

Article

Not peer-reviewed version

Genome-Wide Association Study of Body Size Traits in Yanqi Horses

[Weijun Sun](#)[†], [Zhehong Shen](#)[†], Daoerji Cairen, Penghui Luo, [Xinkui Yao](#), [Jun Meng](#), [Yaqi Zeng](#)^{*}

Posted Date: 14 April 2026

doi: 10.20944/preprints202604.0945.v1

Keywords: body size traits; candidate genes; genome-wide association study; single-nucleotide polymorphism; Yanqi horse



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Genome-Wide Association Study of Body Size Traits in Yanqi Horses

Weijun Sun ^{1†}, Zhehong Shen ^{1†}, Daoerji Cairen ², Penghui Luo ³, Xinkui Yao ^{1,4}, Jun Meng ^{1,4} and Yaqi Zeng ^{1,4,*}

¹ College of Animal Science, Xinjiang Agricultural University, Urumqi 830052, China

² Bazhou Animal Husbandry Workstation, Urumqi 841000, China

³ Animal Husbandry Station of Xinjiang Uygur Autonomous Region, Urumqi 830004, China

⁴ Xinjiang Key Laboratory of Horse Breeding and Exercise Physiology, Urumqi, 830052, China

* Correspondence: xjauzengyaqi@163.com

† These authors contributed equally to this work.

Simple Summary

This study conducted a GWAS on four body size traits (withers height, body length, heart girth, cannon bone circumference) in 183 Yanqi horses to identify associated molecular markers and candidate genes, aiming to accelerate the breed's selective breeding. 185 significant SNPs and 359 candidate genes were identified, with GABRB1, FIGN, GABRA4, ENSECAG00000051747 and COX7B2 core to growth. GO/KEGG analyses showed these genes enrich in muscle cytoskeleton, actin regulation and neuroactive ligand-receptor pathways. The study uncovers novel body size-related genomic markers for Yanqi horses, providing valuable insights for functional gene research and molecular-assisted breeding, which boosts the genetic improvement and sustainable development of this indigenous Xinjiang horse breed.

Abstract

Horse body size serves as a direct production indicator that reflects growth status, with body size traits frequently utilized as principal selection criteria and extensively applied in monitoring animal growth and development as well as evaluating the efficacy of genetic selection. The identification of molecular markers associated with body size traits has the potential to expedite animal breeding programs. The Yanqi horse, an important indigenous breed in Xinjiang, is primarily distributed in the Bayingolin Mongol Autonomous Prefecture. However, molecular markers linked to body size in Yanqi horses remain uncharacterized. In the present study, a genome-wide association study was performed on withers height, body length, heart girth, and cannon bone circumference traits in 183 Yanqi horses to identify genomic variants associated with body size characteristics. A total of 185 single-nucleotide polymorphisms were significantly associated with body size traits, and 359 candidate genes were annotated within 200 kb upstream and downstream of the significant loci. Among these, five genes, GABRB1, FIGN, GABRA4, ENSECAG00000051747, and COX7B2, may be implicated in the growth and development of Yanqi horses. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analyses indicated that these genes are primarily involved in cytoskeletal structures within muscle cells, regulation of the actin cytoskeleton, and neuroactive ligand-receptor interaction pathways. In summary, this study presents novel markers and candidate gene sets associated with body size traits in Yanqi horses, offering valuable insights for functional gene investigations and presenting substantial potential for accelerating the breeding of Yanqi horses.

Keywords: body size traits; candidate genes; genome-wide association study; single-nucleotide polymorphism; Yanqi horse

1. Introduction

Over the past decade, numerous genome-wide association studies (GWAS) have been performed in livestock species, including investigations of various traits in Simmental [1], black goats [2], Cervus nippon [3], Holstein cattle [4], Yunong black pigs [5], Tibetan northwest white cashmere goats [6], Donghu crossbred sheep [7], Dezhou donkeys [8], among others. Whole-genome sequencing has been employed to characterize population structures and to detect polymorphisms potentially associated with economically significant traits in both livestock and poultry. However, genome-wide variation in horses has remained comparatively underexplored.

Therefore, this study was designed to identify single-nucleotide polymorphisms (SNPs) [9] significantly associated with four body size traits [withers height (WH), body length (BL), heart girth (HG), and cannon bone circumference (CBC)] in Yanqi horses through GWAS [10] methodologies. Simultaneously, candidate genes and biological pathways potentially influencing these traits were screened based on the corresponding SNP loci. These findings provide molecular markers valuable for the selective breeding of Yanqi horses.

2. Materials and Methods

2.1. Experimental Animals and Tissue Sample Collection

A total of 183 healthy Yanqi horses were randomly selected as experimental subjects, comprising 36 stallions and 147 mares, aged between 2 months (0 years) and 13 years. All experimental animals originated from a single population maintained at the Baoqi Yanqi Horse Conservation Farm in Bayinbuluke Town, Hejing County, Bayingolin Mongol Autonomous Prefecture. Blood samples (10 ml) were drawn from the jugular vein using disposable vacuum blood collection tubes containing ethylenediaminetetraacetic acid as an anticoagulant. After collection, the samples were immediately mixed thoroughly with the anticoagulant, aliquoted, and stored in liquid nitrogen tanks prior to transport to the laboratory for DNA extraction.

2.2. Phenotype Data Sources and Measurement Methods

Body size traits, including WH, BL, HG, and CBC, were measured on the same group of Yanqi horses by a single individual using a measuring tape and measuring stick. Each trait was measured three times, and the average value was calculated. Descriptive statistical analyses of the phenotypic data were performed using IBM SPSS Statistics 27 software.

Genomic DNA was extracted via the magnetic bead method utilizing the CWE9600 Magbead Blood DNA Kit (Jiangsu Cowin Biotech Co., Ltd.). The procedure involved adding an appropriate quantity of tissue sample to a 2.0 ml lysis tube, followed by the addition of Proteinase K and Buffer WL. The resulting mixture was vortexed for 30 s and incubated at 56 °C in a constant-temperature water bath for 40 min, with intermittent vortexing every 20 min. Upon completion of lysis, the sample was subjected to brief centrifugation. Subsequently, lysate, Buffer KL, and Magbeads PN solution were added. The Spintips Pack (magnetic rod sleeve) was inserted into a 96 DW deep-well plate, and the DNA extraction program was initiated to carry out gDNA extraction.

2.3. Quality Control of Genomic DNA

Genomic DNA was isolated from the collected blood samples using the standard phenol-chloroform extraction method. The concentration of the extracted DNA was quantified using a NanoDrop2000 ultra-micro spectrophotometer, while its integrity was evaluated through agarose gel electrophoresis. All DNA samples were preserved at -80 °C.

2.4. Whole Genome Resequencing

Qualified genomic DNA samples were transported on dry ice to Tianjin Kangpusen Agricultural Technology Co., Ltd. for whole-genome sequencing. A library with an insert size of 350 bp was

prepared, and paired-end sequencing was performed with a read length of 150 bp using the DNBSEQ7 sequencing platform.

2.5. Quality Control

To ensure the reliability of downstream information analysis, fastp software (<https://github.com/OpenGene/fastp>) was employed to conduct rigorous quality control processing on the raw data files (FASTQ format) generated by paired-end sequencing. The parameters `fastp -i -I -o -O -w 4 -q 20 -n 2 -u 30` were configured to remove low-quality reads and adapter sequences, thereby producing clean reads and providing a foundation for subsequent data analysis. The criteria applied for data quality control were as follows:

- (1) Reads containing adapter sequences were discarded;
- (2) Paired reads were removed if the proportion of ambiguous bases (N) in a read exceeded 2% of the total base count;
- (3) Paired reads were discarded if the proportion of low-quality bases ($Q \leq 20$) in a read exceeded 30% of the total bases.

2.6. Alignment to Reference Genome

After quality control of sequencing data, the BWA MEM algorithm in BWA software (<https://github.com/lh3/bwa>) was employed to align the high-quality clean reads (FASTQ format) to the horse reference genome (EquCab 3.0). The alignment parameters were configured as `-t 4 -M -R '@RG\tID:$i\tLB:$i\tPL\tSM:$i'`, and BAM format files containing the alignment results were generated. Subsequently, the SortSam and MarkDuplicates modules of PICARD software (<https://broadinstitute.github.io/picard/>) were utilized to sort the unsorted BAM files by genomic coordinates by specifying the `SORT_ORDER=coordinate` parameter, while the `REMOVE_DUPLICATES=true` parameter was applied to eliminate duplicate reads, thereby yielding high-quality reads (BAM format). The `bamqc` module of QualiMap was then used to perform quality assessment on the processed BAM files. Finally, the BaseRecalibrator and ApplyBQSR modules of Genome Analysis Toolkit (GATK) were executed to carry out base quality score recalibration, thereby enhancing the accuracy of sequencing data and ensuring reliable variant calling for subsequent genetic analyses.

