veterinary sciences

MDPI

Article

Genome-Wide Characterization of Copy Number Variations and Their Associations with Early Body Weight and Body Measurements in Zhongwei Goats

Luli Qi [1,†], Mingxuan Han [1,†], Youjun Rong [1], Zhanqiang Zhao [2], Haidong Du [2], Dun Li [2], Xinle Wang [1], Yong Ye [2], Bing Zhang [2], Xiaofang Ao [1], Fangzheng Shang [1], Ruijun Wang [1,*] and Yanjun Zhang [1,3,4,5,*]

[1] College of Animal Science, Inner Mongolia Agricultural University, Hohhot 010018, China; qll18365738991@163.com (L.Q.); hanmx1997@163.com (M.H.); rongyoujun1@163.com (Y.R.); wangxinle163@163.com (X.W.); aoxiaofang666@163.com (X.A.); sfznnd2020@126.com (F.S.)
[2] Zhongwei Goat Breeding Farm, Zhongwei 755006, China; 13995182407@163.com (Z.Z.); 13909581032@163.com (H.D.); 18169500573@163.com (D.L.); 13629535097@163.com (Y.Y.); 13739563638@163.com (B.Z.)
[3] Key Laboratory of Mutton Sheep Genetics and Breeding, Ministry of Agriculture, Hohhot 010018, China
[4] Key Laboratory of Goat and Sheep Genetics, Breeding and Reproduction in Inner Mongolia Autonomous Region, Hohhot 010018, China
[5] China-Mongolia Biomacromolecule Application "Belt and Road" Joint Laboratory, Hohhot 010018, China
* Correspondence: ruijunw@imau.edu.cn (R.W.); yanjunz@imau.edu.cn or imauzyj@163.com (Y.Z.)
† These authors contributed equally to this work.

Simple Summary

Zhongwei goats are valued for the distinctive white, curly coats of their young, but the genetic factors associated with early body weight and body measurements remain poorly understood. Some sections of DNA may be missing or repeated, and these changes may be associated with differences in body size among animals. We examined 367 Zhongwei goat kids aged 38–52 days and identified 28 potential associations between DNA deletions and body weight or body measurements, including two regions showing signals for more than one trait. Because some associations involved relatively few animals, these results should be interpreted cautiously. Laboratory testing of selected candidate DNA deletion regions showed relative copy-number changes consistent with the predictions, although the results were insufficient to independently confirm their actual copy-number states. Some genes within or near these regions have previously been linked to muscle development, bone formation, and fat metabolism. These findings provide valuable clues for further investigating the genetic basis of early body weight and body measurements in Zhongwei goats, although their potential value for breeding requires confirmation in independent studies.

Abstract

Copy number variations (CNVs) in Zhongwei goats and their associations with early body weight and body measurements remain poorly characterized. We analyzed whole-genome resequencing data from 367 Zhongwei goat kids aged 38–52 days, using CNVpytor and GCTA fastGWA to detect CNVs and test associations with body weight, body height, body length (BL), and chest circumference. We identified 12,660 CNV variant records (71.26% deletions) and 1176 CNV regions spanning 53.34 Mb (2.16% of autosomal length). At $p < 0.01$, 28 exploratory deletion CNV–trait signals were identified; none passed four-trait false discovery rate correction. Two regions showed cross-trait overlap.

Academic Editor: Xueming Zhang

Received: 14 October 2026
Revised: 17 November 2026
Accepted: 19 November 2026
Published: 20 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Substantial genotype missingness and small effective sample sizes warrant caution. Positional annotation identified 287 nonredundant candidate genes; enrichment highlighted transport and lipid metabolism but was sensitive to regional gene clustering. *ODC1*, *RBFOX2*, *ECM2*, *OGN*, *OMD*, *ASPN*, *TNXB*, and *ELOVL5* were prioritized for further study. Under internal relative calibration, quantitative polymerase chain reaction of 10 candidate deletion CNVs in 20 goats classified 190 of 200 preselected deletion-predicted sample–locus combinations (95.0%) as deletions, without independently confirming absolute copy number. These findings characterize CNVs and nominate candidates for independent validation.

