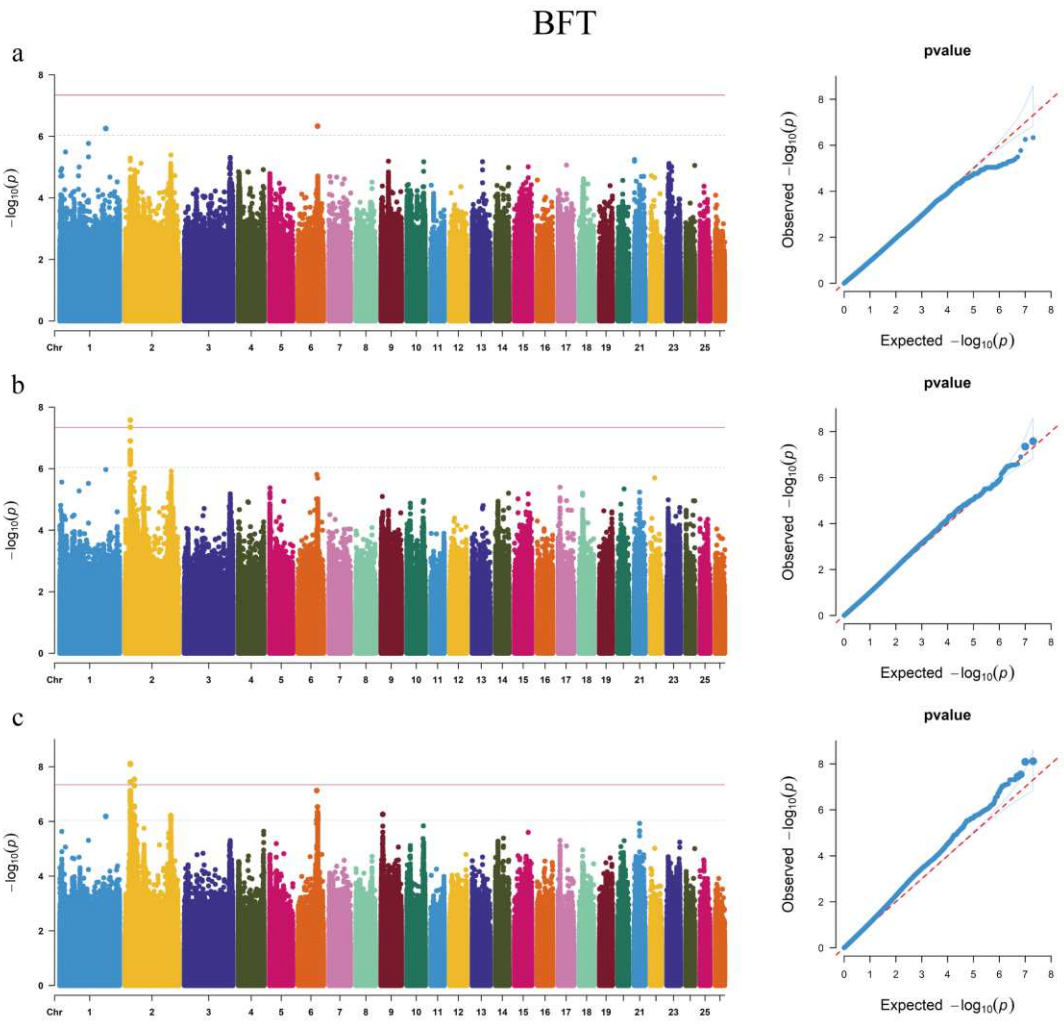

Fig. 3 Manhattan plots and QQ plots of the three GWAS methods in BFT. (a) Analysis results using the GCTA-fastGWA-MLM method; (b) Analysis results using the GEMMA-MLM method; (c) Analysis results using the BLINK method. On the left side of each subfigure is a Manhattan plot, where the horizontal axis represents the genomic positions of chromosomes and SNP loci, and the vertical axis represents the $-\log_{10}$(P) values from association tests; different colors denote different chromosomes. The solid red line indicates the genome-wide significance threshold (0.05/1,091,290), while the gray dashed line indicates the suggestive significance threshold (1/1,091,290). On the right side is a quantile–quantile (QQ) plot, where the horizontal and vertical axes represent the expected and observed $-\log_{10}$(P-value), respectively; the red dashed line depicts the theoretical distribution under the null hypothesis.

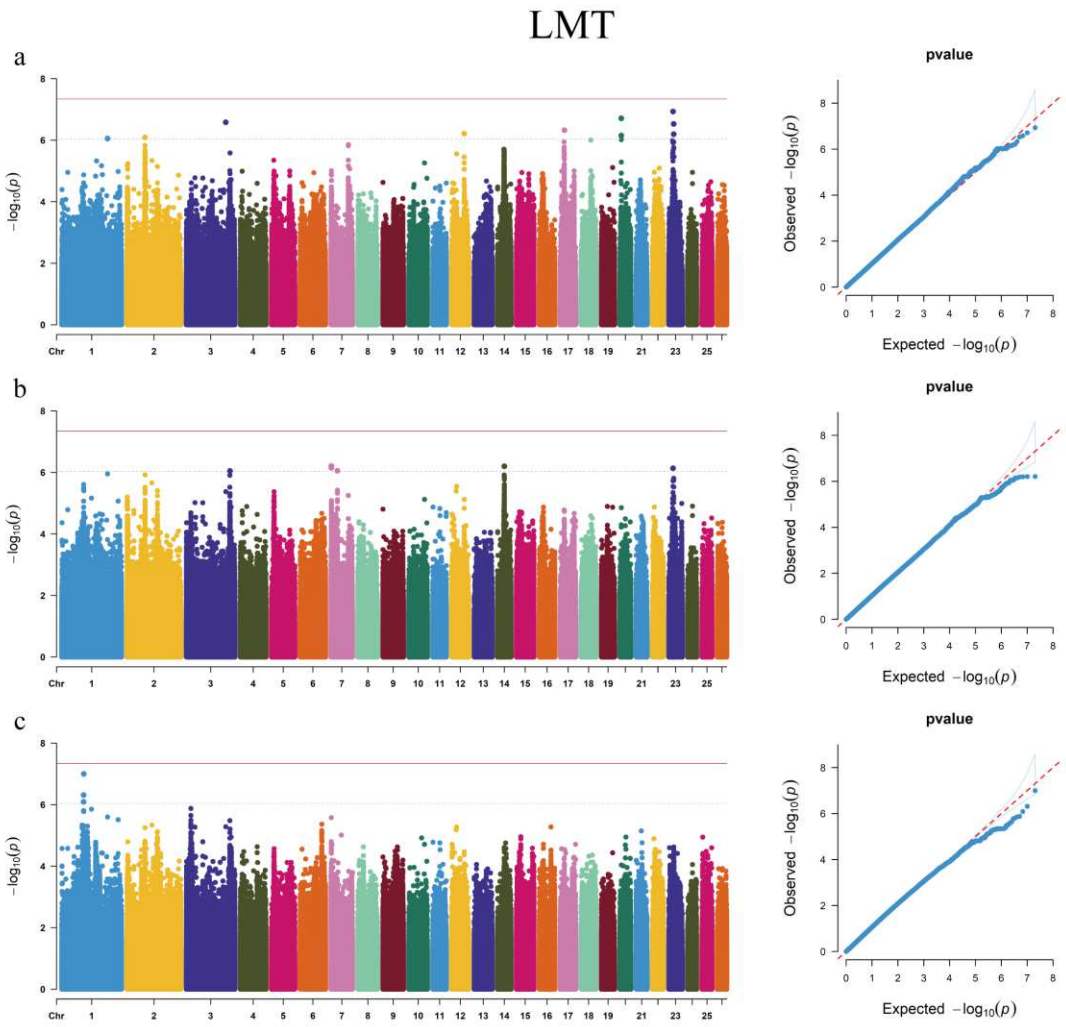

Fig. 4 Manhattan plots and QQ plots of the three GWAS methods in LMT. (a) Analysis results using the GCTA-fastGWA-MLM method; (b) Analysis results using the GEMMA-MLM method; (c) Analysis results using the BLINK method. On the left side of each subfigure is a Manhattan plot, where the horizontal axis represents the genomic positions of chromosomes and SNP loci, and the vertical axis represents the $-\log_{10}$(P) values from association tests; different colors denote different chromosomes. The solid red line indicates the genome-wide significance threshold (0.05/1,091,290), while the gray dashed line indicates the suggestive significance threshold (1/1,091,290). On the right side is a quantile–quantile (QQ) plot, where the horizontal and vertical axes represent the expected and observed $-\log_{10}$(P-value), respectively; the red dashed line depicts the theoretical distribution under the null hypothesis.