2.7. SNP Detection

Raw SNPs were identified and genotyped using the HaplotypeCaller, GenotypeGVCFs, and SelectVariants modules of GATK. Initially, variant information was called from BAM files through the HaplotypeCaller module, producing GVCF (Genome VCF) format files to store the variant data. To ensure the accuracy of variant calling, the minimum mapping quality threshold was set to `--minimum-mapping-quality 30`. Subsequently, the GenomicsDBImport tool in GATK was executed to import multiple GVCF files into a GenomicsDB database, thereby enabling centralized storage and facilitating subsequent processing via the GenotypeGVCFs module. Parameters were configured as `--intervals 1 --reader-threads 1 --batch-size 50` to enhance the efficiency of joint genotyping. Genotypes were then generated from the GVCF files using the GenotypeGVCFs module with EquCab 3.0 as the reference genome, and the parameter `max-alternate-alleles 1` was specified to limit the number of alternate alleles to one. Finally, the SelectVariants module was employed to extract and filter variant sites from the VCF files, resulting in raw VCF files (SNPs). Quality control was conducted using the VariantFiltration module, in accordance with GATK-recommended hard filtering criteria, ultimately yielding clean SNPs.

2.8. Variant Annotation

Functional annotation of clean SNPs was carried out using the `table_annoar.pl` module of ANNOVAR software, based on the horse reference genome annotation file

(https://ftp.ensembl.org/pub/release-113/gtf/equus_caballus/Equus_caballus.EquCab3.0.113.gtf.gz). During the annotation process, the parameter `-buildver UCD1.2UseName` was specified to define the genome build version, while `-protocol refGene` and `-operation g` were selected to designate the annotation protocol. This approach enabled the precise determination of whether variants were located within genes and identified the specific genes in which the variants were present.

2.9. Quality Control

Raw data were subjected to effective loci screening using PLINK software [11], based on the following filtering criteria: minor allele frequency (MAF) less than 5%, individual call rate below 90%, and SNP missing rate under 90%. The parameters were configured as “`--maf 0.05 --mind 0.1 --geno 0.1 --chr-set 31`”.

2.10. GWAS

GWAS analysis was performed using the mixed linear model (MLM) [12] implemented in GEMMA software [13], which integrates multiple association analysis algorithms to enhance the detection of reliable association loci through cross-validation across methods. The MLM equation is as follows:

$$Y = \text{SNP} + \text{PCs} + \text{Kinship} + e$$

Where Y denotes the phenotypic vector, SNP represents the fixed effect vector, PCs indicate the principal components employed to correct for population structure, Kinship denotes the kinship matrix, and e refers to the random residual effect vector. The K matrix represents kinship relationships derived from SNP markers.

To minimize the inflated false positive rate introduced by multiple testing, a multiple testing correction was applied to the GWAS results. A Bonferroni-adjusted significance threshold of $1E-7$ and a suggestive threshold of $1E-6$ were adopted. Finally, Manhattan and quantile–quantile (Q–Q) plots were generated using the CMplot package in R for visualization of GWAS outcomes.

2.11. Candidate Gene and Functional Enrichment Analysis

Based on the EquCab 3.0 horse reference genome in Ensembl, the 200 kb regions upstream and downstream of significant loci were annotated using ANNOVAR software [14]. The resulting candidate genes were annotated and subsequently subjected to Gene Ontology (GO) enrichment analysis via the DAVID platform (<https://david.ncifcrf.gov/summary.jsp>) [15].

3. Results

3.1. Descriptive Statistics of Body Size Traits in Yanqi Horses

The body size traits assessed in this study included WH, BL, HG, and CBC. Descriptive statistical analysis was conducted on the phenotypic data following dataset organization (Table 1). The coefficients of variation (CV%) for the body size traits of 183 Yanqi horses were all below 1.5%, suggesting a high degree of uniformity and minimal phenotypic variation within the population. Among these traits, HG exhibited slightly greater variation (CV% = 1.13%), reflecting individual differences in muscle development.

Table 1. Descriptive statistics of body size traits.

Traits	Number	Mean	SD	Max	Min	CV%
Withers height (cm)	183	132.07	0.96	152.00	98.00	0.73
Body length (cm)	183	131.28	1.30	160.00	90.00	0.99
Heart girth (cm)	183	150.01	1.70	180.00	90.00	1.13

Cannon bone circumference (cm)	183	16.64	0.13	20.00	12.00	0.78
-----------------------------------	-----	-------	------	-------	-------	------

3.2. Genomic data statistics

A total of 35,721,465,026 clean reads were generated in this study, with an average mapping rate of 99.84% and an average sequencing depth of 11.39 \times . Detailed information is presented in Supplementary Table 1. Through sequencing, a total of 27,310,794 SNPs were identified, with an overall genotyping rate of 99.92%, indicating exceptionally high coverage. Following quality control, 13,366,672 SNPs were retained and found to be evenly distributed across the 31 autosomes of Yanqi horses, rendering them suitable for subsequent GWAS analysis (Figure 1). SNPs located within gene regions were found in intronic and exonic regions, accounting for 5,616,877 and 122,236 SNPs, respectively, corresponding to 42.03% and 0.91% of the total SNPs. Within the exonic regions, 59,243 non-synonymous and 57,597 synonymous mutations were identified (Figure 2).

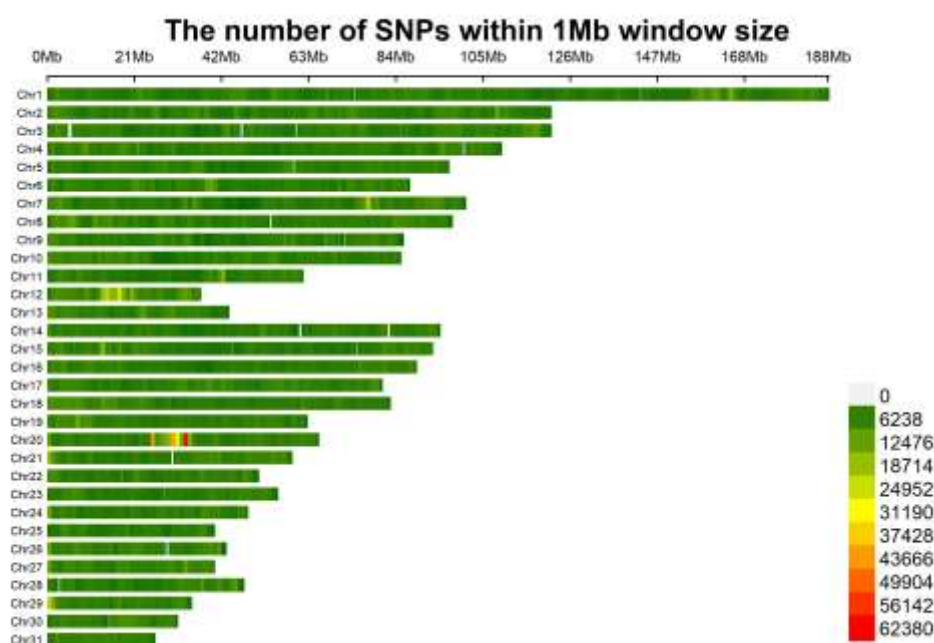


Figure 1. Distribution of SNP loci on autosomes in Yanqi horses.

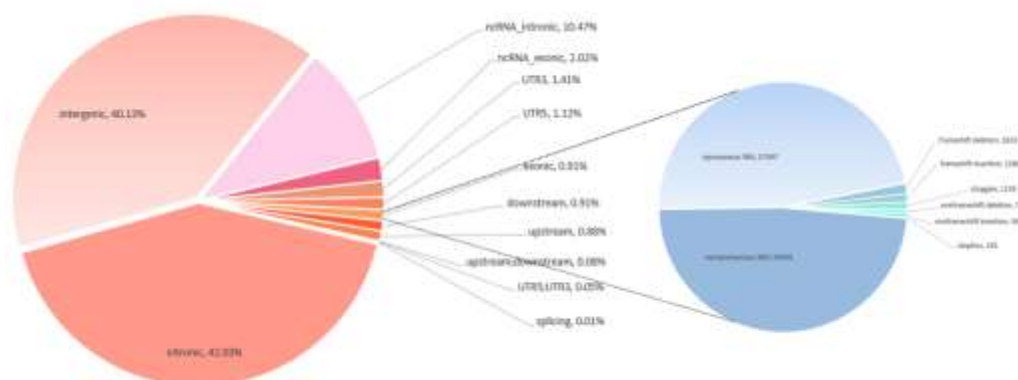


Figure 2. Functional annotation of variant loci in Yanqi horses.

