

RESEARCH ARTICLE

Integrating genomics and transcriptomics reveals candidate genes affecting loin muscle area in Huaxi cattle

Qingqing Xue^{1,2*}, Lili Du^{2*}, Tianyu Deng², Mang Liang², Keanning Li², Li Qian², Shiyuan Qiu², Yan Chen², Xue Gao², Lingyang Xu², Zezhao Wang², Caihong Zheng², Lupei Zhang², Junya Li², Huijiang Gao^{2*}

1 Department of College of Animal Science and Veterinary Medicine, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang, China, **2** Department of Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

☞ These authors contributed equally to this work.

* gaohuijiang@caas.cn



OPEN ACCESS

Citation: Xue Q, Du L, Deng T, Liang M, Li K, Qian L, et al. (2025) Integrating genomics and transcriptomics reveals candidate genes affecting loin muscle area in Huaxi cattle. PLoS One 20(5): e0322026. <https://doi.org/10.1371/journal.pone.0322026>

Editor: Sayed Haidar Abbas Raza, South China Agricultural University, CHINA

Received: November 28, 2024

Accepted: March 14, 2025

Published: May 9, 2025

Copyright: © 2025 Xue et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The RNA-seq data has been submitted to the figshare database at the following DOIs: (<https://doi.org/10.6084/m9.figshare.28414973.v1> and <https://doi.org/10.6084/m9.figshare.28408877.v1>).

Funding: This research was funded by the Science and Technology Project of Inner

Abstract

Loin muscle area (LMA) is an indicator of carcass composition and is related to weight gain, animal musculature and meat quality traits. Therefore, integrating multi-omics data to reveal candidate genes affecting LMA has attracted extensive attention. We used the combined analysis method of GWAS and RNA-seq to find the candidate genes that affect the size of LMA. The association of 770K SNPs with the LMA captured four significant SNPs within or near three genes. Additionally, seven overlapping genes regarding LMA were determined via the analysis of differentially expressed genes (DEGs) and weighted gene co-expression network analysis (WGCNA). There is an overlapping gene (*CD93*) between the results of GWAS and DEGs. Through functional enrichment analysis of the above genes, candidate genes were identified as *THBD*, *CD93*, *RIMS2*, *PLP1*, *SNCA*, and *NDUFS8*, and it was found that they mainly affected the size of LMA by affecting muscle fiber diameter, muscle cell development, differentiation, and function. The findings provide valuable molecular insights into the mechanisms that influence LMA content in beef cattle.

Introduction

Beef is a source of high-quality protein, beneficial fatty acids, and various trace elements in the human diet, which plays an essential role in maintaining the healthy diet [1]. Improving meat quality traits in beef cattle is a key issue facing the beef cattle breeding industry and a hot research topic for beef cattle breeding scholars. Loin muscle area (LMA) is the surface area of the longissimus dorsi muscle (ribeye) at the 12th to 13th rib interface of beef cattle. Previous studies have estimated the heritability of LMA to range from 0.24 to 0.32 [2–4], indicating that it is a moderately heritable trait that can be improved through selective breeding. The identification of genes that

Mongolia Autonomous Region, China (2020GG0210), the Program of National Beef Cattle and Yak Industrial Technology System (CARS-37) and STI2030-Major Projects (2023ZD04048 and 2023ZD0404803 to C.Z.). The funders are involved in data collection.

Competing interests: The authors have declared that no competing interests exist.

influence LMA in beef cattle is important, especially in the improvement of meat quality and breeding. LMA is closely related to net meat yield and slaughter rate, implying that net meat yield and slaughter rate can be increased by increasing the LMA, which is essential for improving the economic value of beef cattle.

Huaxi cattle, is a new specialized beef cattle breed in China, which consists of 24/32 Northern America beef Simmental cattle, 5/32 of dual-purpose Simmental cattle from Germany and Austria, 1/32 Charolais cattle, 1/32 Sanhe cattle, and 1/32 Mongolian cattle. Since the beginning of breeding in 1978, Huaxi cattle have shown good performance in growth rate, slaughter rate and net meat rate [5].

Genome-wide association analysis (GWAS) refers to the analysis of the association between the molecular marker data obtained from the scan and phenotypic traits through genome-wide scanning, to identify genes that affect certain phenotypic or quantitative traits. As an effective molecular genetics method, GWAS plays an important role in the identification of functional genes for quantitative traits and the study of genetic variation of important economic traits in livestock and poultry [6]. A series of genetic loci significantly associated with beef carcass traits have been identified using GWAS. For example, *CARTPT* gene affects LMA by encoding multiple peptides that affect tissue growth and development [7]. The *CHKA* gene [8] influences backfat thickness (BFT) by influencing lipid biosynthesis.

RNA sequencing (RNA-seq) enables the study of gene function and regulatory mechanisms at the global level and the discovery of candidate genes associated with specific functions [9]. The analysis of differentially expressed genes (DEGs) and weighted gene co-expression network analysis (WGCNA) were two mainstream strategies for transcriptomics analysis [10]. The analysis of DEGs is a basic method to compare the differences and similarities of gene expression between two groups of different samples, which can obtain the genes that are significantly up-regulated and down-regulated in one group of samples compared with the other group. Further study of the function of these DEGs will eventually detect the vital genes that affect the target traits. The WGCNA approach aims to find co-expressed gene modules and to explore the associative relationship between the gene network and the phenotype of interest, as well as the core genes in the network. Santos Silva used RNA-seq technology to identify DEGs in the longissimus dorsi muscle of Neolore cattle and predicted that three key genes were related to LMA [11]. Wang H [12] used the WGCNA method to detect six gene modules significantly related to sex between obese boars and sows and finally identified seven DEGs affecting sex. Yang M [13] identified the *DBI* gene as a candidate gene affecting intramuscular fat in Beijing black pigs using a combined analysis of GWAS and RNA-seq. The *FGL1* gene had a significant effect on LMA [14]. GWAS was performed on Simmental beef quality traits to discover 19 candidate SNP loci and 11 candidate genes associated with five meat quality traits such as meat color and LMA, among which *KCTD16* and *LOC506594* were related to LMA [15]. Although numerous studies on LMA have performed, DEGs affecting LMA in Huaxi cattle still need to be further determined. Therefore, this study aimed to identify candidate genes affecting the size of LMA in Huaxi cattle using the combined analysis of GWAS and RNA-seq.

Materials and methods

Institutional review board statement

The animal study protocol was approved by the Animal Welfare and Ethics Committee of Institute of Animal Science, Chinese Academy of Agricultural Sciences (IAS2020–21; Beijing, China).

Samples collection and measurement

The 1,507 experimental individuals in this study were from the Chinese Huaxi cattle resource group, which was established by the Bovine Genetic Breeding Team of the Institute of Animal Sciences of the Chinese Academy of Agricultural Sciences. Blood samples were collected for genotyping during routine farm quarantine procedures. All experimental animals were slaughtered in strict accordance with GB/T 27642–2011. Cattle to be slaughtered are electrocuted and then bled for slaughter. At the time of slaughter, Samples of the left longissimus dorsi muscle from all male individuals were collected and placed into 5 mL freezing tubes, which were immediately frozen in liquid nitrogen and subsequently transferred to a liquid nitrogen tank for long-term storage. These samples will be used for transcriptome sequencing. The section of 12 rib Loin muscle of Huaxi cattle was drawn with sulfuric acid paper [11,16]. Subsequently, the drawn LMA is placed on the measuring board, the number of dots in the measuring board is counted, and the LMA is calculated according to the number of dots, where one dot represents an area of 1 cm². Bulls with an average body weight of 665 kg and an average age of 25 months were selected and divided into two groups for differential expression analysis: a high loin muscle area group (HLMA, n=6) and a low loin muscle area group (LLMA, n=6). Additionally, WGCNA analysis was conducted on bulls (n=20) with the same average body weight and age.