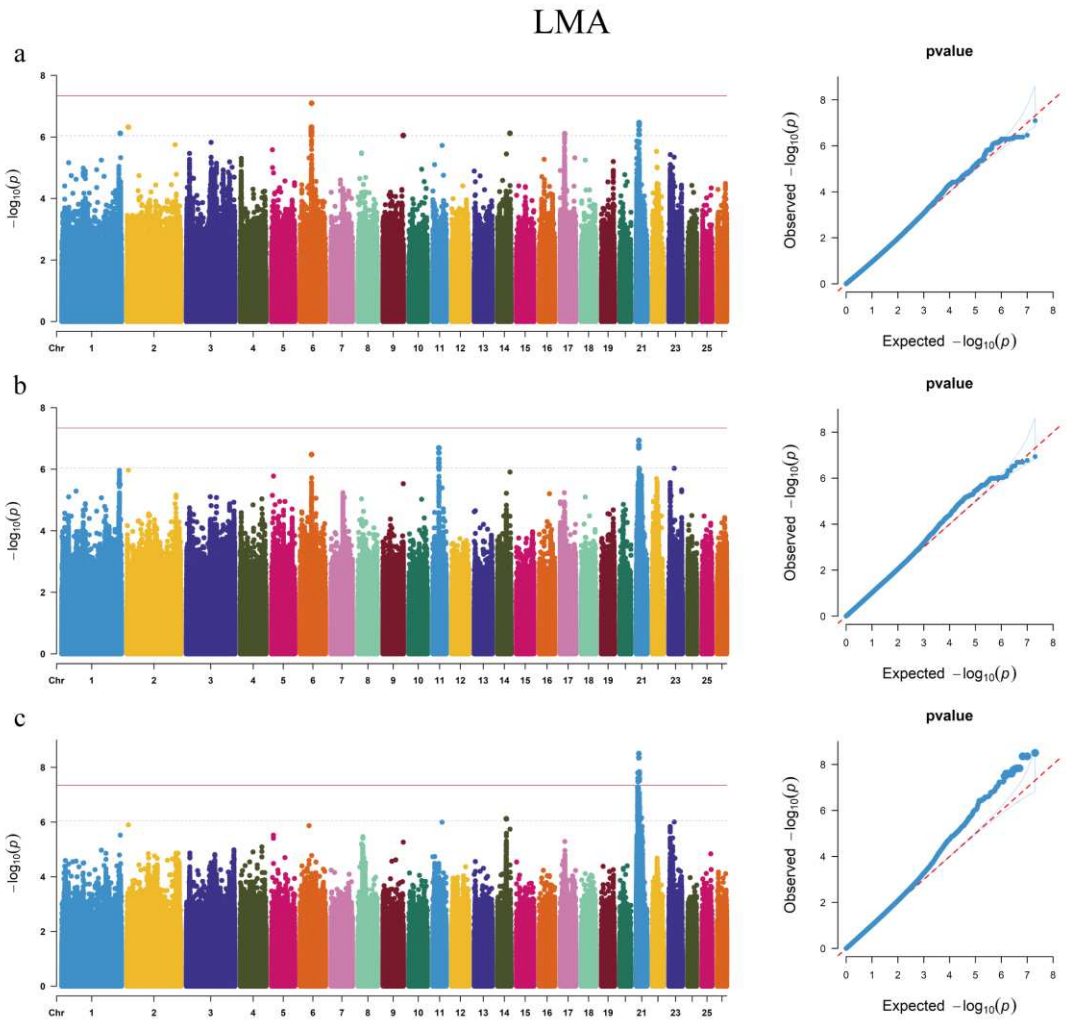

Fig .5 Manhattan plots and QQ plots of the three GWAS methods in LMA. (a) Analysis results using the GCTA-fastGWA-MLM method; (b) Analysis results using the GEMMA-MLM method; (c) Analysis results using the BLINK method. On the left side of each subfigure is a Manhattan plot, where the horizontal axis represents the genomic positions of chromosomes and SNP loci, and the vertical axis represents the $-\log_{10}$(P) values from association tests; different colors denote different chromosomes. The solid red line indicates the genome-wide significance threshold (0.05/1,091,290), while the gray dashed line indicates the suggestive significance threshold (1/1,091,290). On the right side is a QQ plot, where the horizontal and vertical axes represent the expected and observed $-\log_{10}$(P-value), respectively; the red dashed line depicts the theoretical distribution under the null hypothesis.

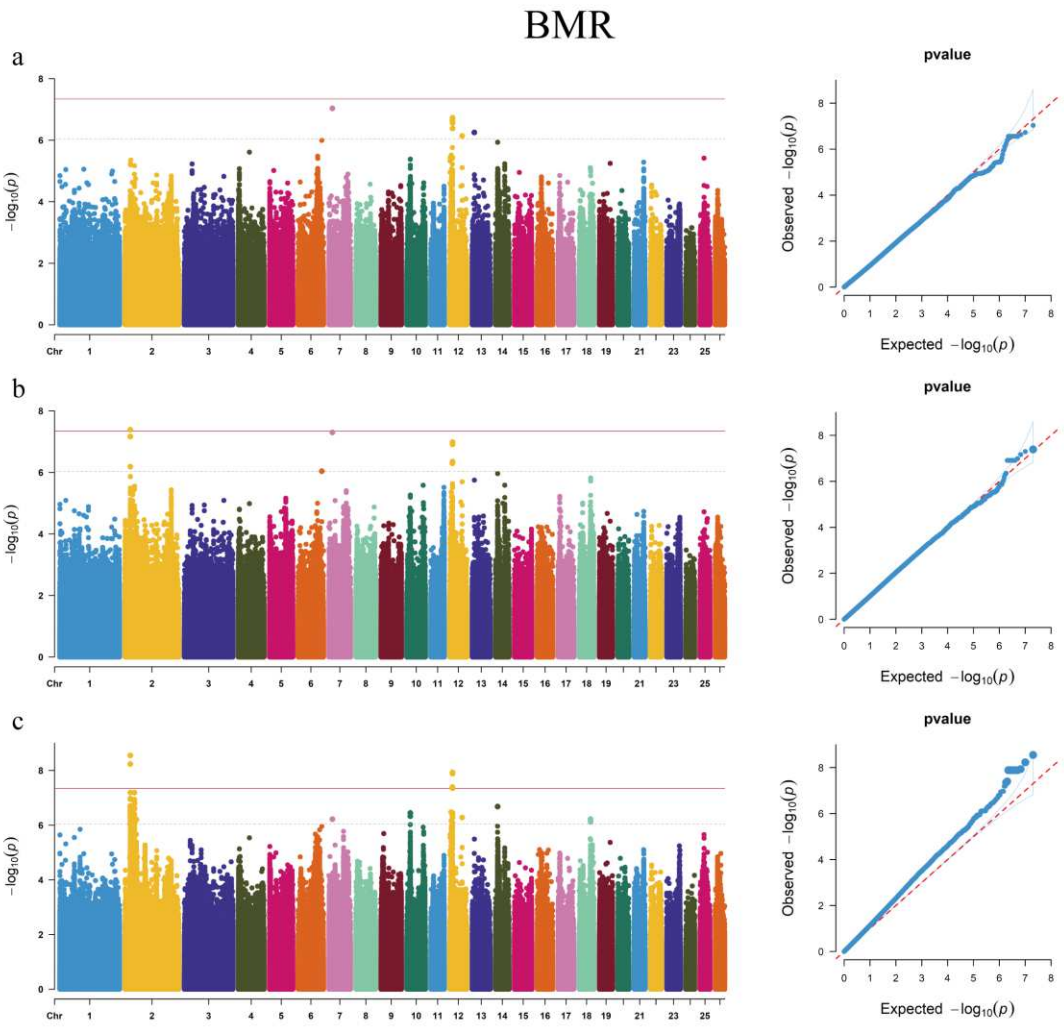

Fig 6. Manhattan plots and QQ plots of the three GWAS methods in BMR. (a) Analysis results using the GCTA-fastGWA-MLM method; (b) Analysis results using the GEMMA-MLM method; (c) Analysis results using the BLINK method. On the left side of each subfigure is a Manhattan plot, where the horizontal axis represents the genomic positions of chromosomes and SNP loci, and the vertical axis represents the $-\log_{10}(P)$ values from association tests; different colors denote different chromosomes. The solid red line indicates the genome-wide significance threshold (0.05/1,091,290), while the gray dashed line indicates the suggestive significance threshold (1/1,091,290). On the right side is a QQ plot, where the horizontal and vertical axes represent the expected and observed $-\log_{10}$(P-value), respectively; the red dashed line depicts the theoretical distribution under the null hypothesis.