3.3. GWAS of WH Traits in Yanqi Horses

The GWAS results indicated that 45 SNPs fell within the suggestive significance threshold ($P < 1 \times 10^{-6}$), co-localized with 79 candidate genes. Among these, six SNP loci exhibited significant associations ($P < 1 \times 10^{-7}$), with notable peaks observed on chromosomes Chr1, Chr3, Chr7, Chr14, Chr18, and Chr21 (Figure 3). These 45 significant SNPs were further analyzed and annotated using ANNOVAR software within a 200 kb window, both upstream and downstream of the SNP loci significantly associated with WH. The most significant SNP was located at position 128,108.446 kb on chromosome 1, with a P-value of $1.340744E-09$, and was annotated to seven genes: HACD3, DENND4A, DPP8, IGDC3, IGDC4, INTS14, and SLC24A1. Additionally, another significantly associated SNP on the same chromosome was mapped to the VPS13C gene. Annotations for the remaining significant loci are provided in Supplementary Table 2.

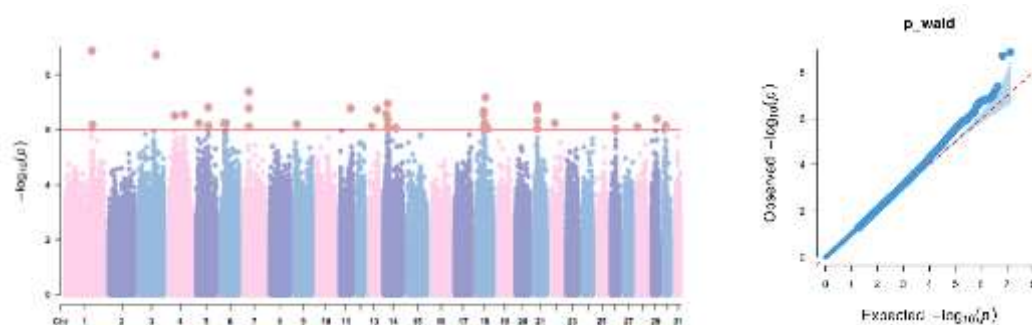


Figure 3. Manhattan plot and quantile-quantile plot for the withers height trait in Yanqi horses.

3.4. GWAS of BL Traits in Yanqi Horses

The GWAS results identified a total of 25 SNPs within the suggestive significance threshold ($P < 1 \times 10^{-6}$), co-localized with 79 candidate genes. Among these, six SNP loci exhibited significant associations ($P < 1 \times 10^{-7}$), with prominent signals detected on chromosomes Chr1, Chr3, Chr18, and Chr30 (Figure 4). For comparative analysis and gene annotation, the 25 significantly associated SNPs were examined using ANNOVAR software within a 200 kb window upstream and downstream of the BL-associated loci. The most significant SNP was located at position 128,108.446 kb on chromosome 1, with a P-value of $3.50987E-08$, and was annotated to seven genes: HACD3, DENND4A, DPP8, IGDC3, IGDC4, INTS14, and SLC24A1. Annotations for the remaining significant loci are listed in Supplementary Table 3.

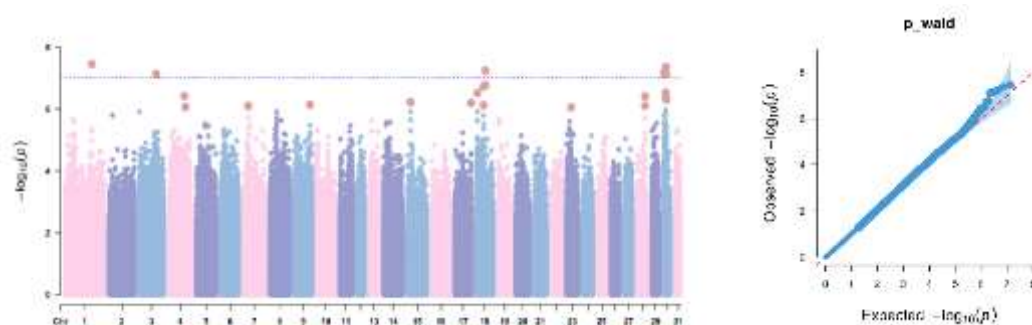


Figure 4. Manhattan plot and quantile-quantile plot for body length trait in Yanqi horses.

3.5. GWAS of HG Traits in Yanqi Horses

The GWAS results identified a total of 51 SNPs within the suggestive significance threshold ($P < 1 \times 10^{-6}$), co-localized with 109 candidate genes. Among these, 13 SNP loci exhibited significant associations ($P < 1 \times 10^{-7}$), with elevated signals observed on chromosomes Chr1, Chr3, Chr9, Chr10, Chr11, Chr13, Chr17, Chr18, Chr28, Chr29, and Chr30 (Figure 5). These 51 significant SNPs were

subjected to comparative analysis and gene annotation using ANNOVAR software, based on a 200 kb region upstream and downstream of each HG-associated SNP locus. The most significant locus was located at position 23,147.732 kb on chromosome 29, with a P-value of 7.81193E-09, and was annotated to the calcium/calmodulin-dependent protein kinase ID gene, CAMK1D. Annotations for the remaining significant loci are provided in Supplementary Table 4.

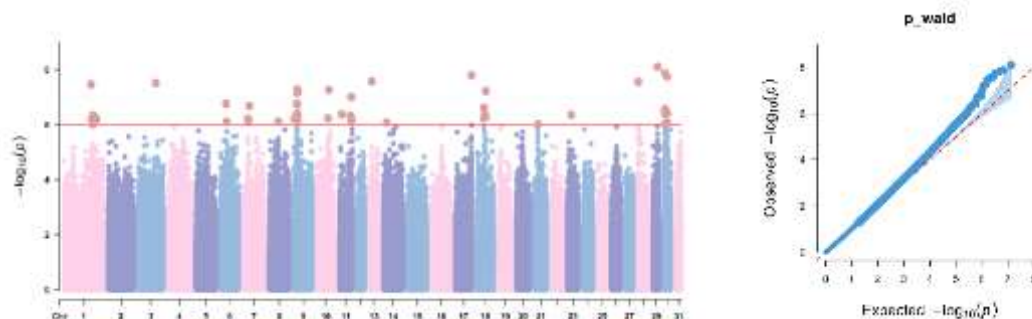


Figure 5. Manhattan plot and quantile-quantile plot for heart girth trait in Yanqi horses.

3.6. GWAS of CBC Traits in Yanqi Horses

The GWAS results identified a total of 64 SNPs within the suggestive significance threshold ($P < 1 \times 10^{-6}$), co-localized with 92 candidate genes. Among these, seven SNP loci exhibited significant associations ($P < 1 \times 10^{-7}$), with prominent signals observed on chromosomes Chr3, Chr6, Chr14, and Chr18 (Figure 6). These 64 significant SNPs were subjected to comparative analysis and gene annotation using ANNOVAR software, based on a 200 kb region both upstream and downstream of each CBC-associated SNP locus. The most significant locus was located at position 28,069.889 kb on chromosome 6, with a P-value of 8.53322E-09, and was annotated to four genes: A0A5F5PYI1_HORSE, PEX26, TUBA8, and USP18. Additionally, seven other significantly associated SNP loci mapped to the same chromosome were also annotated to the same four genes. Annotations for the remaining significant loci are listed in Supplementary Table 5.

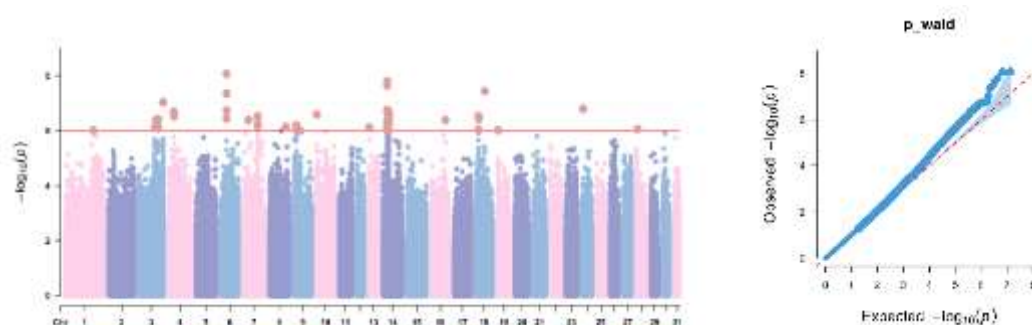


Figure 6. Manhattan plot and quantile-quantile plot for cannon bone circumference trait in Yanqi horses.