Genotyping and quality control

Genomic DNA was extracted from blood samples using a TIANamp Blood DNA Kit (Tiagen Biotech Company Limited, Beijing, China). DNA with 260/280 ratios between 1.8 and 2.0 was further analyzed and genotyped by 770K SNPs Illumina BovineHD BeadChip (Illumina Inc., San Diego, California, USA).

Using PLINK (v1.90) for individuals and SNPs quality control. The quality control criteria are as follows: SNPs with deletion rates greater than 0.05, individuals with genotype deletion greater than 10% [17], SNPs with minor allele frequency (<0.05) and Hardy-Weinberg balance test ($P < 1e-6$) [18], and individuals with heterozygosity levels higher than 3 standard deviations were excluded [19]. Finally, 1,488 samples and 605,671 SNPs were remained for subsequent analysis.

Single-trait GWAS analysis

In this study, according to the actual situation of resource groups, the significance test was conducted on the variables influencing the original phenotypic data, and the significant variables were used to adjust the phenotypic values accordingly. The model is as follows [10]:

$$y_{ijk} = \mu + sex_i + field_j + year_k + e_{ijk} \quad (1)$$

y_{ijk} is the phenotypic value; μ is the population mean; sex , $field$, and $year$ as fixed factors; e is the residual after the normal distribution $N(0, I\sigma^2)$, where I is the identity matrix. Through the significance test, sex and $year$ were selected as fixed factors to be incorporated into the linear model, and the phenotypic values were corrected.

Subsequently, PLINK (v1.90) software was used to calculate the principal components to correct the population stratification, and then using the twstats method to test the significance of the principal components. The results showed that the P-values of the first three principal components were less than 0.05, thus the first three principal components were selected to be as covariables.

This study selected the first three principal components to control for population stratification effects. The Fixed and Random Model Circulating Probability Unification (FarmCPU) method was used to evaluate the population structure, as this model offers higher statistical power and computational efficiency compared to general linear models and mixed linear models. The rMVP package (v1.0.5) was used to explore the association between genomic variations and phenotypes [20].

The significance threshold for GWAS was set at $P=0.05/N$, where N is the number of quality-controlled SNPs. SNPs with p -values below this threshold were considered significant. These significant SNPs were then mapped to the bovine reference genome ARS-UCD 1.2, and linked genes within 100Kb upstream and downstream of these SNPs were identified using the BioMart module on the Ensembl website. These genes were further evaluated by combining literature reviews and gene function annotations to determine their potential influence on LMA trait.

RNA-seq library preparation

Following the manufacturer's instructions, the TRIzol reagent (Invitrogen, Life Technologies) was used to extract total RNA from LMA of Huaxi. The concentration, quality, and integrity of RNA samples were examined by Nano photometer Spectrophotometer (Thermo Fisher Scientific, MA, USA) and RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). To ensure suitability for constructing sequencing libraries, all RNA samples had OD260/OD280 values between 1.8 and 2.0, and an RNA integrity number (RIN) greater than 7.0. The qualified samples were then sent to the Illumina NovaSeq 6000 platform for paired-end RNA sequencing (read length 150 bp). The RNA sequencing was completed by Beijing Novogene Technology Co., Ltd.

RNA-seq data preprocessing

After the sequencing data were obtained, FastQC (v0.11.4) software was used to detect the quality of original Reads (mass fraction, GC content, length distribution, etc.), and Trimmomatic (v0.39) was employed for quality control including removing low-quality sequences and splitter sequences, resulting high-quality clean reads for further analysis. The index was constructed using Hisat2 (v2.2.1) software and clean reads were compared to the bovine reference genome (ARS-UCD 1.2), and then the SAM files were converted to BAM files by SAMtools (v1.11) software. Finally, the raw count file of gene expression was obtained by using FeaturesCounts (v1.5.2) software. The reading numbers of all samples were combined. Per kilobases per million read (FPKM) values were calculated using R.

Identification of differential gene expression

The high (H) and low (L) LMA groups each consisted of six Huaxi cattle, with one group having high LMA and the other having low LMA. The raw count of individuals in the two groups was analyzed by DESeq2 (v1.18.0) software. The criteria for DEGs were $|\log_2\text{FoldChange}| \geq 1$ and $\text{FDR} < 0.05$. Subsequently, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment [21] and Gene Ontology (GO) annotation [22] were performed for these genes using the clusterProfiler R package (v4.12.0) [23]. GO annotation covers three categories: Biological process (BP), Cellular Component (CC), and Molecular function (MF). The significance threshold of enrichment was set as $P < 0.05$.

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) is used to identify clusters (modules) of highly correlated genes, summarize these modules using either the module eigengene or an intramodular hub gene, and explore relationships between modules, as well as their associations with external sample traits through eigengene network analysis. Additionally, WGCNA allows for the calculation of module membership measures [24]. This experiment used the WGCNA R package (v1.72) to construct a gene co-expression network of all genes and LMA trait in 20 samples. Hub genes in the

significance module were defined with Gene significance ($|GS|$) was greater than 0.2 and the absolute value of Module membership ($|MM|$) was greater than 0.8.

Construction of protein interaction network and further identification of core genes

Constructing protein-protein interaction (PPI) networks for Hub genes within each significant module [25] via the STRING database. Then, the intersection and visualization of the top ten significant Hub genes of each significant module based on Maximum Clique Centrality (MCC), Neighborhood Component Centrality (MNC), and Degree were identified by the CytoHubba plug-in of Cytoscape software [26]. Finally, The selected genes were further intersected with DEGs for additional analysis. KEGG pathway enrichment and GO annotation were subsequently conducted on the KOBAS website for both these genes and the genes identified through GWAS.

Results

GWAS results

After quality control of the selected samples, 1,488 samples as well as 605,671 SNPs were subjected to GWAS analysis.

The heritability of LMA was calculated as 0.43 using GCTA [27], which is high heritability, indicating that the trait can be improved by genetic methods. The Manhattan plot in Fig 1A showed that four significant SNP loci were screened at the genome-wide level ($P=0.05/605,671$); and the QQ plot in Fig 1B indicated that there was a consistent relationship