Keywords: Zhongwei goats; copy number variation; whole-genome resequencing; genome-wide association study; early body weight; body measurements; quantitative PCR

1. Introduction

Goats are important small-ruminant livestock that provide meat, milk, fiber, hides, and other products and play important roles in food supply, rural economies, and sustainable animal agriculture [1]. Growth rate and body conformation are closely related to production potential and breeding efficiency and are therefore important targets for genetic improvement [2,3]. Zhongwei goats are renowned for their high-quality white curly kidskin, and previous studies have investigated the mechanisms underlying hair curliness and hair follicle development at the epigenetic, transcriptomic, and proteomic levels [4,5]. In contrast, the genetic basis of body weight and body measurements in this breed remains relatively understudied. More recently, Han et al. conducted a SNP-based GWAS using whole-genome sequencing (WGS) data and identified 342 SNPs significantly associated with growth traits in Zhongwei goats. They also identified candidate genes involved in bone formation, muscle development, and lipid metabolism, providing SNP-level evidence for the genetic basis of growth traits in this breed [6].

Copy number variation (CNV) is an important class of genomic structural variation, characterized primarily by gains or losses of DNA segments relative to a reference genome [7]. Although CNVs are generally less numerous than SNPs, they often span larger genomic regions and may affect both coding genes and regulatory elements, thereby influencing gene expression and complex phenotypes through altered gene dosage, disruption of gene structure, or regulatory changes [8]. In addition, some CNVs show limited linkage disequilibrium with nearby SNPs; therefore, analyses based solely on SNP markers may not fully capture the genetic variation represented by these CNVs [9,10]. CNV analysis can thus complement existing SNP-based studies and broaden our understanding of the genetic basis of early body weight and body measurements in Zhongwei goats from a structural-variation perspective.

Early genome-wide CNV studies relied mainly on array comparative genomic hybridization (aCGH) and high-density SNP arrays. Their detection capacity was constrained by probe location, genomic coverage, and resolution, making some novel or low-frequency CNVs difficult to identify [11]. With the development of whole-genome sequencing (WGS), continuous genomic coverage and abundant read information have reduced dependence on predefined probes and provided a stronger basis for population-level CNV detection [12]. Previous studies have shown that goat populations and sheep breeds differ in CNV/CNVR number, frequency, and distribution [8,13], suggesting that CNV landscapes can be influenced by genetic background. Consequently, CNV maps generated in other goat breeds may not fully represent the structural-variation landscape of Zhongwei goats, supporting the need for breed-specific characterization.

Integrating CNV information with phenotypic records enables further evaluation of the contribution of structural variation to complex traits. Liu et al. investigated the associations of common CNVs with multiple growth and body-size traits in African indigenous goats. After false discovery rate (FDR) correction, two significantly associated CNVs were identified; CNV27 was associated with both chest width and ischial width, and the two CNVs overlapped *ADCY1* and *SNX29*, respectively [14]. More recently, a CNV-based GWAS using whole-genome resequencing data from 461 Inner Mongolian cashmere goats identified 11 CNVs significantly associated with early growth traits. Some loci were associated with multiple traits, and *SOX15*, *FGF11*, and *DVL2* were among the candidate genes identified [15]. In addition, a study of CNVs in *KCNJ15* found that associations with body weight, body height, body length, and chest circumference differed among Chinese goat breeds, further indicating that the phenotypic effects of CNVs may depend on population genetic background [16]. Similar associations have also been reported in sheep, cattle, and pigs [13,17,18]. Collectively, these studies suggest that CNVs represent a potentially important source of genetic variation associated with livestock growth traits and that their genomic distributions and trait associations may vary among breeds or populations, supporting the need for independent CNV association studies in individual breeds.