3.7. Functional Annotation and Enrichment Analysis of Candidate Genes for WH Traits in Yanqi Horses

To further investigate the biological roles of candidate genes significantly associated with WH traits in Yanqi horses, GO functional annotation was conducted. The annotation framework was categorized into three domains: biological process, cellular component, and molecular function. As illustrated in Figure 8 and ranked by P-value, the enriched functional categories within each domain were as follows: (1) Biological process: muscle contraction, chloride transmembrane transport, actin-mediated cell contraction, etc. (2) Cellular component: myosin filament, myofibril, myosin II complex, etc. (3) Molecular function: microfilament motor activity, cytoskeletal motor activity, chloride channel activity, etc.

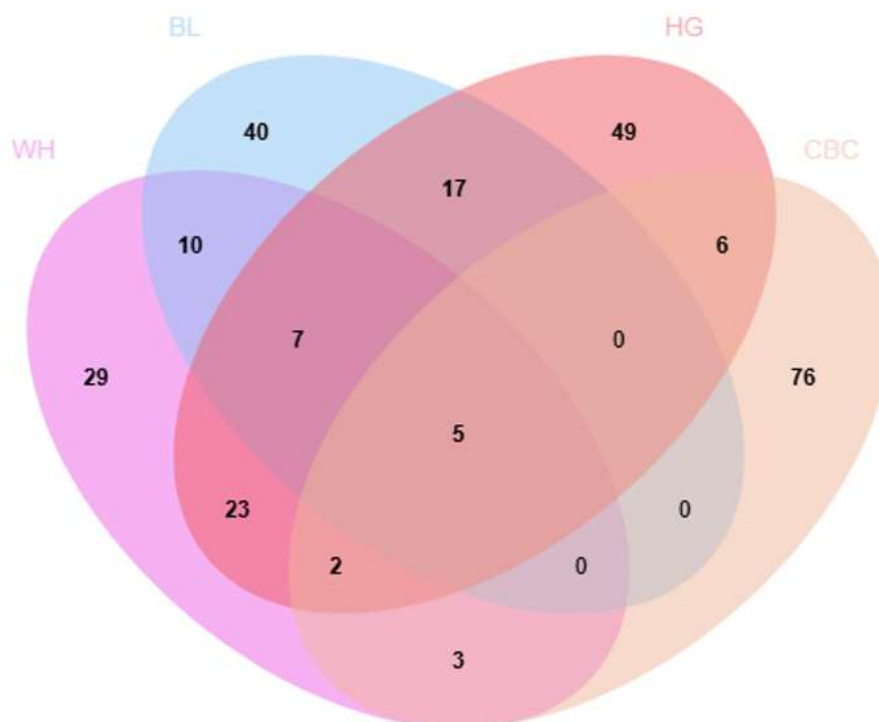


Figure 7. Venn diagram of candidate genes in Yanqi horses.

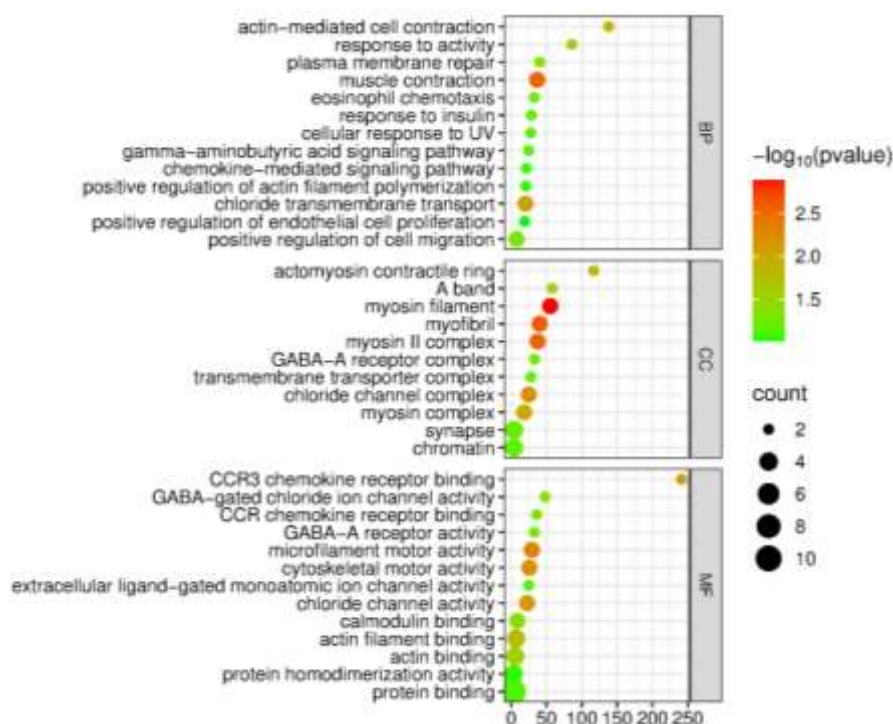


Figure 8. GO bubble enrichment plot of candidate genes for the withers height trait in Yanqi horses.

To elucidate the metabolic pathways involving these candidate genes *in vivo*, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed on the gene cluster significantly associated with WH in Yanqi horses. A bubble enrichment plot was generated based on statistically significant pathways, as shown in Figure 9. These genes were found to be significantly enriched in three primary pathways: morphine addiction, cytoskeleton in muscle cells, and retrograde endocannabinoid signaling. Among these, the highest number of genes was enriched in the cytoskeleton in muscle cells pathway.

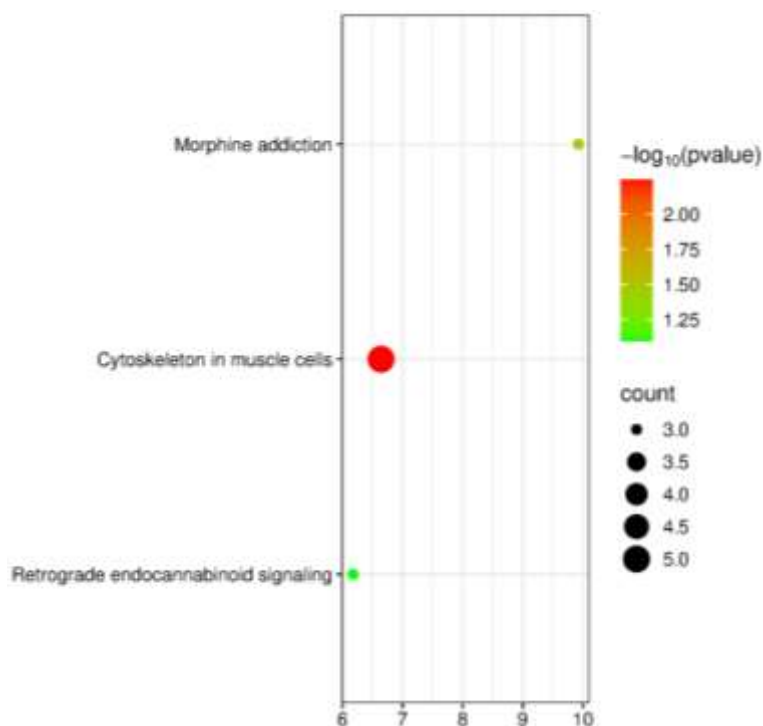


Figure 9. KEGG bubble enrichment plot of candidate genes for WH traits in Yanqi horses.

3.8. Functional Annotation and Enrichment Analysis of Candidate Genes for BL Traits in Yanqi Horses

To further investigate the functions of candidate genes significantly associated with BL traits in Yanqi horses, GO functional annotation was conducted. The annotation structure was categorized into three domains: biological process, cellular component, and molecular function. As illustrated in Figure 10 and ranked by P-value, the enriched GO categories within each domain were as follows: (1) Biological process: DNA cytosine deamination, cytidine to uridine editing, negative regulation of single-stranded virus via double-stranded DNA intermediate, etc. (2) Cellular component: P-body, chloride channel complex, PRC1 complex, etc. (3) Molecular function: cytidine deaminase activity, structural constituent of myelin sheath, nitrate reductase activity, etc.