Genetic annotation was performed on the previously selected SNP loci, yielding a total of 103 candidate genes. Among these, SNPs associated with BFT were annotated to 23 candidate genes, including *BTC* (β-cellulin) and *DIRAS2* (*DIRAS* family GTP-binding protein 2); SNPs associated with LMT were annotated to 21 candidate genes; SNPs associated with LMA were annotated to 39 candidate genes, including Rab family small GTPase *RAB30*; and SNPs associated with the BMR were annotated to 29 candidate genes, including *BRINP3* (BMP/retinoic acid-induced protein) and kinesin *KIF23*. Notably, SNPs linked to both BFT and BMR shared eight common candidate genes, including *DIRAS2*, while SNPs associated with LMT and LMA traits were jointly annotated to the *PCDH18* gene.

The detection results from three GWAS methods were integrated, and SNPs jointly detected by at least two analytical methods were identified as key candidate loci. The results showed that for the BFT, LMT, LMA, and BMR traits, 19,1, 12, and 14 loci, respectively, were jointly detected by at least two methods; among these, 10 loci for the BMR trait were simultaneously detected by all three methods (Figure 7). Further comparison of key candidate loci across different traits revealed that BFT and BMR shared 3 SNP loci. Full information on the shared SNPs identified by the three GWAS methods is shown in Supplementary Table 1(Table. S1).

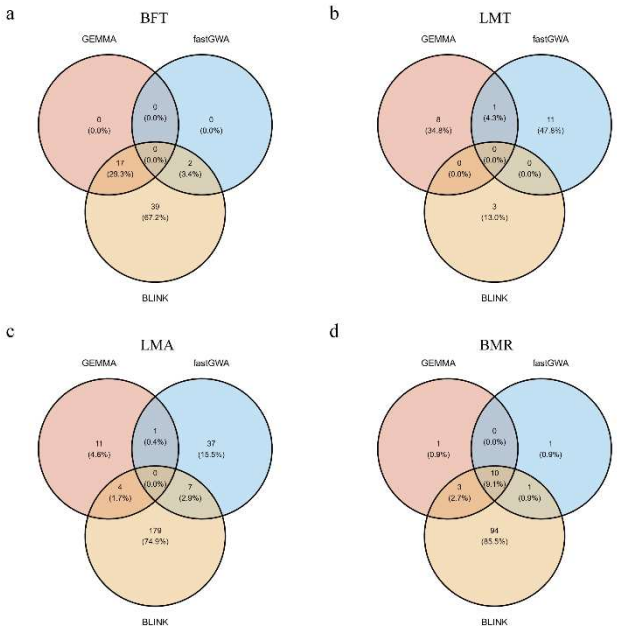

Fig 7. Venn diagrams of overlapping associated SNP loci for each trait detected using different GWAS methods.

Gene Ontology Functional Annotation and Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis of Candidate Genes

Further GO and KEGG functional enrichment analyses were performed on the aforementioned 103 candidate genes. GO enrichment analysis revealed that the candidate genes belonged to three categories of terms: cellular components, molecular functions, and biological processes. Among cellular components, enriched terms included plasma membrane, clathrin-coated vesicles, phosphatidylinositol 3-kinase complexes, and transcription factor TFIID complex. In terms of molecular functions, the primary enriched terms were RNA polymerase II general transcription initiation factor activity and 1-phosphatidylinositol-3-kinase regulatory activity. Regarding biological processes, enriched terms included RNA polymerase II transcription initiation complex assembly, receptor aggregation, establishment or maintenance of epithelial cell apical-basal polarity, and positive regulation of MAP kinase activity. KEGG pathway enrichment analysis indicated that the candidate genes were predominantly enriched in the phospholipase D signaling pathway, platelet activation, and basal transcription factors. Specifically, genes such as *PIK3R5*, *PIK3R6*, and *SYK*

were involved in the phospholipase D signaling pathway and platelet activation, while *TAF2* was associated with basal transcription factors (Figure 8).

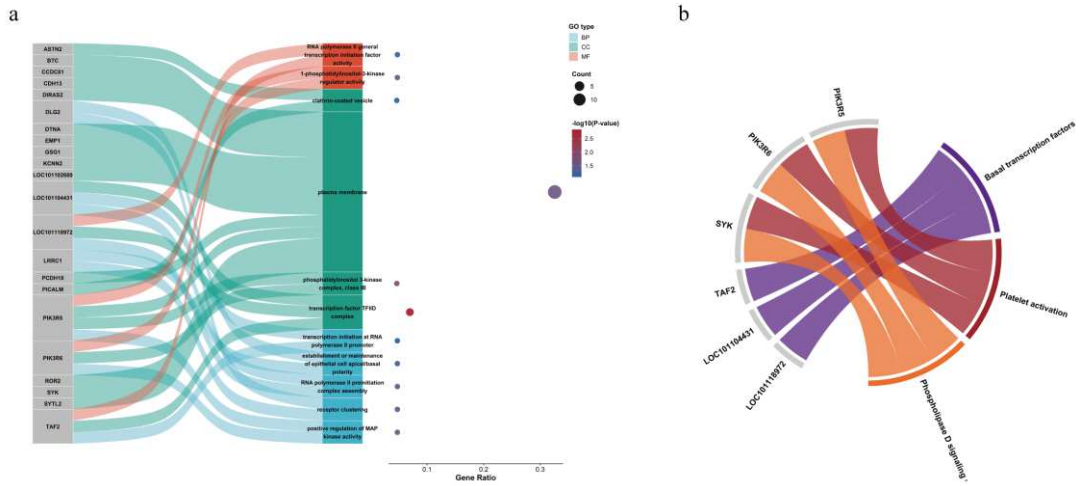

Fig. 8 Results of GO functional enrichment analysis and KEGG pathway enrichment analysis for candidate genes. (a) GO functional enrichment heatmap. Different colors represent the biological processes (BP), cellular components (CC), and molecular functions (MF) categories to which the GO entries belong. (b) KEGG pathway enrichment network diagram. different colors are used to distinguish between distinct KEGG pathways.

Analysis of significant locus linkage disequilibrium and identification of candidate genes

The significantly associated loci identified in this study exhibited a regional-clustering pattern across the genome, and multiple loci demonstrated high levels of linkage disequilibrium along with consistent patterns of genetic effects. Relying solely on individual SNPs is insufficient to comprehensively elucidate the complex genetic architecture underlying in-vivo carcass quality traits. Accordingly, we performed linkage disequilibrium analyses for closely positioned SNPs of shared candidate loci on the same chromosome and constructed haplotype blocks based on these results.

The 17 SNP loci associated with BFT formed block A (Figure 9), yielding 18 haplotypes. Among these, two haplotypes—A1 (0.6183) and A2 (0.3217)—exhibited higher frequencies. Haplotype-phenotype association analysis revealed no significant differences in BFT among the haplotype combinations; however, the A2/A2 haplotype combination exhibited lower BFT compared with other haplotype combinations, making it the superior haplotype combination for the BFT trait. Three genes were annotated within block A: *DIRAS2*, *LOC105606997*, and *LOC105611657*. The majority of this haplotype block was located within the ncRNA-intronic region of *LOC105606997*, with *LOC105611657* situated upstream and *DIRAS2* downstream.