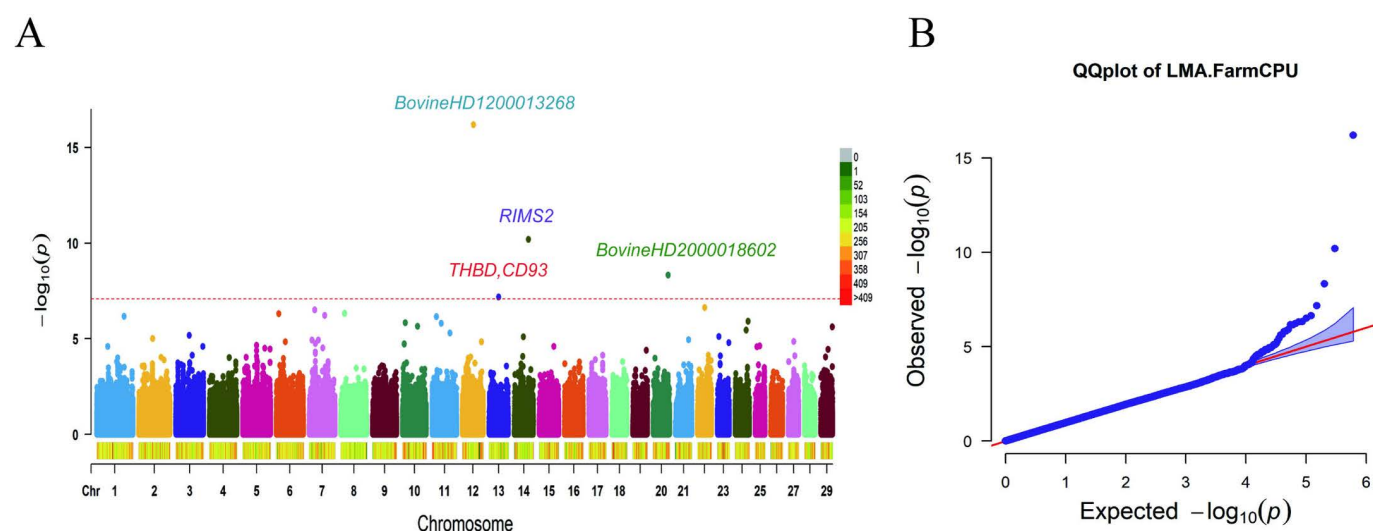


Fig 1. Genome-wide association analysis of LMA trait in the Huaxi cattle population. (A) Manhattan plot. The red dotted lines indicated the significant threshold ($P=0.05/605,671$). (B) QQ plot.

<https://doi.org/10.1371/journal.pone.0322026.g001>

Table 1. The annotation of significant SNPs.

SNP ID	Chr	POS	REF	ALT	SE	P.adj
BovineHD1200013268	12	48368923	A	G	0.568644443	6.41E-17
BovineHD1300012261	13	42247767	A	G	0.418439684	6.68E-08
BovineHD1400017455	14	62769117	A	C	0.91433653	6.31E-11
BovineHD2000018602	20	65011159	A	G	0.752846917	4.68E-09

<https://doi.org/10.1371/journal.pone.0322026.t001>

between the actual and theoretical values for most of the loci, which effectively demonstrated the effectiveness of this method in population control [28]. The basic information of these four significant SNP is shown in Table 1.

Based on the BioMart module, these significant SNPs were identified to be located in or close to three coding genes. BovineHD1300012261 on Bos taurus autosomes (BTA) 13 was located in or close to the *THBD* gene and *CD93* gene; BovineHD1400017455 on BTA 14 was located in the *RIMS2* gene. The two significant SNPs located on chromosome 12 as well as 20 did not detect reveal any associated genes.

Results of transcriptome sequencing data comparison

S1 Table in S1 File demonstrated the basic statistics for the LMA in the transcriptome 12 experimental Huaxi cattle. After quality control of the transcriptome sequencing data, an average of 21,848,620 clean reads were obtained for each sample. The comparison results showed that the multiple mapping rate, the unique mapping rate, the total mapping rate, and the unmapped rate of the 12 samples were 2.74%~3.79%, 90.08%~91.83%, and 95.97%~97.12%, and 5.16%~6.69%, respectively (S2 Table in S1 File), which indicated that the sequencing results were better to be used for the subsequent analysis.

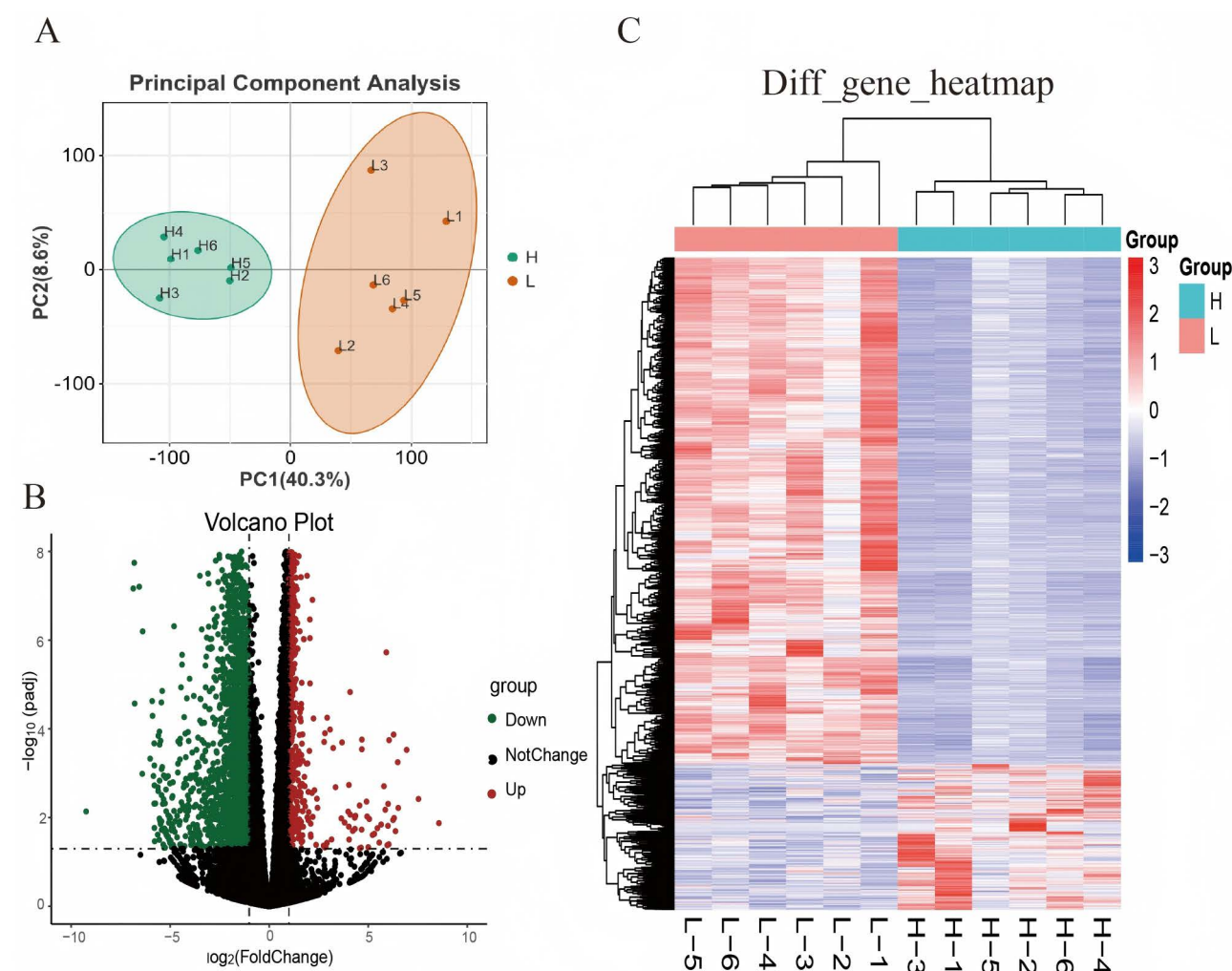


Fig 2. Identification of DEGs in the high LMA and low LMA groups. (A) PCA of the samples. Green and red dots indicated samples with high LMA and low LMA, respectively. (B) Volcano plot of DEGs. Red dots represented significantly up-regulated genes; dark-green dots indicated significantly down-regulated genes; black dots indicated genes with no significant effect. (C) Heatmap of DEGs. H: high LMA group; L: low LMA group.