Although the genetic basis of body conformation in Zhongwei goats has been partly characterized at the SNP level, the genome-wide CNV/CNVR landscape and its associations with early body weight and body measurements have not been systematically investigated. Therefore, using whole-genome resequencing data from 367 Zhongwei goats, we constructed a CNV/CNVR map and characterized its genomic distribution. We then performed CNV-GWAS for BW, BH, BL, and CC and compared shared and trait-specific candidate association signals across the four traits. Genes within or near candidate regions were then annotated, and the resulting candidate gene set was subjected to functional enrichment analysis. Selected candidate deletion CNVs were subsequently evaluated by quantitative PCR (qPCR) using an internal relative calibration approach. These analyses provide a basis for subsequent replication in independent populations and the functional investigation of key CNVs.

2. Materials and Methods

2.1. Experimental Animals and Phenotypic Measurements

A total of 367 Zhongwei goat kids (189 males and 178 females) from the Zhongwei Goat Breeding Farm in the Ningxia Hui Autonomous Region, China, were included in this study. Of these, 240 goats were also included in the previous SNP-based GWAS by Han et al. (2025) [6], with overlapping whole-genome sequencing data and records of early body weight and body measurements. Phenotypic measurements and ear-tissue sampling were conducted in spring 2023, when the kids were 38–52 days old. Given the extended measurement period, limited field personnel, and rapid growth at this age, age at measurement was categorized into three groups (38–42, 43–47, and 48–52 days) for adjustment in the association analyses. All kids were single-born and were raised with their dams during the suckling period under uniform husbandry and nutritional conditions. The kids were housed with their dams in group pens connected to outdoor exercise areas, where they were allowed to suckle freely. They were provided with alfalfa hay, age-appropriate pelleted concentrate, and clean drinking water. No obvious clinical signs of disease were observed at the time of phenotypic measurements and ear-tissue sampling. The measurement procedures for the four phenotypes are presented in Table 1.

Following phenotypic measurements, the biopsy site along the posterior margin of the ear was disinfected, and local infiltration anesthesia was administered using 0.15–0.25 mL of 2% lidocaine hydrochloride (Jilin Huamu Animal Health Products Co., Ltd., Changchun, China) per animal. After waiting at least 5 min and confirming the onset of local

anesthesia, approximately 0.5 g of ear tissue was collected using biopsy forceps. The samples were immediately placed in cryovials containing 75% ethanol (Tianjin Zhiyuan Chemical Reagent Co., Ltd., Tianjin, China) and stored at −80 °C until genomic DNA extraction. Hemostasis was achieved after sampling, and the kids were returned to their original herd after recovery. Their general health and biopsy sites were monitored for the following 2–3 days.

WPS Spreadsheets (v12.1.0.29690; Kingsoft Office Software) was used to organize the phenotypic data, record sample size, and calculate the mean, standard deviation, minimum, maximum, and coefficient of variation. Kolmogorov–Smirnov tests were performed in Origin 2021 to assess the normality of body weight, body height, body length, and chest circumference.

Table 1. Measurement methods for body weight and body measurements in Zhongwei goats.

Trait	Measurement Method	Instrument	Unit
Body weight (BW)	Measured before morning feeding	Platform scale	kg
Body height (BH)	Vertical distance from the highest point of the withers to the ground	Measuring stick	cm
Body length (BL)	Straight-line distance from the point of shoulder to the posterior end of the ischial tuberosity	Measuring stick	cm
Chest circumference (CC)	Circumference of the chest measured immediately behind the scapula	Measuring tape	cm

Note: During body measurement, all goats were maintained in a natural standing position on a firm and level surface.