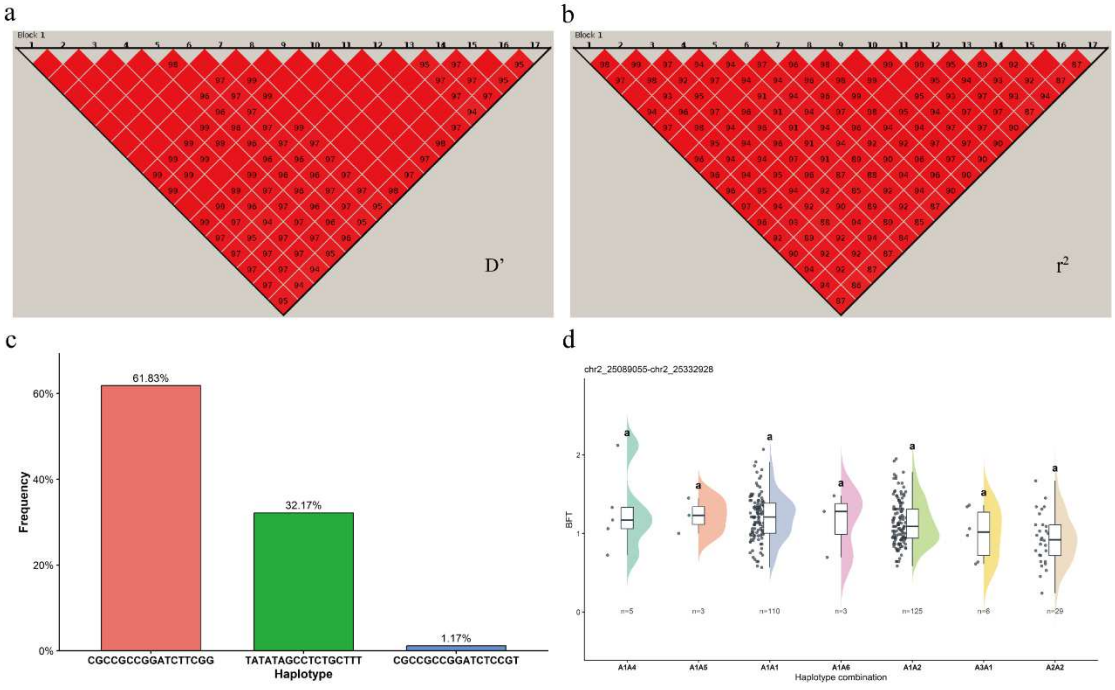

Fig. 9 Results of linkage disequilibrium analysis in regions associated with BFT, as well as haplotype-phenotype association analyses. (a) Heatmap of LD analysis based on D′ values; (b) Heatmap of LD analysis based on r^2 values; (c) Frequency distribution of major haplotypes, with the horizontal axis representing haplotype sequences and the vertical axis representing the frequency of each haplotype in the population; (d) Phenotypic distribution of BFT across different haplotype combinations. Different

letters indicate statistically significant differences between groups (P <0.05), while identical letters indicate no significant difference between groups (P> 0.05).

Among the SNPs significantly associated with LMA, four constituted block B and seven constituted block C, respectively (Figure 10). Block B comprised five haplotypes, among which B1 (0.7933) and B2 (0.1650) exhibited higher haplotype frequencies. Association analysis revealed that the LMA of the B1/B1 haplotype combination was significantly larger than that of the B2/B2 haplotype combination; however, no significant differences were observed between the B1/B1 haplotype combination and the B3/B1 or B2/B1 haplotype combinations, nor between the B3/B1 or B2/B1 haplotype combinations and the B2/B2 haplotype combination. Block C comprised eight haplotypes, with C1 (0.7583) and C2 (0.2167) showing higher frequencies. Multiple comparison tests indicated that the LMA of the C1/C1 and C3/C1 haplotype combinations was significantly larger than that of the C2/C1 and C2/C2 haplotype combinations; no significant differences were found between the C1/C1 haplotype combination and the C3/C1 haplotype combination, nor between the C2/C1 haplotype combination and the C2/C2 haplotype combination. All loci in block B were located within the intronic region of the RAB30 gene, whereas block C was situated within the ncRNA-intronic region of LOC105604002. In summary, the haplotype combinations B1/B1 and C1/C1 represented the optimal haplotype combinations for LMA traits.

Among all SNPs significantly associated with BMR, three SNPs formed block D and eight SNPs formed block E, respectively (Figure 11). Block D comprised seven haplotypes, among which D1 (0.4367) and D2 (0.4) exhibited the highest frequencies; the BMR of the D3/D1 haplotype combination was significantly higher than that of the D3/D2 and D2/D2 haplotype combinations;

288 the BMR of the D4/D1 haplotype combination was significantly higher than that of the D2/D2
289 haplotype combination; no significant differences in BMR were observed among the haplotype
290 combinations D4/D1, D1/D1, D1/D5, D1/D2, and D3/D2; likewise, no significant differences were
291 found between the D3/D1 haplotype and each of the haplotype combinations D4/D1, D1/D1, D1/D5,
292 and D1/D2. Block E comprised four haplotypes, with E1 (0.9267) and E2 (0.0617) having the
293 highest frequencies; no significant differences in BMR were observed among the haplotype
294 combinations, although the E1E2 haplotype combination exhibited the highest BMR, while the
295 E1/E1 haplotype combination exhibited the lowest. All loci in block E are located within the
296 *LOC105606997* ncRNA_intronic region, and block E itself resides in the intronic region of *BRINP3*.
297 In conclusion, the haplotype combinations D2/D2 and E1/E1 are the superior haplotype
298 combinations for the BMR trait.

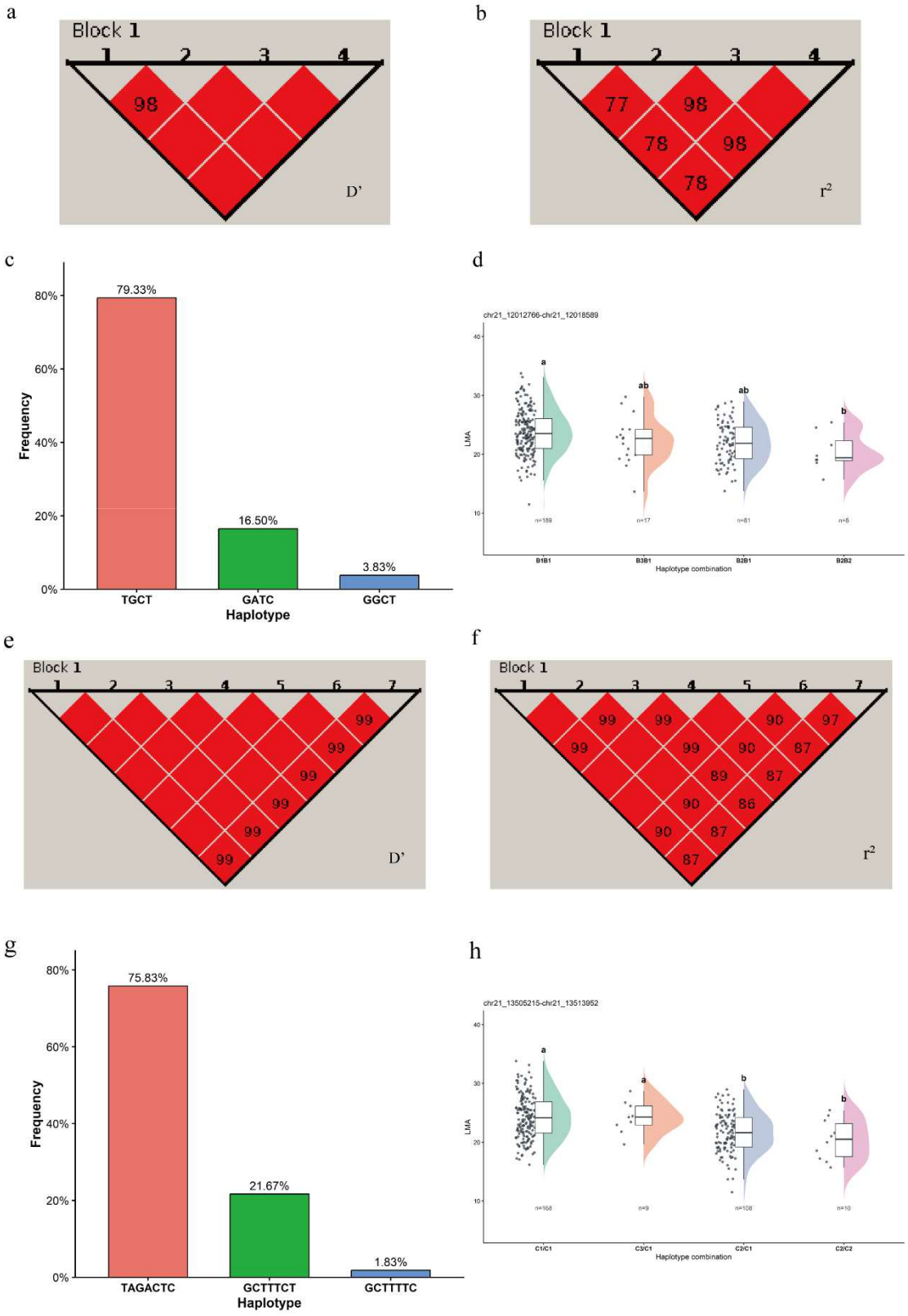

Fig. 10 Results of linkage disequilibrium analysis in regions associated with LMA, as well as haplotype-phenotype association analyses. (a) Heatmap of LD analysis based on D′ values; (b) Heatmap of LD analysis based on r^2 values; (c) Frequency distribution of major haplotypes, with the horizontal axis