<https://doi.org/10.1371/journal.pone.0322026.g002>

Identification of DEGs

The individuals in high LMA and low LMA groups were screened according to the values of LMA trait. Principal component analysis was performed on the gene expression data of the two groups of individuals after quality control. Fig 2A showed the sample cluster graph based on the first two principal components. PC1 and PC2 explained 40.3% and 8.6% of the gene expression variation, respectively. DEGs were screened with $|\log_2\text{FoldChange}| \geq 1$ and $\text{FDR} < 0.05$. Finally, a total of 2,896 DEGs were identified in the comparison group of HLMA and LLMA, including 647 up-regulated genes and 2,249 down-regulated genes (Fig 2B). Heatmap demonstrated that the expression patterns of these differential genes were similar within groups while different between groups (Fig 2C).

Functional enrichment analysis of DEGs

To further understand the biological function of DEGs, 2,896 DEGs were enriched by GO and KEGG. The results showed that DEGs were enriched in 268 GO terms, including 203 biological processes (BP), 36 cellular components (CC), and 29 molecular functions (MF). These DEGs were mainly concentrated in cell adhesion (GO:0007155), actin cytoskeleton organization (GO:0030036), and muscle cell differentiation (GO:0042692) (Fig 3A). Furthermore, these DEGs were enriched in 103 KEGG pathways (Fig 3B), among which the pathways related to muscle development included MAPK signaling pathway (bta04010), Wnt signaling pathway (bta04310), and mTOR signaling pathway (bta04150) could be served as important pathways affecting LMA trait.

Construction of weighted gene co-expression network

Firstly, the hclust function was used to perform hierarchical clustering on 21,297 genes in 20 samples to eliminate outlier samples [24]. As shown in Fig 4A, 19 samples were retained for subsequent analysis. Second, the scale-free co-expression gene network was established with the one-step network construction function, and the soft threshold (β) was calculated using the R function “pickSoftThreshold” (Fig 4B). Then, the adjacency matrix was transformed into topological overlap matrix (TOM) and the corresponding dissimilarity (1-TOM) was calculated. Using hierarchical clustering and dynamic tree cutting to detect the modules in the network. Modules with a height value lower than 0.25 were merged, resulting in 19 distinct modules. Correlation analysis between these 19 modules and LMA traits revealed that the sienna3 and lightcyan1 modules were strongly positively correlated with LMA.

Hub genes screening in the sienna3 module and lightcyan1 module

Hub genes in the sienna3 module and lightcyan1 module were defined with GS was greater than 0.2 and MM greater than 0.8. The results showed there were 18 Hub genes in the sienna3 module (Fig 5A) and 105 Hub genes in the lightcyan1 module (Fig 5B), respectively. Then, the top 10 genes of hub gene in lightcyan1 module and the top 10 genes of all genes in sienna3 module were selected according to MCC, MNC and Degree calculated by Cytohubba plug-in in Cytoscape software. At the same time, the intersection of genes obtained by the three algorithms was taken (Fig 5C and 5D). Subsequently, these initially selected genes were further intersected with DEGs and the overlapping genes including *PLP1*, *SNCA*, *MRPL36*, *NDUFS8*, *NDUFB7*, *NDUFA11*, and *MRPL40* were determined for further analyzed (Fig 5E).

Discussion

LMA is the cross-sectional area of the longest dorsal muscle of livestock, which is particularly important in breeding because of its strong correlation with livestock meat production performance. In general, a larger LMA indicates a relatively higher pure meat percentage and better meat quality, while cattle with a larger LMA tend to have fuller and more developed muscles, which may also imply that they grow faster. Previous studies combined GWAS with RNA-seq to

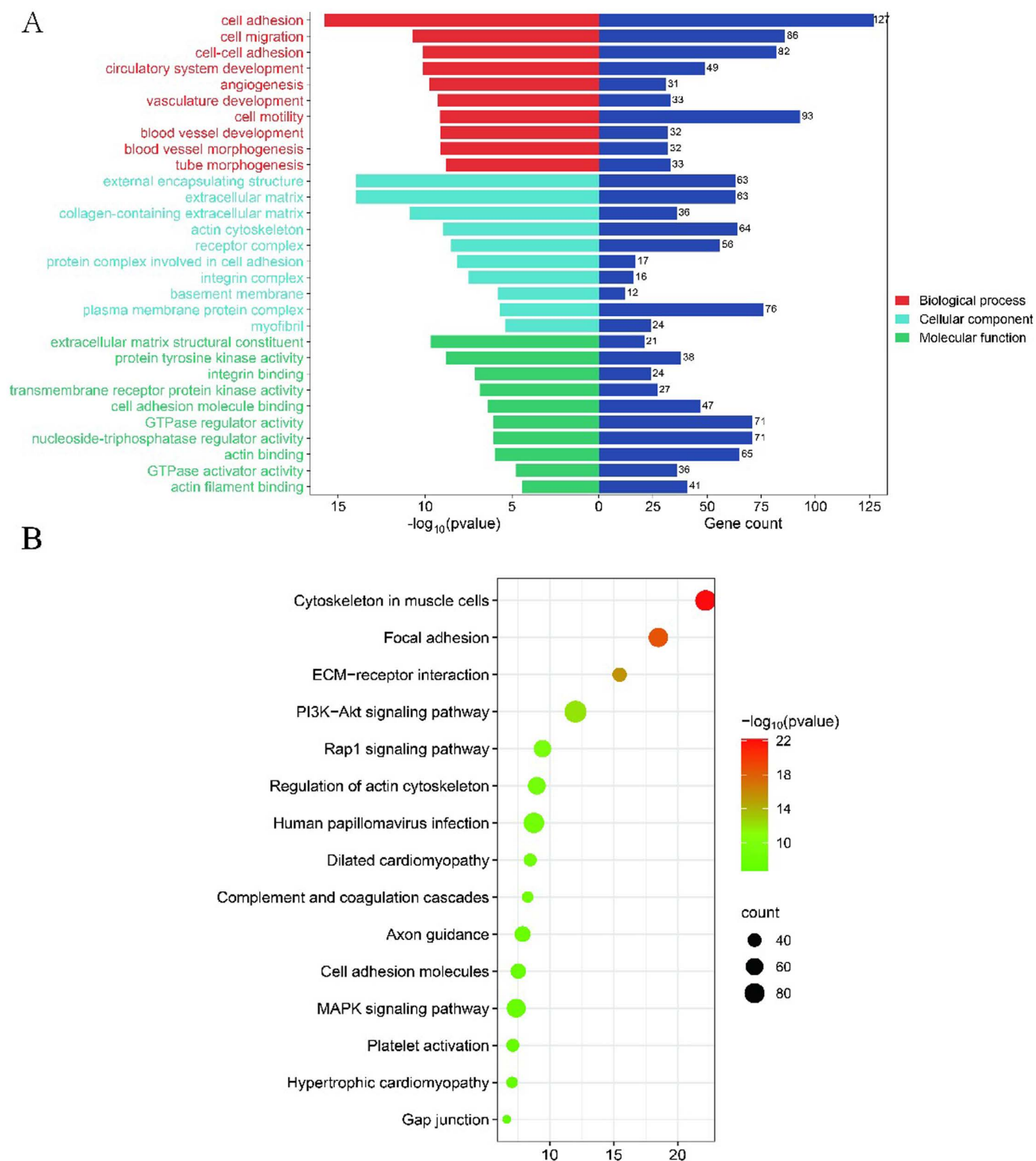


Fig 3. Functional enrichment analysis of DEGs. (A) BP, CC, and MF showed the top ten terms, respectively; (B) The top 15 KEGG pathways enrichment of significant DEGs.