2.2. Whole-Genome Resequencing and Data Quality Control

Genomic DNA was extracted from ear tissue using the phenol–chloroform method. DNA integrity and purity were assessed by 2% agarose gel electrophoresis and a NanoDrop 2000 UV spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), respectively. DNA samples that passed quality control were randomly fragmented to approximately 350 bp using a Covaris sonicator (Covaris, LLC, Woburn, MA, USA), followed by end repair, A-tailing, adapter ligation, purification, and PCR amplification for library construction. After Qubit quantification (Thermo Fisher Scientific, Waltham, MA, USA), 150-bp paired-end sequencing was performed on the MGI-T7 platform (MGI Tech Co., Ltd., Shenzhen, China). The whole-genome resequencing data for all 367 Zhongwei goats are available in the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA1121220.

Raw sequencing reads were processed using fastp (v0.20.0) to remove low-quality reads and obtain clean reads [19]. The clean reads were then aligned to the goat ARS1 reference genome (GCF_001704415.1) using BWA (v0.7.17) [20]. SAMtools (v1.9) was used for SAM/BAM conversion and sorting [21], and PCR duplicate reads were identified and removed using Picard MarkDuplicates (v2.17.5). Finally, BAM files were indexed with SAMtools for subsequent CNV detection.

2.3. Whole-Genome CNV Detection and Filtering

Genome-wide CNVs were detected from the processed BAM files using CNVpytor (v1.3.2) [22]. Following the read-depth (RD) workflow, RD signals were extracted, binned, corrected for GC bias, and segmented using the mean-shift algorithm for CNV calling. Considering sequencing coverage, the target CNV detection scale (≥1 kb), and RD signal stability, a bin size of 1 kb was used. To improve call reliability, CNVs meeting any of the following criteria were excluded: CNVpytor e-val1 > 0.01, q0 > 0.5, length < 1 kb, pN > 0.5, or dG < 100 kb. Here, e-val1 is a read-depth-based statistic used by CNVpytor for CNV

detection; q0, pN, and dG denote the proportion of reads with mapping quality 0, the proportion of N bases in the reference genome, and the distance to the nearest genomic gap, respectively. The same CNVpytor calling parameters and quality-control criteria were applied independently to all 367 Zhongwei goats. Copy-number states in the exported VCF files were assigned according to the normalized read-depth values and the built-in classification rules of CNVpytor (v1.3.2). For deletion calls, normalized read-depth values below 0.25 were encoded as CN0, whereas values above 0.25 were encoded as CN1. For duplication calls, values ≤1.75 were encoded as CN3, values >1.75 and ≤2.25 as CN4, and values >2.25 as the nearest integer to twice the normalized read-depth value. No independently confirmed CN2 state was assigned to animals without a corresponding CNV call. The filtered CNV calls from individual animals were merged into a multi-sample VCF file using bcftools (v1.21) [23]. A custom Python script (Python v3.8.20) was used to summarize CNV length and chromosomal distribution, with CNV length calculated as end position − start position + 1. For the visualization of the CNV length distribution, CNVs ≤300 kb were grouped into seven size intervals: 1–2, 2–5, 5–10, 10–20, 20–50, 50–100, and 100–300 kb. GraphPad Prism (v9.5.0) was used to visualize CNV distributions across chromosomes and size intervals.

2.4. Construction of Copy Number Variation Regions (CNVRs)

A custom Python script was used to merge overlapping CNVs across individuals into nonredundant copy number variation regions (CNVRs). CNVs overlapping by at least 1 bp were merged into the same CNVR, with the boundaries defined by the minimum start and maximum end coordinates of the constituent CNVs [8]. To reduce the influence of sporadic calls, only merged regions containing at least one constituent CNV record carried by three or more individuals were retained [24]. Based on the constituent CNV types, CNVRs were classified as deletion, duplication, or mixed, with mixed CNVRs containing both deletion and duplication events. The number, cumulative length, type, and size distribution of CNVRs were summarized for each chromosome. CNVRs were grouped into six size intervals: 1–2, 2–5, 5–10, 10–20, 20–50, and 50–2000 kb. For calculating the proportion of the autosomal genome covered by CNVRs, the total length of the 29 autosomes in the ARS1 reference genome was set to 2,466,191,353 bp. To visualize their genome-wide distribution, CNVRs were assigned to non-overlapping 1-Mb windows according to their midpoint, and the number of CNVRs in each window was counted. The resulting statistics were visualized using GraphPad Prism (v9.5.0) and the CMplot package (v4.5.1) in R (v4.5.3). CNVRs were constructed solely to characterize population-level CNV distributions and were not used as markers in subsequent association analyses.