representing haplotype sequences and the vertical axis representing the frequency of each haplotype in the population; (d) Phenotypic distribution of BFT across different haplotype combinations. Different letters indicate statistically significant differences between groups ($P < 0.05$), while identical letters indicate no significant difference between groups ($P > 0.05$).

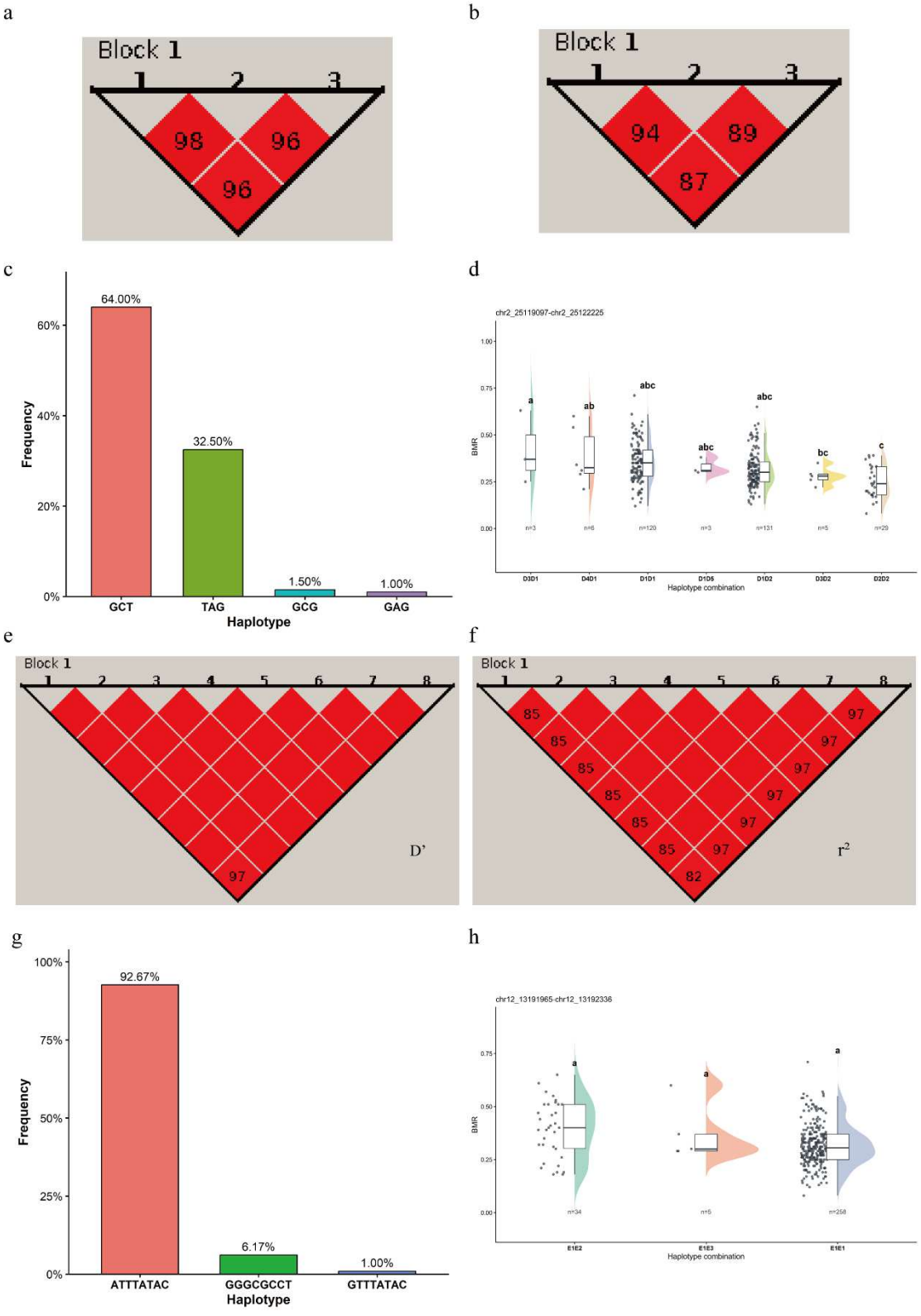

Fig. 11 Construction of haplotype blocks for significant SNPs associated with BMR in Suffolk sheep. a D' values, r^2 values and haplotype frequencies for the block formed by 19 significant SNPs on

chromosome 6 for the BMR trait; b D' values, r^2 values and haplotype frequencies for the BMR in a block formed by 28 significant SNPs on chromosome 6; c D' values, r^2 values and haplotype frequencies for the BMR in a block formed by 12 significant SNPs on chromosome 17; d D' values, r^2 values and haplotype frequencies for the BMR in a block formed by 7 significant SNPs on chromosome 21.

Discussion

Carcass quality traits of Suffolk sheep exhibit a complex genetic basis, influenced jointly by polygenic effects, small-effect genetic variants, and environmental factors[17]. In this study, three GWAS approaches were employed to conduct association analyses for BFT, LMT, LMA, and BMR. By integrating the results from these different methods, candidate loci detected by at least two methods were selected to enhance the robustness of the association signals. Phenotypic correlation analysis revealed a significant positive correlation between BFT and BMR ($r = 0.795$, $P < 0.001$). Given that BMR is calculated as the ratio of BFT to LMT, its correlation with BFT is partially influenced by this mathematical relationship; however, three shared candidate loci were detected between BFT and BMR, indicating that some genetic variants influencing BFT may also affect phenotypic variation in BMR. No shared candidate loci were identified between LMT and BMR. This observation may be attributed to the small number of loci detected for LMT and the consequent limited availability of reliable, comparable association signals. Therefore, the current findings indicate that, in the present study population and under the screening conditions employed, BMR and BFT exhibit a more pronounced overlap in their association signals.

Pearson correlation analysis revealed a weak and non-significant correlation between BFT and LMT ($r = 0.046$, $P > 0.05$), indicating that these two traits exhibit a relatively high degree of independence at the phenotypic level. For carcass composition improvement in meat-type sheep, while ensuring adequate fat coverage, breeders typically aim to reduce backfat thickness and increase lumbar muscle thickness. The weak phenotypic correlation between BFT and LMT suggests that, during selective breeding, it is possible to simultaneously reduce fat deposition and improve muscle development. BMR, defined as the ratio of BFT to LMT, comprehensively reflects an individual's

fat deposition and muscle development levels; therefore, a lower BMR value aligns better with the breeding objective of increasing the carcass lean meat proportion. Based on this rationale, the D2/D2 and E1/E1 haplotype combinations — characterized by lower BMR values — were identified in this study as promising haplotype combinations of great interest, and these combinations were subsequently designated as candidate molecular markers for the BMR trait.