<https://doi.org/10.1371/journal.pone.0322026.g003>

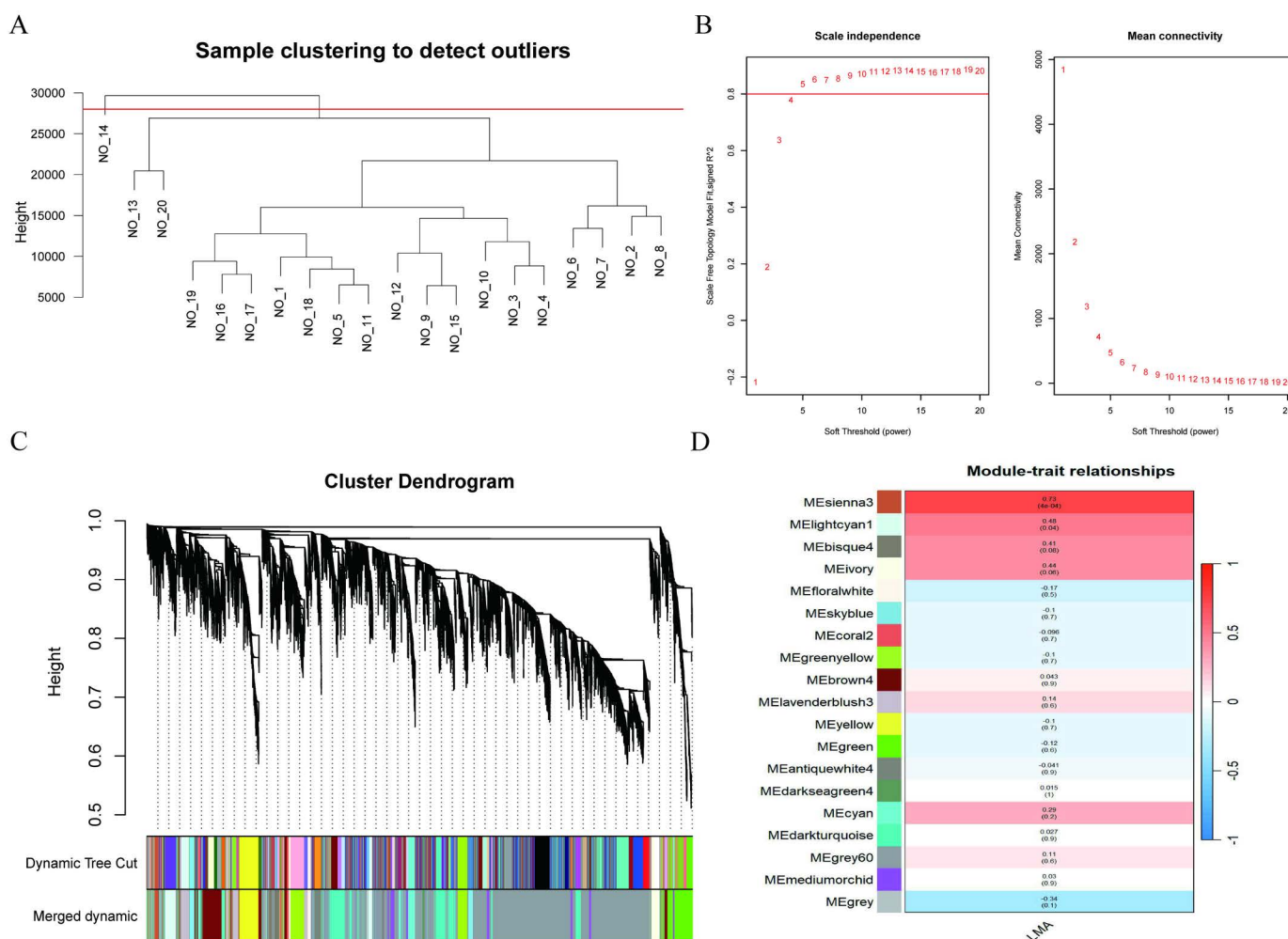


Fig 4. WGCNA analysis. (A) Sample clustering tree. (B) Soft threshold estimation based on adjacency matrix. (C) Dynamic cutting tree algorithm to partition gene modules. (D) Modular-trait heatmap of correlation between gene modules and LMA. Each module contains the corresponding correlation coefficient and p-value.

<https://doi.org/10.1371/journal.pone.0322026.g004>

reveal that *HDAC9* is a candidate gene affecting LMA in Beijing black pigs [29]; Two quantitative trait loci (QTL) related to LMA were identified on SSC1 and SSC2 by multi-species genome-wide association analysis. The candidate gene for the QTL on SSC1 is *FUBP3* (far upstream element-binding protein 3) [30]. Zhao X et al [31] identified one (*DOCK5*), three (*PID1*, *PITX2*, *ELOVL6*), and three (*CCR1*, *PARP14*, *CASR*) genes as candidate genes for LMA, loin muscle depth (LMD), and BFT using GWAS.

In this study, four significant SNPs were screened at the genome-wide level and annotated to be located in or close to three coding genes. The *THBD* gene encodes Thrombomodulin, a protein primarily expressed on the surface of endothelial cells. The *THBD* gene is enriched in GO pathways related to calcium ion binding and blood coagulation. The *THBD* gene played vital roles in the muscle regulatory networks in the Yorkshire breed [32], which is consistent with the results of this study. *CD93*, also known as complement component C1q receptor, plays a key role in cell proliferation and migration [33]. The *CD93* gene exacerbated cell proliferation, angiogenesis, and immune evasion in osteosarcoma by triggering the PI3K/AKT pathway [34]. The *RIMS2* gene encodes the regulatory synaptic membrane exocytosis 2 protein and affects