2.5. Population Structure Assessment and Genomic Relationship Matrix Construction

Principal component analysis (PCA) was performed on the quality-controlled CNV genotype data using PLINK (v1.90) [25], and the first three principal components (PC1–PC3) were extracted. The proportion of variance explained by each principal component was calculated by dividing its eigenvalue by the sum of the diagonal elements of the variance-standardized relationship matrix used for PCA. A two-dimensional scatter plot of PC1 versus PC2 was generated in R (v4.5.3) to assess population structure. A genomic relationship matrix (GRM) was constructed from the CNV genotypes using GCTA (v1.94.1), and a sparse GRM was generated using a relatedness threshold of 0.05. The sparse GRM was then provided as an input file for the subsequent GCTA association analyses [26].

2.6. CNV-Based Genome-Wide Association Analysis (CNV-GWAS)

A multi-sample VCF file containing the quality-controlled CNV calls was converted into binary genotype files using PLINK (v1.90) for subsequent association analyses. Unlike the CNVRs used to characterize the population-level distribution of CNVs, GWAS markers were derived from distinct variant records in the VCF file, and overlapping CNV records with different start coordinates were not further merged. Genotype conversion was based on the GT field in the VCF file rather than treating the CN values as continuous copy-number dosages. For example, GT:CN entries of 1/1:0 and 0/1:1 represented CN0 and CN1 deletions, respectively, whereas 0/1:3 represented a duplication and ./.:. indicated a missing genotype. Missing or uncertain genotypes, including partially missing GT entries, were retained as missing values, and individuals without a detected CNV were not automatically assigned a normal diploid genotype (CN = 2). Genotype quality control was performed using GCTA, with --geno 0.9 used to exclude markers with genotype missingness exceeding 90% and a default minor allele frequency (MAF) threshold of 0.0001. No additional minor allele count (MAC) filtering threshold was applied. The effective sample size for each CNV marker was obtained from the N column of the fastGWA output. No genotypes representing a confidently established diploid state (CN = 2) were available among the retained CNV markers. Consequently, the association tests with valid statistics primarily compared CN0 and CN1 deletion genotypes among animals with non-missing calls, rather than deletion carriers and confirmed diploid non-carriers across the entire cohort of 367 goats.

CNV-GWAS analyses were performed separately for BW, BH, BL, and CC using the fastGWA-mlm option in GCTA (v1.94.1) [26,27]. Sex and age group (38–42 d, 43–47 d, and 48–52 d) were included as categorical fixed effects, whereas the first three principal components (PC1–PC3) were included as continuous covariates. Based on the model-fitting results, GCTA automatically switched to linear regression for the association tests. The statistical model was as follows:

$$y = Xb + Z_{\mathrm{cnv}}\beta_{\mathrm{cnv}} + e$$

Here, y is an $n \times 1$ vector of phenotypic values; X is the design matrix for fixed effects and continuous covariates, while b is the corresponding coefficient vector; Z_{cnv} is the genotype vector for the CNV being tested, while β_{cnv} is its effect on the target trait; and e is the residual vector.

The estimated proportion of phenotypic variance explained (PVE) by each candidate CNV was calculated using the method described by Shim et al., based on the effect estimate, standard error, allele frequency, and actual sample size reported by fastGWA [28]. To standardize the reported frequency and effect direction of deletion CNVs, for loci with A1 = N and A2 = <DEL>, deletion allele frequency was calculated as $1 - \mathrm{AF1}$ and the deletion-allele effect was expressed as $\beta_{\mathrm{DEL}} = -\mathrm{BETA}$. Here, BETA denotes the effect estimate for the A1 allele in the original fastGWA output. This transformation was used only to standardize the direction of presentation for the deletion allele and did not alter the standard error (SE), p value, or estimated PVE. These PVE estimates were derived from marker-specific subsets of animals with non-missing genotypes and should not be interpreted as population-level estimates of the proportion of phenotypic variance explained across all 367 goats.