The SNP chr6_89305500, which shows a significant association with backfat thickness, is annotated to the *BTC* gene. *BTC* is a member of the epidermal growth factor family; it can bind to and activate EGFR and other ERBB family receptors, thereby participating in biological processes such as cell proliferation and differentiation[18–20]. Previous GWAS studies have demonstrated that genetic variants affecting backfat thickness are involved in diverse biological processes, including energy homeostasis, lipogenesis, fatty acid metabolism, and insulin signaling[21]. Consequently, *BTC* has a potential biological role in subcutaneous adipose tissue development and lipid deposition. The SNP chr2_25332928 and block A are annotated to the *DIRAS2* gene, which belongs to the Ras-associated small GTPase family and exhibits GTP-binding activity and mediates signal transduction. Existing research indicates that *DIRAS2* can regulate signaling pathways such as AKT1-mTOR and RAS-MAPK[22]. Among these pathways, the AKT–mTOR axis plays pivotal roles in cell growth, energy metabolism, and lipid synthesis [23, 24], whereas the RAS–MAPK pathway is primarily involved in biological processes including cell proliferation, differentiation, and adipocyte formation [25, 26]. Furthermore, in recent multi-omics studies on pig growth and carcass traits, *DIRAS2* was integrated into a candidate gene regulatory network associated with growth and lipid deposition in Yorkshire pigs; these studies suggest that *DIRAS2* may be implicated in lipid metabolism, cell proliferation, and tissue remodeling[27]. Accordingly, *BTC* and *DIRAS2* are proposed as candidate genes for backfat thickness.

RAB30, which is associated with LMA, encodes a small GTPase of the *RAB* family, primarily localized to the Golgi apparatus, and participates in both structural maintenance of the Golgi apparatus and intracellular membrane-trafficking processes. Previous studies have demonstrated that members of the *RAB* family are involved in processes such as myogenic cell differentiation, myotube formation, and skeletal muscle regeneration; among these[28], *Rab5a* can promote

365 myogenic cell differentiation by regulating the IGF–AKT–mTOR signaling pathway[29]. Given the
366 intracellular membrane-trafficking function of *RAB30* and the roles of other *RAB* family members
367 in myogenesis, it is hypothesized that *RAB30* may influence muscle fiber development by
368 modulating intracellular material transport and related signaling pathways, thereby affecting the
369 LMA trait. Accordingly, this study identified *RAB30* as a candidate gene for the LMA trait in Suffolk
370 sheep.

371 Both LMA and LMT traits were jointly annotated to the gene *PCDH18*, which belongs to the
372 protocadherin family and is primarily involved in intercellular adhesion and the establishment of
373 cell junctions[30]. Previous studies have demonstrated that members of the cadherin family play
374 critical roles in myoblast adhesion, myotube formation, and skeletal muscle development[30, 31].
375 Given that *PCDH18* was associated with both LMT and LMA traits in the present study, this
376 observation suggests that this gene may contribute to the formation of lumbar-muscle traits in
377 Suffolk sheep by modulating cell adhesion and muscle tissue development. Therefore, in this study,
378 *PCDH18* was designated as a common candidate gene for both LMT and LMA traits.

379 The SNP loci and haplotype block E associated with the BMR were both mapped to the *BRINP3*
380 gene. Previous studies employing combined GWAS and transcriptome analyses have identified
381 *BRINP3* as a candidate gene linked to growth rate in sheep, and have demonstrated in Hu sheep
382 populations that the *BRINP3* g.16903465 T>C polymorphism is significantly correlated with growth
383 traits, suggesting that this gene may participate in the growth and development processes of
384 sheep[32]. Based on our present findings, we further suggest that, besides its association with
385 growth traits, this gene may also be linked to carcass composition traits in sheep. *KIF23* belongs to
386 the kinesin family and encodes the mitotic kinesin *MKLP1*, which plays critical roles during mitotic
387 telophase and cytokinesis[33]. Therefore, in this study, *BRINP3* and *KIF23* were designated as
388 candidate genes for the BMR trait in Suffolk sheep.

389 Single-locus GWAS treats each SNP as an independent marker for testing. Owing to linkage

disequilibrium, multiple SNPs are co-inherited as haplotype blocks, which better dissect multi-locus synergistic effects on complex quantitative traits[15, 34]. Herein, haplotype reconstruction and haplotype–phenotype association were conducted on shared SNPs. The favorable haplotype combinations of the five haplotype blocks were homozygous with stable genetic transmissibility. Hence, these five haplotype blocks can be utilized for molecular marker-assisted breeding targeting ultrasound-measured carcass quality traits in Suffolk sheep.

Conclusions

In this study, three statistical models were employed to conduct genome-wide association analyses (GWAS) on four ultrasound-measured carcass quality traits in 300 Suffolk sheep, identifying a total of 46 candidate SNPs significantly associated with these traits: 19 for backfat thickness (BFT), 1 for Longissimus dorsi muscle thickness (LMT), 12 for Longissimus dorsi muscle area (LMA), and 14 for Backfat to muscle thickness ratio (BMR). Through functional annotation screening, six key candidate genes were identified: *BTC* and *DIRAS2* as candidate genes for BFT; *RAB30* as a candidate gene for LMA; *PCDH18*, which influences both LMA and LMT; and *BRINP3* and *KIF23* as candidate genes for BMR. Utilizing these candidate SNPs, this study constructed five haplotype blocks associated with ultrasound-measured carcass quality traits in Suffolk sheep. These findings provide a theoretical foundation and candidate molecular resources for marker-assisted breeding of carcass quality in Suffolk sheep.

Materials and methods

Data Collection

In this study, 300 12-month-old Suffolk sheep (254 females and 46 males) were used as experimental animals, and all individuals were provided with whole-genome resequencing data and phenotypic data. Phenotypic data were provided from Inner Mongolia Siziwangqi Sainuo Sheep Breeding Technology Co., Ltd. The ultrasound-measured carcass quality traits data in this study were measured between 2020 and 2024, including BFT, LMT, LMA, and BMR. The measurement

procedures were as follows: (1) BFT was measured using a B-mode ultrasonic thickness meter to determine the thickness of subcutaneous fat between the 12th and 13th ribs, with the probe positioned 5 cm parallel to the dorsal midline of the sheep. (2) LMT was measured using the same ultrasonic device to determine the thickness of the longissimus dorsi muscle between the 12th and 13th ribs, with the probe positioned 5 cm parallel to the dorsal midline of the sheep. (3) LMA was measured using the same ultrasonic device to determine the area of the longissimus dorsi muscle between the 12th and 13th ribs, with the probe positioned 5 cm perpendicular to the dorsal midline of the sheep. (4) BMR was calculated as the quotient of BFT and LMT. All phenotypic measurements described above were performed on live sheep without causing any injury to the animals.

Source of Genotypic Data

Blood samples were collected from the jugular vein of live Suffolk sheep and transferred into EDTA anticoagulated blood collection tubes. The tubes were gently inverted for mixing and stored at -80℃ for subsequent analyses. DNA was isolated using a commercial genomic DNA extraction kit for blood samples. Whole-genome resequencing was carried out by Shijiazhuang Boridi Biotechnology Co., Ltd.

Library Construction and Sequencing

DNA (200 ng) passing quality control was used for library construction with the GenoBaits DNA Library Prep Kit. Following DNA fragmentation, end repair, A-tailing, adapter ligation, fragment size selection, and PCR amplification, sequencing libraries with insert sizes of approximately 300-350 bp were generated. Library quantification and quality checking were performed using Qubit 2.0 and qPCR. Qualified libraries were subjected to paired-end sequencing (PE150) on the DNBSEQ-T7 platform.

SNP Detection and Quality Control

Raw reads were processed with fastp (v0.20.0) for adapter and low-quality base filtering[35]. BWA (v0.7.17)[36] was applied to map clean reads against the sheep reference genome oar_v4.0 (GCF_000298735.2). SAMtools (v1.90) was used to convert SAM files to BAM files[37]. Qualimap and SAMtools were utilized to evaluate mapping status, coverage and sequencing depth. GATK (v1.90)was used for SNP calling to generate the VCF file[38].

To ensure the reliability of GWAS results, this study performed rigorous quality control analysis on the raw data using PLINK(v1.9) software[39]. The specific quality control criteria were as follows:

(1) Sample call rate greater than 90%; (2) SNP call rate greater than 98%; (3) Minor allele frequency greater than 5%; (4) P-value from the Hardy-Weinberg equilibrium test greater than 10^{-6}. Following quality control, a total of 20,182,599 high-quality SNPs were retained. The CMplot package in R (v4.5.2) was used to visualise the genomic distribution characteristics of the quality-controlled SNPs.

Population Genetic Structure Analysis and Visualisation

PLINK (v1.9) was used to calculate the principal components of the Suffolk sheep population and to construct the Suffolk sheep population identity-by-state (IBS) matrix and genomic relationship matrix (G-matrix), which were then visualised using R (v4.5.2).