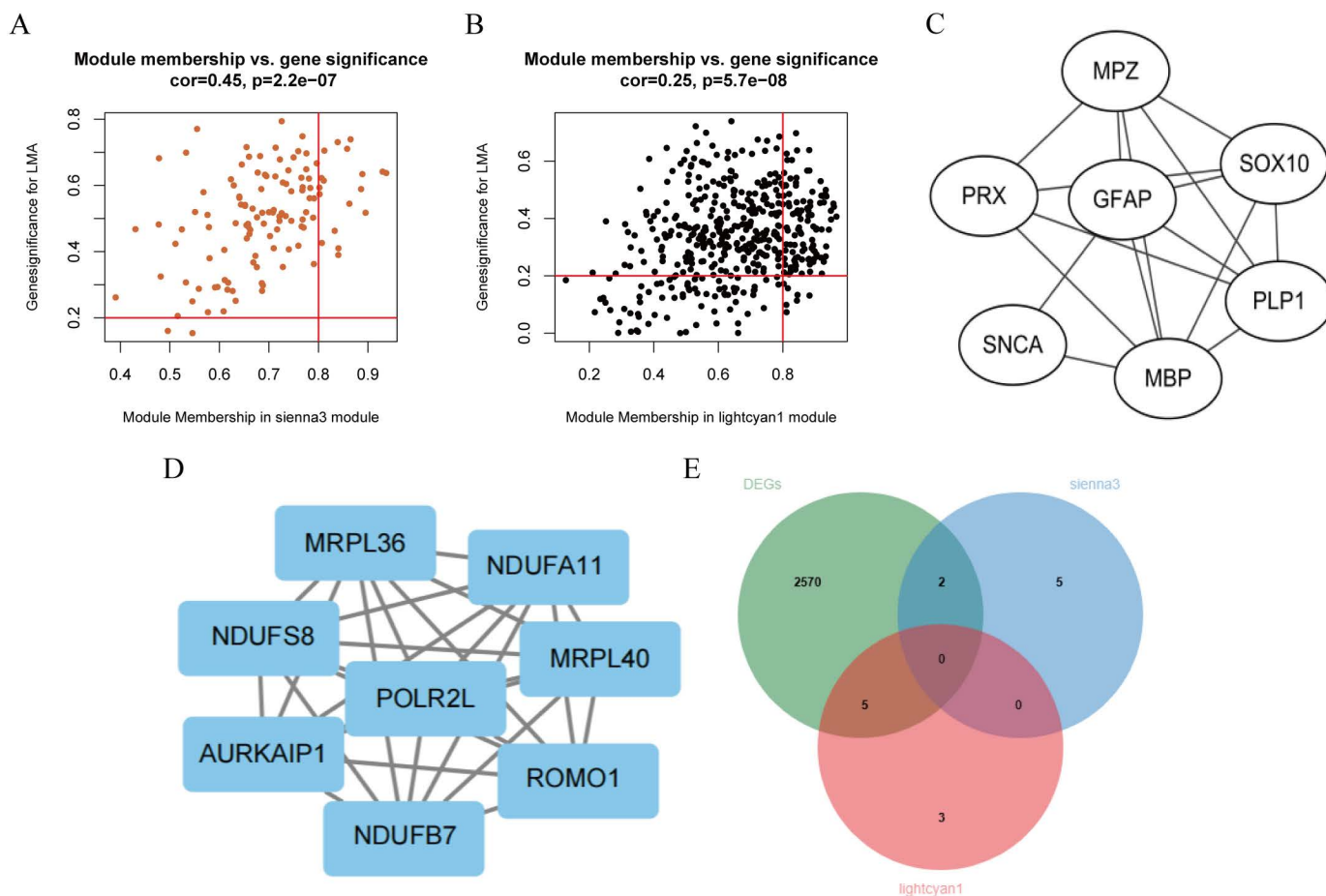


Fig 5. Intersection gene screening. (A) Hub gene within the sienna3 module. (B) Hub gene within the lightcyan1 module. (C) Intersection genes obtained by MCC, MNC, and Degree algorithms in the sienna3 module. (D) Intersection genes obtained by MCC, MNC, and Degree algorithms in the lightcyan1 module. (E) The intersection of genes selected by three algorithms and DEGs.

<https://doi.org/10.1371/journal.pone.0322026.g005>

metabolic pathways related to cell differentiation and proliferation. It is also associated with the regulation of the immune system.

In this study, seven candidate genes were screened at the transcriptome level, with one overlapping gene with the GWAS results. Pathway enrichment analysis showed proteolipid protein I (*PLP1*) was mainly enriched in the signaling pathway associated with mental illness, and was also involved in the long-chain fatty acid biosynthetic process (GO:0042759). *PLP1* was the main component of myelin, and its mutation will lead to progressive neurodegeneration and eventually death due to severe protein defects [35], thus its dysfunction will directly affect the motor function of muscles. *PLP1* was also involved in the regulation of intestinal motility and barrier function [36]. The α -synuclein (*SNCA*) was involved in Parkinson's disease (hsa05012), actin binding (GO:0003779) and other related pathways. *SNCA* may play a role in the compartmentalization of acetylcholine at the neuromuscular junction and in the fine control of skeletal muscle activity [37].

MRPL36 and *MRPL40* both belong to mitochondrial ribosomal proteins, which are important components of the structural and functional integrity of the mitochondrial complex [38]. The overexpression of *MRPL36* seems to improve the efficiency of mitochondrial translation [39]. Ying Liu et al. found that *MRPL40* may be a potential target for the treatment of cryptorchidism and decreased sperm motility and number [40]. Moreover, the downregulation of *MRPL40* expression can

eliminate the effect of interleukin-8 (IL-8) on promoting the proliferation and migration of rectal cancer [41]. Currently, there was no direct evidence to show that *MRPL36* and *MRPL40* were involved in the genetic regulation of muscle development. However, as key components of mitochondrial ribosomal proteins, their contribution to mitochondrial function and energy metabolism may indirectly affect muscle development and maintenance.

NDUFS8, *NDUFB7*, and *NDUFA11* were three genes related to the mitochondrial respiratory chain, and the proteins encoded by these three genes were all important components of the mitochondrial respiratory chain Complex I (NADH). All three genes were involved in biological processes such as citric acid (TCA) cycling respiratory electron transport (R-HSA-1428517) and respiratory electron transport (R-HSA-611105). The expression level of *NDUFS8* in myoblasts decreased significantly with the increase of age, and experimental results showed that the expression of *NDUFS8* in myoblasts promoted differentiation, self-renewal, and apoptosis resistance [42]. Knockout of *NDUFS8* inhibits cell proliferation, migration, and capillary formation in cultured endothelial cells [43]. *NDUFB7* and other genes were involved in regulating lipid synthesis and fatty acid β -oxidation, to maintain normal energy supply under low oxygen conditions [44]. In chickens with woody breast meat, the expression of the *NDUFB7* gene, which was associated with mitochondrial and ATP synthesis, tends to decrease. Additionally, other studies have shown that *NDUFA11* expression was significantly reduced in the blood of patients with acute myocardial infarction, making it a potential biomarker for diagnosis of myocardial infarction [45].

The genes *PPP3R1*, *FAM129B*, and *UBE2G1* are key genes influencing the LMA of uncastrated Nellore males [11]. These genes do not overlap with those identified in this study, which is likely due to differences in breeds.

Conclusions

This study described the genomic variants and candidate genes associated with LMA in Huaxi cattle. Through genome-wide association analysis, the *THBD* gene and the *CD93* gene on BTA 13, and the *RIMS2* gene on BTA 14 were detected to be crucial in affecting the LMA trait. Transcriptome analysis showed that three other candidate genes (*PLP1*, *SNCA*, and *NDUFS8*) were also identified as candidate genes, which regulate LMA by participating in actin synthesis and muscle development. Summarily, the candidate genes related to LMA identified in this study were *THBD*, *CD93*, *RIMS2*, *PLP1*, *SNCA*, and *NDUFS8*. This result could provide new insights into the molecular mechanisms regulating LMA in beef cattle.