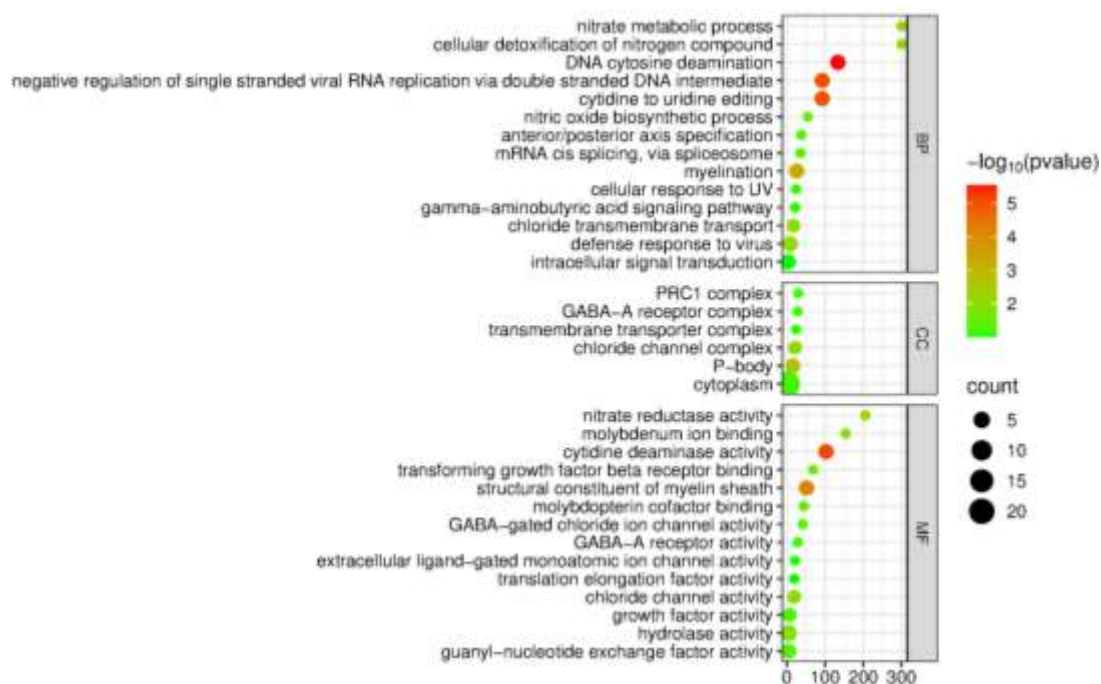


Figure 10. GO bubble enrichment plot of candidate genes for body length in Yanqi horses.

To elucidate the metabolic pathways in which the candidate genes participate *in vivo*, KEGG pathway enrichment analysis was conducted on the gene cluster significantly associated with BL in Yanqi horses. A bubble enrichment plot was generated based on statistically significant pathways, as illustrated in Figure 11. These candidate genes were significantly enriched in six pathways: Viral life cycle - HIV-1, TGF-beta signaling pathway, signaling pathways regulating pluripotency of stem cells, AMPK signaling pathway, human immunodeficiency virus 1 infection, and regulation of actin cytoskeleton. Among these, the highest number of genes was enriched in the human immunodeficiency virus 1 infection pathway.

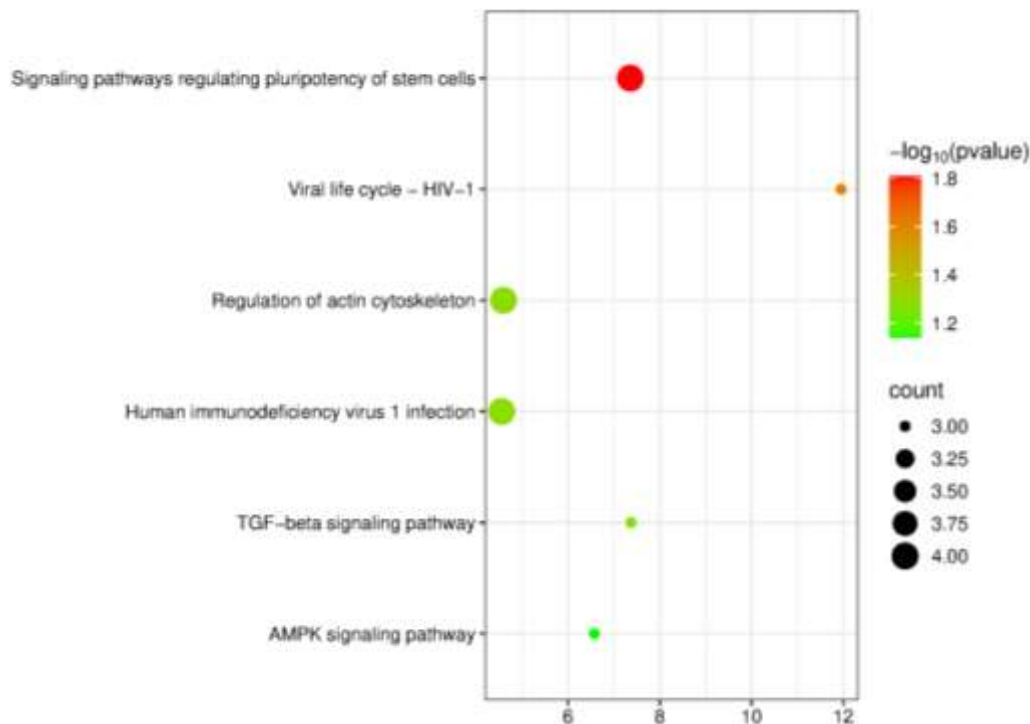


Figure 11. KEGG bubble enrichment plot of candidate genes for body length traits in Yanqi horses.

3.9. Functional Annotation and Enrichment Analysis of Candidate Genes for Hg Traits in Yanqi Horses

To further investigate the biological functions of candidate genes significantly associated with HG traits in Yanqi horses, GO functional annotation was conducted. The annotation structure was classified into three domains: biological process, cellular component, and molecular function. As illustrated in Figure 12 and ranked by P-value, the enriched GO categories within each domain were as follows: (1) Biological process: muscle contraction, nitrate metabolic process, cellular detoxification of nitrogen compound, etc. (2) Cellular component: myofibril, myosin complex, myosin filament, etc. (3) Molecular function: microfilament motor activity, cytoskeletal motor activity, etc.

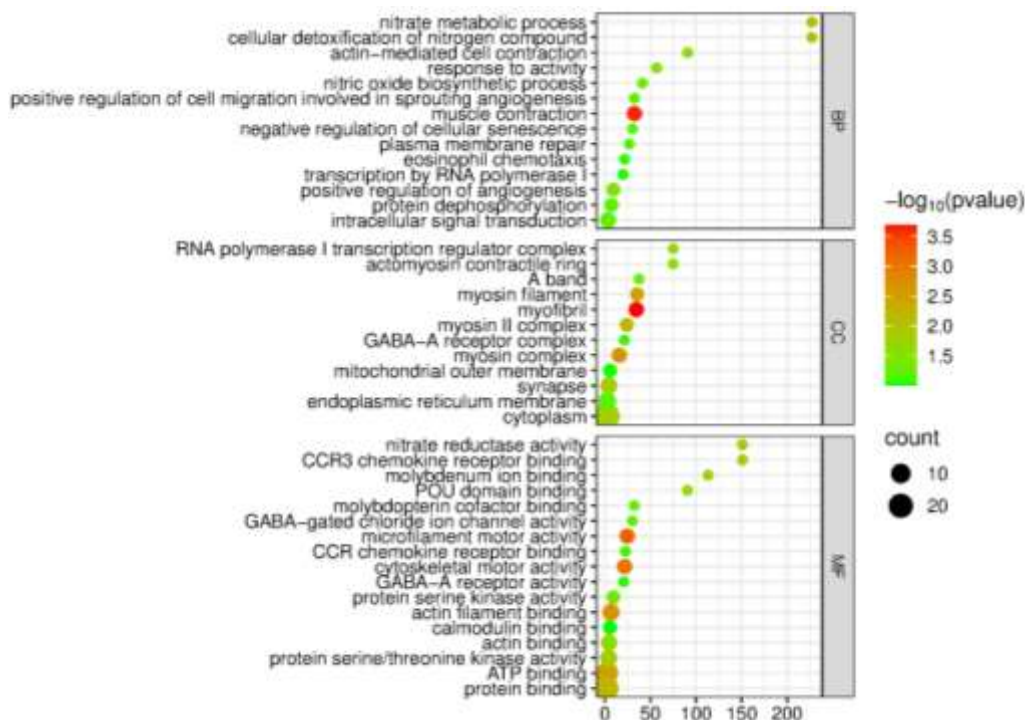


Figure 12. GO bubble enrichment plot of candidate genes for the heart girth trait in Yanqi horses.