Genome-wide association analysis

To compare different genome-wide association study (GWAS) methods, three distinct models—GCTA-fastGWA-MLM, GEMMA-MLM and BLINK—were used for GWAS.

The fastGWA-MLM model is based on the mixed linear model and was implemented using

GCTA (v1.94.1) [10]. The model is shown below:

$$y = X\alpha + M\beta + Zu + e$$

where y denotes the phenotypic value of the trait, α denotes fixed effects, β denotes SNP marker effects, u denotes additive effects, X denotes the fixed effects matrix, M denotes the SNP marker effects matrix, Z denotes the random effects matrix, and e denotes the random error. The first three principal components (PCs), sex and year were incorporated into the model. Since the farm adopted group housing by year, the farm effect was not included in the model.

The GEMMA-MLM approach used the univariate linear mixed-effects model implemented in GEMMA (v0.98.5)[12]; the model is as follows:

$$y = W\alpha + x\beta + u + \varepsilon$$

Here, y denotes the individual phenotype; W denotes the covariate matrix comprising the intercept, sex, age, and the first three principal components; α denotes the fixed effect coefficients for these covariates; x denotes the genotype vector of the SNP under investigation; and β denotes the SNP effect value.

u denotes the random effects of an individual arising from whole-genome genetic background, genetic relatedness, and other factors; it follows a normal distribution.

$$u \sim N_n\left(0, \sigma_g^2 K\right)$$

Here, K denotes the genomic relatedness matrix, and σ_g^2 denotes the genetic variance.

ε denotes the residual, which follows a normal distribution.

$$\varepsilon \sim N_n(0, \ \sigma_e^2 I_n)$$

Here, I_n denotes the identity matrix, and σ_e^2 denotes the residual variance.

The BLINK model was based on BLINK-C (v0.01)[13], the model is presented below:

$$y = W\alpha + x_j\beta_j + \sum_{k=1}^{i} M_k \gamma_k + \varepsilon$$

Here, y denotes the vector of phenotypic values, W denotes the fixed-effect design matrix, α denotes the fixed-effect vector corresponding to sex, age and the first 10 principal components, x_j denotes the genotype vector of the j-th SNP under testing, β_j denotes its genetic effect, M_k denotes the genotype vector of the k-th pseudo QTN, γ_k denotes the corresponding genetic effect of the pseudo QTN, i denotes the total number of pseudo QTNs included in the model, and ε denotes the random residual.

GO Functional Annotation and KEGG Enrichment Analysis of Candidate Genes

Based on the physical positions of significantly and suggestively associated SNPs, ANNOVAR (version 2018Apr16)[40] was used for functional annotation of these SNPs with the sheep reference genome Oar_v4.0 (GCF_000298735.2), Afterwards, the DAVID database[41] was applied to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of candidate genes, to explore the biological processes and signaling pathways potentially regulated by these candidate genes.

Chain imbalance analysis

Haplotype blocks for candidate SNP loci were constructed Using Haploview software[42], and R (v4.5.2) was employed to perform haplotype-phenotype association analyses and multiple comparisons. To ensure the reliability of the statistical analyses, haplotype combinations with fewer than three samples were excluded.

Abbreviations

GWAS, Genome-Wide Association Study; SNPs, Single Nucleotide Polymorphism; BFT, backfat thickness; LMT, Longissimus dorsi muscle thickness; LMA, Longissimus dorsi muscle area; BMR, Backfat to muscle thickness ratio; MLM, Mixed Linear Model; GRM, genomic relationship matrix; BIC, Bayesian Information Criterion; LD, Linkage Disequilibrium; QQ plot, Quantile-quantile plot; IBS matrix, identity by state genetic distance matrix; GO functional annotation, Gene Ontology biological function annotation; KEGG Enrichment Analysis Kyoto Encyclopedia of Genes and Genomes pathway enrichment; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; DNA, Deoxyribonucleic Acid; qPCR, quantitative real-time polymerase chain reaction; QTN, Quantitative Trait Nucleotide; PCA, Principal Component Analysis

Acknowledgments

The authors are grateful to the College of Animal Science, Inner Mongolia Agricultural University for the use of experimental facilities, and to Inner Mongolia Siziwang Banner Saino Breeding Sheep Technology Co., Ltd. for providing the experimental samples.

Authors' contributions

Y.H. conceptualized this study; J.L. helped in the methodology; X.W., C.Z., D.W. and J.P. helped in sample collection; X.W. and L.Z. helped with software; Y.R. helped in formal analysis; S.Z. and Y.Z. helped in visualization; X.W., C.Z. and H.Z. helped in data curation; Y.H., J.L. helped in writing the original draft. Y.Z., R.W. and Y.L. helped in writing-reviewing and editing the manuscript. Y.L. and R.W. helped in funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding

This reported work was supported by the Inner Mongolia Autonomous Region “Leading the Charge with Open Competition” topic (2022JBGS0024), the Inner Mongolia Autonomous Region Breeding Joint

Research Project (YZ2023011), the Fundamental Research Funds of Directly Subordinate Universities of Inner Mongolia Autonomous Region (BR251201), the Inner Mongolia Education Department Special Research Project For First Class Disciplines (YLXKZX-NND-007), the Science and Program of Inner Mongolia Autonomous Region (2025KYPT0103), the Fundamental Research Funds of Directly Subordinate Universities of Inner Mongolia Autonomous Region (BR261013), the Conservation and Innovative Utilization of Sheep Genetic Resources (BR261201) and the Fundamental Research Funds of Directly Subordinate Universities of Inner Mongolia Autonomous Region (BR251004).

Availability of data and material

All raw genome sequencing data generated during the current study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database with the BioProject accession numbers PRJNA1483776 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1483776).

Declarations

Ethics approval and consent to participate

All animal experiments were performed in accordance with the Guidelines for Experimental Animals of the Ministry of Science and Technology (Beijing, China) and were approved by the Scientific Research and Academic Ethics Committee of Inner Mongolia Agricultural University and the Biomedical Research Ethics of Inner Mongolia Agricultural University (Approval No. [2020] 056). Written informed consent was obtained from the owners for the participation of animals in this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Wang W, Zhang X, Wei H, Wang S, Ye Y, He L, et al. Effects of feeding systems on the growth performance, carcass characteristics, and meat quality in sheep: A meta-analysis. Animals. 2024;14:2738. https://doi.org/10.3390/ani14182738.

2. Prache S, Schreurs N, Guillier L. Review: Factors affecting sheep carcass and meat quality attributes. Animal. 2022;16:100330. https://doi.org/10.1016/j.animal.2021.100330.

3. Rocha AO, Saleem AH, Gunes HY, De Carvalho FE, Wilson CS, Freking BA, et al. Estimates of genetic parameters and genetic trends for body weight, body composition, and reproductive traits in U.S. suffolk sheep. J Anim Sci. 2026;:skag124. https://doi.org/10.1093/jas/skag124.

4. Xu H, Zhang Z, Zhao W, Zhuang Y, Hou X, He Y, et al. Deep learning-driven automated carcass segmentation and composition quantification in live pigs via large-scale CT imaging and its application in genetic analysis of pig breeding. Smart Agric Technol. 2025;11:100898. https://doi.org/10.1016/j.atech.2025.100898.

5. Tan X, He Z, Fahey AG, Zhao G, Liu R, Wen J. Research progress and applications of genome-wide association study in farm animals. Anim Res One Health. 2023;1:56–77. https://doi.org/10.1002/aro2.14.

6. Veller C, Coop GM. Interpreting population- and family-based genome-wide association studies in the presence of confounding. PLOS Biol. 2024;22:e3002511. https://doi.org/10.1371/journal.pbio.3002511.

7. Tan L, Zhao Y. Principle, optimization and application of mixed models in genome-wide association study. Sci Agric Sin. 2023;56:1617–32. https://doi.org/10.3864/j.issn.0578-1752.2023.09.001.

8. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the

572 application of mixed-model association methods. Nat Genet. 2014;46:100–6.
573 https://doi.org/10.1038/ng.2876.

574 9. Runcie DE, Crawford L. Fast and flexible linear mixed models for genome-wide
575 genetics. PLOS Genet. 2019;15:e1007978.
576 https://doi.org/10.1371/journal.pgen.1007978.