Supporting information

S1 File. **S1 Table.** Phenotype statistics of two groups of samples. **S2 Table.** The annotation of significant SNPs (± 100 kb). **S3 Table.** The information of sequencing reads alignment to *Bos taurus* reference genome. **S4 Table.** The significant GO terms and KEGG pathways for differentially expressed genes. **S5 Table.** The enriched GO terms and KEGG pathways for the hub genes in two significant module. **S6 Table.** The detailed information of 6 genes associated with loin muscle area identified by GWAS and RNA-seq strategies. (XLSX)

Acknowledgments

The author would like to thank all members of the Cattle Breeding Innovative Research Team for sample collection and data statistics, as well as the teachers and senior sisters for their careful guidance.

Author contributions

Conceptualization: Yan Chen, Xue Gao, Lingyang Xu, Huijiang Gao.

Data curation: Qingqing Xue, Lili Du, Lupei Zhang, Junya Li.

Investigation: Qingqing Xue, Lili Du, Tianyu Deng, Mang Liang.

Methodology: Yan Chen, Xue Gao, Lingyang Xu, Zezhao Wang, Caihong Zheng, Huijiang Gao.

Software: Keaning Li.

Supervision: Huijiang Gao.

Validation: Keaning Li, Li Qian, Shiyuan Qiu.

Writing – original draft: Qingqing Xue, Lili Du, Tianyu Deng, Mang Liang.

Writing – review & editing: Huijiang Gao.

References

1. Murimi M. What is the Place of Beef in a Healthful Sustainable Homo sapiens Diet? *Am J Health Promot.* 2022;36(5):898–900. <https://doi.org/10.1177/08901171221088661c> PMID: 35531995
2. da Silveira DD, Schmidt PI, Campos GS, de Vargas L, de Souza FRP, Roso VM, et al. Genetic analysis of growth, visual scores, height, and carcass traits in Nelore cattle. *Anim Sci J.* 2021;92(1):e13611. <https://doi.org/10.1111/asj.13611> PMID: 34431165
3. Grigoletto L, Ferraz JBS, Oliveira HR, Eler JP, Bussiman FO, Abreu Silva BC, et al. Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical® Composite Beef Cattle. *Front Genet.* 2020;11:123. <https://doi.org/10.3389/fgene.2020.00123> PMID: 32180796
4. Mamede MMS, Rosa GJM, Eifert EC, Lopes FB, Baldi F, Costa MF, et al. Estimating genetic parameters of reproductive, carcass, and meat quality traits in Polled Nellore cattle. *Trop Anim Health Prod.* 2023;55(2):119. <https://doi.org/10.1007/s11250-023-03541-3> PMID: 36930426
5. Ma J, Gao X, Li J, Gao H, Wang Z, Zhang L, et al. Assessing the Genetic Background and Selection Signatures of Huaxi Cattle Using High-Density SNP Array. *Animals (Basel).* 2021;11(12):3469. <https://doi.org/10.3390/ani11123469> PMID: 34944246
6. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996;273(5281):1516–7. <https://doi.org/10.1126/science.273.5281.1516> PMID: 8801636
7. Arikawa LM, Mota LFM, Schmidt PI, Frezarim GB, Fonseca LFS, Magalhães AFB, et al. Genome-wide scans identify biological and metabolic pathways regulating carcass and meat quality traits in beef cattle. *Meat Sci.* 2024;209:109402. <https://doi.org/10.1016/j.meatsci.2023.109402> PMID: 38056170
8. Medeiros de Oliveira Silva R, Bonvino Stafuzza N, de Oliveira Fragomeni B, Miguel Ferreira de Camargo G, Matos Ceacero T, Noely Dos Santos Gonçalves Cyrillo J, et al. Genome-Wide Association Study for Carcass Traits in an Experimental Nelore Cattle Population. *PLoS One.* 2017;12(1):e0169860. <https://doi.org/10.1371/journal.pone.0169860> PMID: 28118362
9. Shi B, Zhang Z, Lv X, An K, Li L, Xia Z. Screening of Genes Related to Fat Deposition of Pekin Ducks Based on Transcriptome Analysis. *Animals (Basel).* 2024;14(2):268. <https://doi.org/10.3390/ani14020268> PMID: 38254437
10. Du L, Li K, Chang T, An B, Liang M, Deng T, et al. Integrating genomics and transcriptomics to identify candidate genes for subcutaneous fat deposition in beef cattle. *Genomics.* 2022;114(4):110406. <https://doi.org/10.1016/j.ygeno.2022.110406> PMID: 35709924
11. Santos Silva DBD, Fonseca LFS, Magalhães AFB, Muniz MMM, Baldi F, Ferro JA, et al. Transcriptome profiling of muscle in Nelore cattle phenotypically divergent for the ribeye muscle area. *Genomics.* 2020;112(2):1257–63. <https://doi.org/10.1016/j.ygeno.2019.07.012> PMID: 31351181
12. Wang H, Wang X, Li M, Wang S, Chen Q, Lu S. Identification of key sex-specific pathways and genes in the subcutaneous adipose tissue from pigs using WGCNA method. *BMC Genom Data.* 2022;23(1):35. <https://doi.org/10.1186/s12863-022-01054-w> PMID: 35538407
13. Yang M, Zhang R, Liu X, Shi G, Liu H, Wang L, et al. Integrating genome-wide association study with RNA-seq revealed DBI as a good candidate gene for intramuscular fat content in Beijing black pigs. *Anim Genet.* 2023;54(1):24–34. <https://doi.org/10.1111/age.13270> PMID: 36305366
14. Piórkowska K, Żukowski K, Ropka-Molik K, Tyra M. Variations in Fibrinogen-like 1 (FGL1) Gene Locus as a Genetic Marker Related to Fat Deposition Based on Pig Model and Liver RNA-Seq Data. *Genes (Basel).* 2022;13(8):1419. <https://doi.org/10.3390/genes13081419> PMID: 36011329
15. Xia J, Qi X, Wu Y, Zhu B, Xu L, Zhang L, et al. Genome-wide association study identifies loci and candidate genes for meat quality traits in Simmental beef cattle. *Mamm Genome.* 2016;27(5–6):246–55. <https://doi.org/10.1007/s00335-016-9635-x> PMID: 27126640
16. Luan M, Ruan D, Qiu Y, Ye Y, Zhou S, Yang J, et al. Genome-wide association study for loin muscle area of commercial crossbred pigs. *Anim Biosci.* 2023;36(6):861–8. <https://doi.org/10.5713/ab.22.0407> PMID: 36634654
17. Niu Q, Zhang T, Xu L, Wang T, Wang Z, Zhu B, et al. Integration of selection signatures and multi-trait GWAS reveals polygenic genetic architecture of carcass traits in beef cattle. *Genomics.* 2021;113(5):3325–36. <https://doi.org/10.1016/j.ygeno.2021.07.025> PMID: 34314829
18. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res.* 2018;27(2):e1608. <https://doi.org/10.1002/mpr.1608> PMID: 29484742
19. Zepeda-Batista JL, Núñez-Domínguez R, Ramírez-Valverde R, Jahuey-Martínez FJ, Herrera-Ojeda JB, Parra-Bracamonte GM. Discovering of Genomic Variations Associated to Growth Traits by GWAS in Braunvieh Cattle. *Genes (Basel).* 2021;12(11):1666. <https://doi.org/10.3390/genes12111666> PMID: 34828272
20. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics Proteomics Bioinformatics.* 2021;19(4):619–28. <https://doi.org/10.1016/j.gpb.2020.10.007> PMID: 33662620

21. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 1999;27(1):29–34. <https://doi.org/10.1093/nar/27.1.29> PMID: [9847135](#)
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25–9. <https://doi.org/10.1038/75556> PMID: [10802651](#)
23. Hassan SU, Chua EG, Kaur P, Paz EA, Tay CY, Greeff JC, et al. Contribution of the Immune Response in the Ileum to the Development of Diarrhoea caused by Helminth Infection: Studies with the Sheep Model. *Funct Integr Genomics.* 2022;22(5):865–77. <https://doi.org/10.1007/s10142-022-00864-6> PMID: [35576023](#)
24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559. <https://doi.org/10.1186/1471-2105-9-559> PMID: [19114008](#)
25. Yang X, Huang Y, Tang D, Yue L. Identification of key genes associated with acute myocardial infarction using WGCNA and two-sample mendelian randomization study. *PLoS One.* 2024;19(7):e0305532. <https://doi.org/10.1371/journal.pone.0305532> PMID: [39024234](#)
26. Zhu H, Lu H, Li T, Chen J. Identification of the differentially expressed activated memory CD4+ T-cells-related genes and ceRNAs in oral lichen planus. *Heliyon.* 2024;10(12):e33305. <https://doi.org/10.1016/j.heliyon.2024.e33305> PMID: [39022110](#)
27. Wang J, Liu Z, Cao D, Liu J, Li F, Han H, et al. Elucidation of the genetic determination of clutch traits in Chinese local chickens of the Laiwu Black breed. *BMC Genomics.* 2023;24(1):686. <https://doi.org/10.1186/s12864-023-09798-0> PMID: [37968610](#)
28. Liu D, Li X, Wang L, Pei Q, Zhao J, Sun D, et al. Genome-wide association studies of body size traits in Tibetan sheep. *BMC Genomics.* 2024;25(1):739. <https://doi.org/10.1186/s12864-024-10633-3> PMID: [39080522](#)
29. Hou R, Chen L, Liu X, Liu H, Shi G, Hou X, et al. Integrating Genome-Wide Association Study with RNA-Sequencing Reveals HDAC9 as a Candidate Gene Influencing Loin Muscle Area in Beijing Black Pigs. *Biology (Basel).* 2022;11(11):1635. <https://doi.org/10.3390/biology11111635> PMID: [36358336](#)
30. He Y, Ma J, Zhang F, Hou L, Chen H, Guo Y, et al. Multi-breed genome-wide association study reveals heterogeneous loci associated with loin eye area in pigs. *J Appl Genet.* 2016;57(4):511–8. <https://doi.org/10.1007/s13353-016-0351-8> PMID: [27183999](#)
31. Zhao X, Qiu Y, Meng F, Zhuang Z, Ruan D, Wu J, et al. Genome-wide association studies for loin muscle area, loin muscle depth and backfat thickness in DLY pigs. *Anim Genet.* 2024;55(1):134–9. <https://doi.org/10.1111/age.13386> PMID: [38098441](#)
32. Zhao Y, Li J, Liu H, Xi Y, Xue M, Liu W, et al. Dynamic transcriptome profiles of skeletal muscle tissue across 11 developmental stages for both Tongcheng and Yorkshire pigs. *BMC Genomics.* 2015;16(1):377. <https://doi.org/10.1186/s12864-015-1580-7> PMID: [25962502](#)
33. Fantone S, Tossetta G, Di Simone N, Tersigni C, Scambia G, Marcheggiani F, et al. CD93 a potential player in cytotrophoblast and endothelial cell migration. *Cell Tissue Res.* 2022;387(1):123–30. <https://doi.org/10.1007/s00441-021-03543-3> PMID: [34674045](#)
34. Zhang Y, Liu Y, Ma Y, Xu Y, Wang G, Han X. CD93 aggravates cell proliferation, angiogenesis and immune escape in osteosarcoma through triggering the PI3K/AKT pathway. *J Orthop Sci.* 2024:S0949-2658(24)00204-5. <https://doi.org/10.1016/j.jos.2024.10.011> PMID: [39542800](#)
35. Elitt MS, Tesar PJ. Pelizaeus-Merzbacher disease: on the cusp of myelin medicine. *Trends Mol Med.* 2024;30(5):459–70. <https://doi.org/10.1016/j.molmed.2024.03.005> PMID: [38582621](#)
36. Woods C, Flockton AR, Wallace LE, Keenan CM, Macklin WB, Sharkey KA, et al. Proteolipid protein 1 is involved in the regulation of intestinal motility and barrier function in the mouse. *Am J Physiol Gastrointest Liver Physiol.* 2023;324(2):G115–30. <https://doi.org/10.1152/ajpgi.00171.2022> PMID: [36511517](#)
37. Pelkonen A, Yavich L. Neuromuscular pathology in mice lacking alpha-synuclein. *Neurosci Lett.* 2011;487(3):350–3. <https://doi.org/10.1016/j.neulet.2010.10.054> PMID: [21029764](#)
38. Cheong A, Lingutla R, Mager J. Expression analysis of mammalian mitochondrial ribosomal protein genes. *Gene Expr Patterns.* 2020;38:119147. <https://doi.org/10.1016/j.gep.2020.119147> PMID: [32987154](#)
39. Prestele M, Vogel F, Reichert AS, Herrmann JM, Ott M. Mrpl36 is important for generation of assembly competent proteins during mitochondrial translation. *Mol Biol Cell.* 2009;20(10):2615–25. <https://doi.org/10.1091/mbc.e08-12-1162> PMID: [19339279](#)
40. Liu Y, Fu L-L, Xu H-Z, Zheng Y-M, Li W-X, Qian G-H, et al. Insufficiency of Mrpl40 disrupts testicular structure and semen parameters in a murine model. *Asian J Androl.* 2023;25(5):627–31. <https://doi.org/10.4103/aja2022119> PMID: [36891938](#)
41. Du F, Zhang Y, Ji X, Zhang N, Xu J, Ning T, et al. IL-8-mediated overexpression of ZNF274 promotes the proliferation and migration of colorectal cancer cells through the transactivation of MRPL40. *Heliyon.* 2023;9(8):e19046. <https://doi.org/10.1016/j.heliyon.2023.e19046> PMID: [37636370](#)
42. Motohashi N, Minegishi K, Aoki Y. Inherited myogenic abilities in muscle precursor cells defined by the mitochondrial complex I-encoding protein. *Cell Death Dis.* 2023;14(10):689. <https://doi.org/10.1038/s41419-023-06192-2> PMID: [37857600](#)
43. Xiong Q-W, Jiang K, Shen X-W, Ma Z-R, Yan X-M, Xia H, et al. The requirement of the mitochondrial protein NDUFS8 for angiogenesis. *Cell Death Dis.* 2024;15(4):253. <https://doi.org/10.1038/s41419-024-06636-3> PMID: [38594244](#)
44. Li M, Tian X, Li X, Huang M, Huang S, Wu Y, et al. Diverse energy metabolism patterns in females in *Neodon fuscus*, *Lasiopodomys brandtii*, and *Mus musculus* revealed by comparative transcriptomics under hypoxic conditions. *Sci Total Environ.* 2021;783:147130. <https://doi.org/10.1016/j.scitotenv.2021.147130> PMID: [34088150](#)
45. Qiu J, Gu Y. Analysis of the prognostic value of mitochondria-related genes in patients with acute myocardial infarction. *BMC Cardiovasc Disord.* 2024;24(1):408. <https://doi.org/10.1186/s12872-024-04051-2> PMID: [39103773](#)