To elucidate the metabolic pathways in which these candidate genes participate in vivo, KEGG pathway enrichment analysis was conducted on the gene cluster significantly associated with HG in Yanqi horses. A bubble enrichment plot was generated based on statistically significant pathways, as shown in Figure 13. These genes were significantly enriched in six pathways: GABAergic synapse, morphine addiction, motor proteins, dopaminergic synapse, cytoskeleton in muscle cells, and chemokine signaling pathway. Among these, the highest number of genes was enriched in the motor proteins pathway.

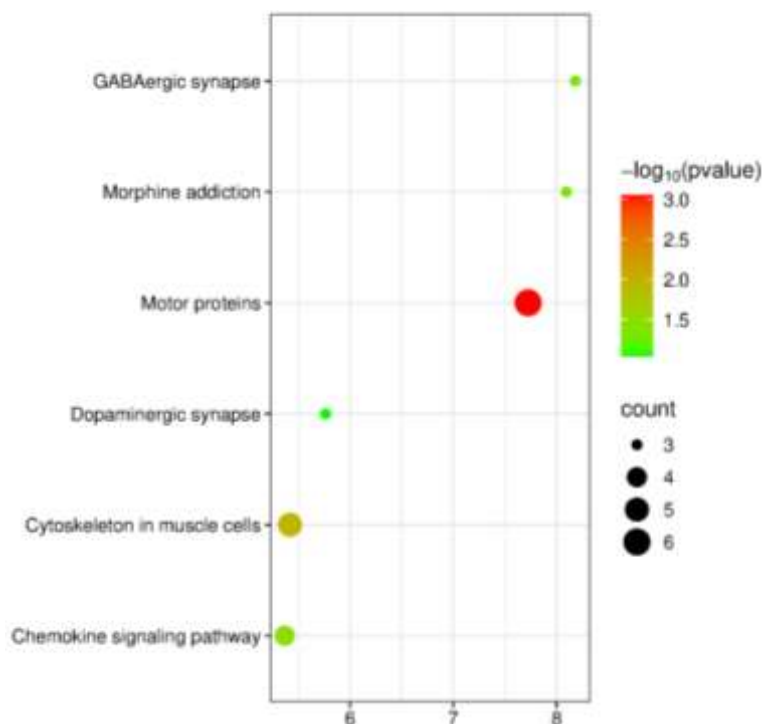


Figure 13. KEGG bubble enrichment plot of candidate genes for the heart girth trait in Yanqi horses.

3.10. Functional Annotation and Enrichment Analysis of Candidate Genes for CBC Traits in Yanqi Horses

To further investigate the biological functions of candidate genes significantly associated with CBC traits in Yanqi horses, GO functional annotation was carried out. The annotation structure was organized into three categories: biological process, cellular component, and molecular function. As illustrated in Figure 14 and ranked by P-value, the enriched GO terms in each category were as follows: (1) Biological process: photoreceptor cell maintenance, gamma-aminobutyric acid signaling pathway, axoneme assembly, etc. (2) Cellular component: neuron projection, synapse, chloride channel complex, etc. (3) Molecular function: serine-type endopeptidase inhibitor activity, chloride channel activity, catalytic activity, etc.

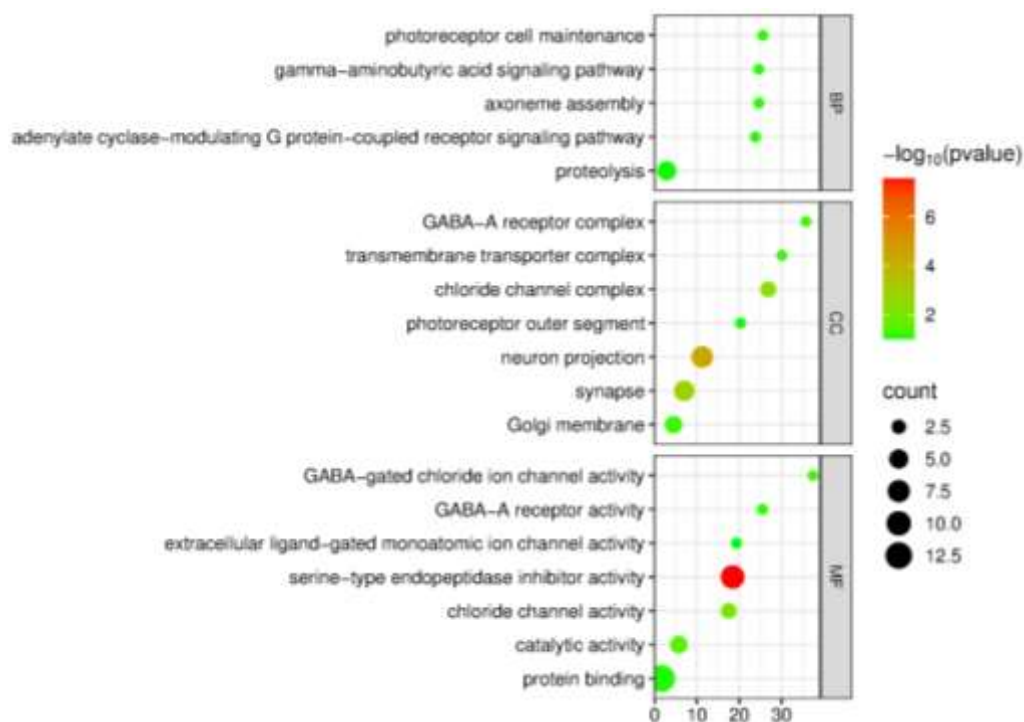


Figure 14. GO bubble enrichment plot of candidate genes for cannon bone circumference in Yanqi horses.

To elucidate the metabolic pathways in which candidate genes participate *in vivo*, KEGG pathway enrichment analysis was conducted on the gene cluster significantly associated with CBC in Yanqi horses. A bubble enrichment plot was generated based on statistically significant pathways, as illustrated in Figure 15. These candidate genes were significantly enriched in two pathways: nicotine addiction and neuroactive ligand-receptor interaction. Among these, the highest number of genes was enriched in the neuroactive ligand-receptor interaction pathway.

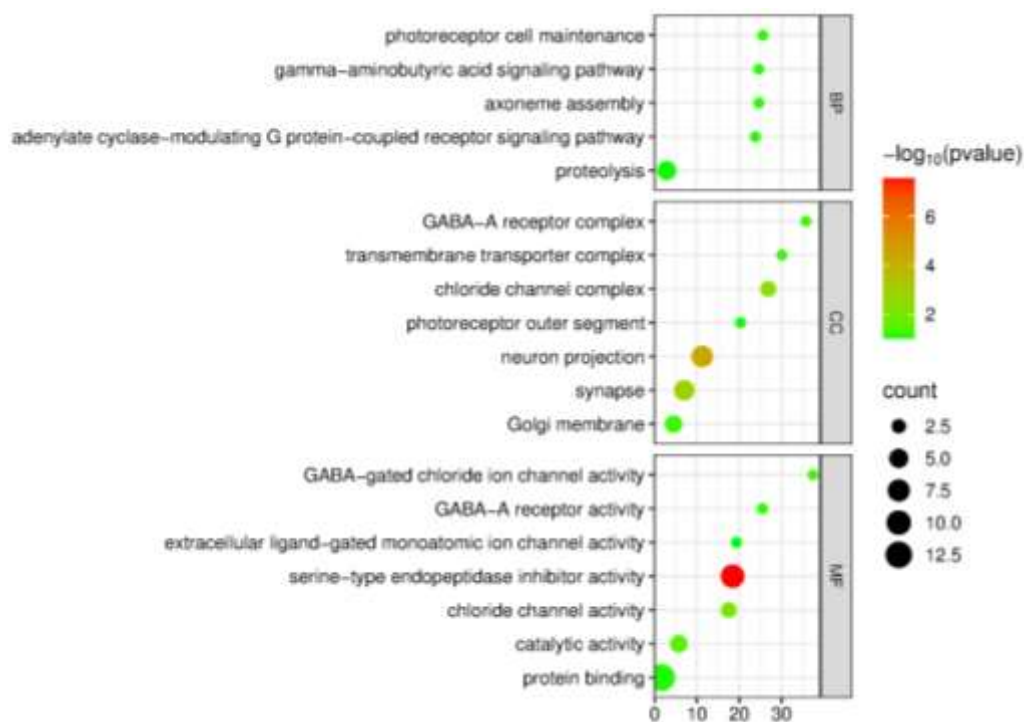


Figure 15. KEGG bubble enrichment plot of candidate genes for cannon bone circumference traits in Yanqi horses.