577 10. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex
578 trait analysis. Am J Hum Genet. 2011;88:76–82.
579 https://doi.org/10.1016/j.ajhg.2010.11.011.

580 11. A resource-efficient tool for mixed model association analysis of large-scale data |
581 nature genetics. https://www.nature.com/articles/s41588-019-0530-8. Accessed 14 Oct
582 2026.

583 12. Genome-wide efficient mixed-model analysis for association studies | nature
584 genetics. https://www.nature.com/articles/ng.2310. Accessed 8 Oct 2026.

585 13. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: A package for the next level
586 of genome-wide association studies with both individuals and markers in the millions.
587 GigaScience. 2019;8:giy154. https://doi.org/10.1093/gigascience/giy154.

588 14. Kaseja K, Lambe N, Yates J, Smith E, Conington J. Genome wide association studies
589 for carcass traits measured by video image analysis in crossbred lambs. Meat Sci.
590 2024;214:109518. https://doi.org/10.1016/j.meatsci.2024.109518.

591 15. Srivastava S, Srikanth K, Won S, Son J-H, Park J-E, Park W, et al. Haplotype-based
592 genome-wide association study and identification of candidate genes associated with
593 carcass traits in hanwoo cattle. Genes. 2020;11:551.
594 https://doi.org/10.3390/genes11050551.

595 16. Zhu S, Guo T, Zhao H, Qiao G, Han M, Liu J, et al. Genome-wide association study
596 using individual single-nucleotide polymorphisms and haplotypes for erythrocyte traits
597 in alpine merino sheep. Front Genet. 2020;11.
598 https://doi.org/10.3389/fgene.2020.00848.

599 17. Rocha AO, Saleem AH, Gunes HY, De Carvalho FE, Wilson CS, Freking BA, et al.
600 Estimates of genetic parameters and genetic trends for body weight, body composition,
601 and reproductive traits in U.S. suffolk sheep. J Anim Sci. 2026;:skag124.
602 https://doi.org/10.1093/jas/skag124.

603 18. Dahlhoff M, Wolf E, Schneider MR. The ABC of BTC: Structural properties and
604 biological roles of betacellulin. Semin Cell Dev Biol. 2014;28:42–8.

https://doi.org/10.1016/j.semcdb.2014.01.002.

19. Dunbar AJ, Goddard C. Structure-function and biological role of betacellulin. Int J Biochem Cell Biol. 2000;32:805–15. https://doi.org/10.1016/S1357-2725(00)00028-5.

20. Chen J, Zeng F, Forrester SJ, Eguchi S, Zhang M-Z, Harris RC. Expression and function of the epidermal growth factor receptor in physiology and disease. Physiol Rev. 2016;96:1025–69. https://doi.org/10.1152/physrev.00030.2015.

21. Genetic architecture and major genes for backfat thickness in pig lines of diverse genetic backgrounds | genetics selection evolution | springer nature link. https://link.springer.com/article/10.1186/s12711-021-00671-w. Accessed 18 Oct 2026.

22. Sutton MN, Yang H, Huang GY, Fu C, Pontikos M, Wang Y, et al. RAS-related GTPases *DIRAS1* and *DIRAS2* induce autophagic cancer cell death and are required for autophagy in murine ovarian cancer cells. Autophagy. 2018;14:637–53. https://doi.org/10.1080/15548627.2018.1427022.

23. mTOR at the nexus of nutrition, growth, ageing and disease | nature reviews molecular cell biology. https://www.nature.com/articles/s41580-019-0199-y. Accessed 13 Oct 2026.

24. Porstmann T, Santos CR, Griffiths B, Cully M, Wu M, Leevers S, et al. SREBP activity is regulated by mTORC1 and contributes to akt-dependent cell growth. Cell Metab. 2008;8:224–36. https://doi.org/10.1016/j.cmet.2008.07.007.

25. Bost F, Aouadi M, Caron L, Binétruy B. The role of MAPKs in adipocyte differentiation and obesity. Biochimie. 2005;87:51–6. https://doi.org/10.1016/j.biochi.2004.10.018.

26. Aubert J, Belmonte N, Dani C. Role of pathways for signal transducers and activators of transcription, and mitogen-activated protein kinase in adipocyte differentiation. Cell Mol Life Sci CMLS. 1999;56:538–42. https://doi.org/10.1007/s000180050450.

27. Zhao A, Qiao J, Kuang R, Chen Z, Tan Y, Zhu M, et al. Multi-omics characterization identifies conserved candidate gene and reveals breed-specific regulatory mechanisms underlying growth-related traits in pigs. BMC Genomics. 2026;27:684. https://doi.org/10.1186/s12864-026-12979-2.

28. Kelly EE, Giordano F, Horgan CP, Jollivet F, Raposo G, McCaffrey MW. Rab30 is required for the morphological integrity of the golgi apparatus. Biol Cell. 2012;104:84–101. https://doi.org/10.1111/boc.201100080.

29. Cong XX, Gao XK, Rao XS, Wen J, Liu XC, Shi YP, et al. Rab5a activates IRS1 to

637 coordinate IGF-AKT-mTOR signaling and myoblast differentiation during muscle
638 regeneration. Cell Death Differ. 2020;27:2344–62. https://doi.org/10.1038/s41418-020-
639 0508-1.

640 30. Mozin E, Massouridès E, Mournetas V, Lièvre C, Bourdon A, Jackson DL, et al.
641 Dystrophin deficiency impairs cell junction formation during embryonic myogenesis
642 from pluripotent stem cells. iScience. 2024;27:110242.
643 https://doi.org/10.1016/j.isci.2024.110242.

644 31. M-cadherin regulates the formation of adherens junctions and secondary myofibers
645 during fetal myogenesis to determine adult myofiber number and muscle mass. Stem
646 Cell Rep. 2026;21:102812. https://doi.org/10.1016/j.stemcr.2026.102812.

647 32. Zhao L, Li F, Zhang X, Yuan L, Tian H, Xu D, et al. Integrating a genome-wide
648 association and transcriptome analysis to provide molecular insights into growth rates
649 in sheep. J Integr Agric. 2026;25:3352–67. https://doi.org/10.1016/j.jia.2024.08.011.

650 33. Liljeholm M, Irvine AF, Vikberg A-L, Norberg A, Month S, Sandström H, et al.
651 Congenital dyserythropoietic anemia type III (CDA III) is caused by a mutation in kinesin
652 family member, *KIF23*. Blood. 2013;121:4791–9. https://doi.org/10.1182/blood-2012-
653 10-461392.

654 34. Sehgal D, Mondal S, Crespo-Herrera L, Velu G, Juliana P, Huerta-Espino J, et al.
655 Haplotype-based, genome-wide association study reveals stable genomic regions for
656 grain yield in CIMMYT spring bread wheat. Front Genet. 2020;11:589490.
657 https://doi.org/10.3389/fgene.2020.589490.

658 35. Chen S, Zhou Y, Chen Y, Gu J. fastp: An ultra-fast all-in-one FASTQ preprocessor.
659 Bioinformatics. 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.

660 36. Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler
661 transform. Bioinformatics. 2009;25:1754–60.
662 https://doi.org/10.1093/bioinformatics/btp324.

663 37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence
664 alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
665 https://doi.org/10.1093/bioinformatics/btp352.

666 38. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
667 genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA
668 sequencing data. Genome Res. 2010;20:1297–303.
669 https://doi.org/10.1101/gr.107524.110.

670 39. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A
671 tool set for whole-genome association and population-based linkage analyses. Am J
672 Hum Genet. 2007;81:559–75. https://doi.org/10.1086/519795.

673 40. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants
674 from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164–e164.
675 https://doi.org/10.1093/nar/gkq603.

676 41. Sherman BT, Panzade G, Dotrang T, Hao M, Xu L, Li X, et al. DAVID: A web server for
677 functional annotation and functional enrichment analysis of gene lists (2025 update).
678 Nucleic Acids Res. 2026;54:W367–73. https://doi.org/10.1093/nar/gkag470.

679 42. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and
680 haplotype maps. Bioinformatics. 2005;21:263–5.
681 https://doi.org/10.1093/bioinformatics/bth457.

682

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials1.docx
- SupplementaryFigure1..pdf
- SupplementaryFigure3..pdf
- SupplementaryFigure2..pdf
- SupplementaryTable.1.xls