4. Discussion

In this study, a GWAS was conducted on four body size traits in 183 Yanqi horses using the MLM based on resequencing data. Following Bonferroni multiple testing correction of the P-values obtained, a total of 185 loci were identified as significantly associated with body size traits in Yanqi horses. Gene annotation was performed within a 200 kb region, both upstream and downstream of these significant loci, leading to the identification of 359 candidate genes. Among these, 45 loci were significantly associated with WH, 25 with BL, 51 with HG, and 64 with CBC. Notably, five candidate genes, GABRB1, FIGN, GABRA4, ENSECAG0000051747, and COX7B2, were annotated across all four traits and are known to be involved in multiple biological processes, with previous studies reporting their association with body size development.

The gamma-aminobutyric acid (GABA) receptor gene GABRB1 may indirectly regulate growth hormone (GH) secretion through the hypothalamic-pituitary-growth axis, thereby modulating skeletal development and body size traits in animals. In a genome-wide association analysis of Dutch Holstein cattle, Harmen P. Doekes [16] et al. reported that genetic markers proximal to GABRB1 were significantly associated with body size characteristics such as withers height and hip width. Similarly, Wossenie Mebratie [17] et al. identified SNPs near GABRB1 in a broiler chicken GWAS, which were associated with tibial length and body weight, and hypothesized that these variants might affect body size traits through GABA signaling pathways regulating skeletal development and the growth axis. Both GABRA4 and GABRB1, members of the GABA receptor gene family, are essential in the nervous system and function by mediating GABAergic inhibitory neurotransmission [18]. In a bovine genome-wide study, Muhammad S. Tahir [19] et al. suggested that GABA signaling transmitted via GABRA4 receptors acts in coordination with N-methyl-D-aspartate and α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors, leading to calcium ion channel activation, elevated intracellular calcium concentrations, and subsequent stimulation of gonadotropin-releasing hormone secretion, ultimately impacting pubertal and body size development. Furthermore, Zhao et al. [20] demonstrated in Simmental cattle that the GABA synaptic pathway was related to live weight and contributed to the regulation of feeding behavior. Thus, GABRA4 and GABRB1 may influence body size traits in Yanqi horses predominantly through mechanisms involving neural (GABA receptor

neurotransmission) and hormonal (growth hormone-releasing hormone secretion) regulation, which warrants further empirical validation. FIGN encodes an ATP-dependent microtubule-severing enzyme of the AAA-ATPase family, which plays a fundamental role in cytoskeletal dynamics, mitosis, and ciliary function. It is hypothesized that FIGN may affect body size by modulating skeletal growth, muscle development, and fat metabolism. G. A. Cox [21] et al. demonstrated that FIGN influences mammalian development, with its mutant form, fidgetin, promoting osteocyte differentiation, thus implicating it as a potential regulator of skeletal and muscular growth. Therefore, it is proposed that FIGN may influence body size development in Yanqi horses through metabolic mechanisms involving skeletal and adipose tissue regulation, a hypothesis that requires further investigation. COX7B2, an auxiliary subunit of cytochrome c oxidase in the mitochondrial electron transport chain, is involved in oxidative phosphorylation and ATP synthesis [22]. Sumona Akter [23] et al. demonstrated that COX7B2 plays a role in the assembly of mitochondrially encoded core subunits of cytochrome c oxidase, and that defects in this assembly process lead to mitochondrial dysfunction, substantially reducing exercise capacity and subsequently impairing growth and development in animals. Given its role in mitochondrial energy metabolism, it is speculated that mutations in COX7B2 may enhance its transcriptional activity, improve energy production efficiency, and thus influence body size development during critical growth stages in Yanqi horses. Currently, limited research is available on ENSECAG00000051747, and its specific biological function remains to be elucidated. Further exploration and functional validation are necessary to determine its relevance to body size traits.

GO analysis revealed that, at the biological process level, the candidate genes were significantly enriched in signal transduction, organic substance response, and specific catalytic activity processes. Analysis of cellular components indicated that these genes were predominantly localized within myosin complexes, myofibrils, and synaptic structures. Molecular function enrichment further identified ion channel activity and membrane protein functionality as prominent characteristics. KEGG pathway analysis demonstrated that significantly enriched signaling pathways were chiefly concentrated in the cytoskeleton in muscle cells, regulation of actin cytoskeleton, and neuroactive ligand-receptor interaction. A growing body of literature has highlighted the essential involvement of the cytoskeletal system in regulating muscle development and body size traits. Petra Gimpel et al. [24] demonstrated that mutations in the cytoskeletal protein Nesprin-1 result in aberrant nuclear positioning within muscle cells, ultimately leading to the manifestation of muscular dystrophy symptoms. This finding elucidated that cytoskeletal proteins contribute not only to cellular architecture maintenance but also directly influence normal muscle development. Complementary findings by Jianyan Zeng et al. [25] indicated that cytoskeletal pathways in muscle cells are engaged in critical biological processes such as myotube fusion and myofiber organization, primarily by regulating actin polymerization and microtubule dynamics, thereby impacting skeletal muscle morphology and physiological function. Mutations in Nesprin-1 may contribute to abnormalities in muscle cell positioning and potentially hinder body size development in Yanqi horses. At the molecular level, Thomas D. Pollard et al. [26] demonstrated that within the actin cytoskeleton regulatory pathway, filamentous actin provides structural support and mechanical propulsion to cells, thereby contributing to cellular shape and motility, which in turn influence organismal growth and phenotypic size variation. Furthermore, Sarah J. Heasman et al. [27] revealed that regulation of the actin cytoskeleton via Rho GTPase signaling modulates cell proliferation and migration, consequently impacting muscle tissue development and expansion. These collective observations highlight the close association between myofibril organization and muscle mass, as well as body size traits. Alterations in myofibril architecture may modulate muscle cell morphology and dynamics, leading to differences in body size traits among Yanqi horses. Beyond the cytoskeletal network, endocrine signaling pathways have also been implicated in the regulation of body size. Andrea M. Hanson et al. [28] reported that the hypothalamic-pituitary-liver axis orchestrates postnatal growth and development through the GHRH-GH-IGF-1 signaling cascade. F. Lupu et al. [29] demonstrated that GH and IGF-1 act both independently and synergistically to drive postnatal growth, with these

pathways serving as key regulators of phenotypic body size differentiation. Thyroid hormone (TH) also plays a pivotal role in skeletal development and linear growth. J. H. Duncan Bassett et al. [30] revealed that TH deficiency halts skeletal maturation and growth, and that neuroactive ligand-receptor interaction pathways influence TH secretion, thereby modulating body size. Therefore, TH may influence skeletal development and growth in Yanqi horses, and TH insufficiency may contribute to delayed skeletal maturation, ultimately resulting in observable differences in body size.

5. Conclusions

Genomic regions and key genes associated with body size traits in Yanqi horses were identified through a GWAS. A total of 185 SNPs related to body size characteristics were discovered, along with 359 putative candidate genes. Among these, *GABRB1*, *FIGN*, *GABRA4*, *ENSECAG00000051747*, and *COX7B2* were suggested to exert pivotal regulatory effects on growth and developmental processes in Yanqi horses. The findings presented in this study warrant further validation through expanded research efforts to ascertain their applicability in the selective breeding of Yanqi horses.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: W.S., Conceptualisation and design, research, drafting, review and revision; Z.S. and D.C.; Visualisation, software, methodology, resources; P.L., Conceptualisation, supervision, funding acquisition, writing—review and editing; X.Y., Methodology, formal analysis; J.M., Methodology, formal analysis; Z.S., Validation, Formal analysis, Data Curation; Y.Z., Conceptualisation; formal analysis; writing—review and editing.

Funding: The Major Science and Technology Project of Xinjiang Uygur Autonomous Region (grant No. 2022A02013-1); Central Guidance Project for Local Science and Technology Development - (Research on the Regulation Mechanism of Horse Breeding and Athletic Performance) (grant No. ZYYD2025JD02); National Natural Science Foundation of China Youth Program (grant No. 32202667); Graduate School-level Scientific Research and Innovation Project of Xinjiang Agricultural University (grant No. XJAUGRI2024021).

Institutional Review Board Statement: All animal procedures in this study were approved by the Animal welfare and Ethics Committee of Xinjiang Agricultural University (approval number: 2023020).

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data for this study is currently being uploaded.

Conflicts of Interest: The authors declare no competing financial interests.

References

1. Limuge Che. Genome-wide Association Study For growth Traits-related Candidate Functional Gene in Chinese Simmental Cattle. Master's Thesis, Inner Mongolia University, 2023.
2. Xueliang Sun. Genetic Diversity and Genome-wide Association Study of Coat Color in Chuanzhong Black Goats. Master's Thesis, Sichuan Agricultural University, 2023.
3. Li Haodong. Study on Whole Genome Selection Technique for Body Weight and other Traits of Sika Deer. Master's Thesis, Chinese Academy of Agricultural Sciences, 2023.
4. Honghong Hu. Genetic Basis and Genomic Selection of Longevity Traits in Holstein Cattle. PhD Thesis, Ningxia University, 2023.
5. Ziyi Wu. Genome-wide Association Study and Genomic Selection for Growth-related Traits in Yunong-black. Master's Thesis, Henan Agricultural University, 2024.
6. Xiaotian Lu. Genome-wide Association Analysis of Fleece Traits in Northwest Xizang White Cashmere Goat. Master's Thesis, Inner Mongolia Agricultural University, 2024.
7. Menghan Wang. Genetic Parameter Estimation and Genome-wide Association Study for Growth Traits in East Friesian-Hu hybrid sheep. Master's Thesis, Northwest A&F University, 2024.

8. Yan Sun. Genome-wide Association Study For Numbers of Thoracic and Lumbar Vertebrae in Dezhou Donkey and Functional Study of Candidate Genes. PhD Thesis, Shandong Agricultural University, 2023.
9. Chuankun Wang. GWAS, ROH and SVs Mining Candidate Genes for Important Traits in Yili Horse. PhD Thesis, Xinjiang Agricultural University, 2024.
10. Ropka-Molik, K.; Stefaniuk-Szmukier, M.; Musiał, A.D.; Velie, B.D. The Genetics of Racing Performance in Arabian Horses. *Int J Genomics* **2019**, *2019*, 9013239, doi:10.1155/2019/9013239.
11. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *Gigascience* **2015**, *4*, 7, doi:10.1186/s13742-015-0047-8.
12. Tang, Y.; Liu, X.; Wang, J.; Li, M.; Wang, Q.; Tian, F.; Su, Z.; Pan, Y.; Liu, D.; Lipka, A.E.; et al. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. *Plant Genome* **2016**, *9*, doi:10.3835/plantgenome2015.11.0120.
13. Zhou, X.; Stephens, M. Genome-Wide Efficient Mixed-Model Analysis for Association Studies. *Nat Genet* **2012**, *44*, 821–824, doi:10.1038/ng.2310.
14. Yang, H.; Wang, K. Genomic Variant Annotation and Prioritization with ANNOVAR and wANNOVAR. *Nat Protoc* **2015**, *10*, 1556–1566, doi:10.1038/nprot.2015.105.
15. Sherman, B.T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M.W.; Lane, H.C.; Imamichi, T.; Chang, W. DAVID: A Web Server for Functional Enrichment Analysis and Functional Annotation of Gene Lists (2021 Update). *Nucleic Acids Res* **2022**, *50*, W216–W221, doi:10.1093/nar/gkac194.
16. Doekes, H.P.; Veerkamp, R.F.; Bijma, P.; Hiemstra, S.J.; Windig, J.J. Trends in Genome-Wide and Region-Specific Genetic Diversity in the Dutch-Flemish Holstein-Friesian Breeding Program from 1986 to 2015. *Genet Sel Evol* **2018**, *50*, 15, doi:10.1186/s12711-018-0385-y.
17. Mebratie, W.; Reyer, H.; Wimmers, K.; Bovenhuis, H.; Jensen, J. Genome Wide Association Study of Body Weight and Feed Efficiency Traits in a Commercial Broiler Chicken Population, a Re-Visitation. *Sci Rep* **2019**, *9*, 922, doi:10.1038/s41598-018-37216-z.
18. Olsen, R.W.; Sieghart, W. GABA A Receptors: Subtypes Provide Diversity of Function and Pharmacology. *Neuropharmacology* **2009**, *56*, 141–148, doi:10.1016/j.neuropharm.2008.07.045.
19. Tahir, M.S.; Porto-Neto, L.R.; Gondro, C.; Shittu, O.B.; Wockner, K.; Tan, A.W.L.; Smith, H.R.; Gouveia, G.C.; Kour, J.; Fortes, M.R.S. Meta-Analysis of Heifer Traits Identified Reproductive Pathways in Bos Indicus Cattle. *Genes (Basel)* **2021**, *12*, 768, doi:10.3390/genes12050768.
20. Zhao, G.; Liu, Y.; Niu, Q.; Zheng, X.; Zhang, T.; Wang, Z.; Xu, L.; Zhu, B.; Gao, X.; Zhang, L.; et al. Runs of Homozygosity Analysis Reveals Consensus Homozygous Regions Affecting Production Traits in Chinese Simmental Beef Cattle. *BMC Genomics* **2021**, *22*, 678, doi:10.1186/s12864-021-07992-6.
21. Cox, G.A.; Mahaffey, C.L.; Nystuen, A.; Letts, V.A.; Frankel, W.N. The Mouse Fidgetin Gene Defines a New Role for AAA Family Proteins in Mammalian Development. *Nat Genet* **2000**, *26*, 198–202, doi:10.1038/79923.
22. Timón-Gómez, A.; Nývltová, E.; Abriata, L.A.; Vila, A.J.; Hosler, J.; Barrientos, A. Mitochondrial Cytochrome c Oxidase Biogenesis: Recent Developments. *Semin Cell Dev Biol* **2018**, *76*, 163–178, doi:10.1016/j.semcdb.2017.08.055.
23. Akter, M.S.; Hada, M.; Shikata, D.; Watanabe, G.; Ogura, A.; Matoba, S. CRISPR/Cas9-Based Genetic Screen of SCNT-Reprogramming Resistant Genes Identifies Critical Genes for Male Germ Cell Development in Mice. *Sci Rep* **2021**, *11*, 15438, doi:10.1038/s41598-021-94851-9.
24. Gimpel, P.; Lee, Y.L.; Sobota, R.M.; Calvi, A.; Koullourou, V.; Patel, R.; Mamchaoui, K.; Nédélec, F.; Shackleton, S.; Schmoranz, J.; et al. Nesprin-1 α -Dependent Microtubule Nucleation from the Nuclear Envelope via Akap450 Is Necessary for Nuclear Positioning in Muscle Cells. *Curr Biol* **2017**, *27*, 2999–3009.e9, doi:10.1016/j.cub.2017.08.031.
25. Zeng, J.; Xi, J.; Li, B.; Yan, X.; Dai, Y.; Wu, Y.; Xiao, Y.; Pei, Y.; Zhang, M. Microtubules Play a Crucial Role in Regulating Actin Organization and Cell Initiation in Cotton Fibers. *Plant Cell Rep* **2022**, *41*, 1059–1073, doi:10.1007/s00299-022-02837-2.
26. Pollard, T.D.; Cooper, J.A. Actin, a Central Player in Cell Shape and Movement. *Science* **2009**, *326*, 1208–1212, doi:10.1126/science.1175862.
27. Heasman, S.J.; Ridley, A.J. Mammalian Rho GTPases: New Insights into Their Functions from in Vivo Studies. *Nat Rev Mol Cell Biol* **2008**, *9*, 690–701, doi:10.1038/nrm2476.

28. Hanson, A.M.; Stodieck, L.S.; Cannon, C.M.A.; Simske, S.J.; Ferguson, V.L. Seven Days of Muscle Re-Loading and Voluntary Wheel Running Following Hindlimb Suspension in Mice Restores Running Performance, Muscle Morphology and Metrics of Fatigue but Not Muscle Strength. *J Muscle Res Cell Motil* **2010**, *31*, 141–153, doi:10.1007/s10974-010-9218-5.
29. Lupu, F.; Terwilliger, J.D.; Lee, K.; Segre, G.V.; Efstratiadis, A. Roles of Growth Hormone and Insulin-like Growth Factor 1 in Mouse Postnatal Growth. *Dev Biol* **2001**, *229*, 141–162, doi:10.1006/dbio.2000.9975.
30. Bassett, J.H.D.; Williams, G.R. Role of Thyroid Hormones in Skeletal Development and Bone Maintenance. *Endocr Rev* **2016**, *37*, 135–187, doi:10.1210/er.2015-1106.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.