For each trait, genome-wide significance was assessed using a Bonferroni-corrected threshold of $p < 0.05/878 = 5.69 \times 10^{-5}$. Following previous CNV association studies, $p < 0.01$ was used as an exploratory threshold for identifying candidate CNVs [29,30]. In addition, the Benjamini–Hochberg procedure was applied to control the false discovery rate (FDR) and further assess the statistical evidence for association [31]. The primary correction included all retained markers, with markers lacking valid association p values assigned a

value of 1 solely for FDR calculation. A sensitivity analysis was also performed using only markers with valid association p values. Multiple-testing corrections were evaluated both within each trait and across all four traits combined.

The genomic inflation factor (λ_{GC}) was calculated from valid association p values using the median-based genomic control method [32]:

$$\lambda_{GC} = \frac{median\{[\Phi^{-1}(P_i/2)]^2\}}{\chi^2_{1,0.5}}$$

where P_i denotes a valid two-sided association p value, Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution, and $\chi^2_{1,0.5}$ is the theoretical median of the chi-square distribution with one degree of freedom (approximately 0.4549364231). Manhattan and quantile–quantile (Q–Q) plots were generated using the CMplot package (v4.5.1) in R (v4.5.3).

2.7. Annotation of Candidate CNVs and Functional Analysis of Candidate Genes

Using the goat ARS1 reference genome (GCF_001704415.1) and its gene annotation [33], BEDTools (v2.30.0) was used to annotate the exploratory candidate CNVs identified by CNV-GWAS [34]. Genes that overlapped a candidate CNV or were located within 500 kb upstream or downstream of the CNV interval were included in the candidate gene set. For genes lacking established gene symbols, the original LOC identifiers were retained, and available predicted gene descriptions were reported without assigning unverified gene names. The ±500 kb window was used for exploratory positional annotation rather than being defined by population-specific linkage disequilibrium estimates. GO and KEGG enrichment analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID; https://davidbioinformatics.nih.gov/) [35,36]. Genes annotated within the same ±500 kb window around CNV markers with valid association p values were used as the primary enrichment background. An alternative background based on all retained CNV markers was also examined. Terms with nominal $p < 0.05$ were recorded, whereas only those with Benjamini-adjusted $p < 0.05$ were considered statistically significant. To assess the influence of locally clustered genes, regional exclusion sensitivity analyses were performed by removing genes from selected genomic regions in both the candidate and background gene sets.

2.8. qPCR-Based Relative Copy-Number Assessment

To assess the relative copy-number profiles of candidate deletion CNVs predicted by CNVpytor, 10 deletion CNVs (S1–S10) were selected from the candidate CNVs and evaluated by real-time quantitative PCR (qPCR) in 20 Zhongwei goats. All 20 animals were selected from the original cohort of 367 Zhongwei goats and were predicted by CNVpytor to carry deletions at all 10 target CNV loci. Genomic sequences corresponding to the chromosomal positions of the target CNVs in the goat ARS1 reference genome were retrieved, and specific primers were designed using NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) [37]. Primers were synthesized by Beijing Xinnuo Jinda Gene Technology Co., Ltd. (Beijing, China). Primer sequences and amplicon information are provided in Supplementary Table S1.

Genomic DNA from Zhongwei goats was used as the template, with *β-actin* as the reference gene. qPCR was performed on a QuantStudio™ 3 System (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) using TB Green® Premix Ex Taq™ II (Takara Biotechnology (Dalian) Co., Ltd., Dalian, China). Each target CNV was assayed in three technical replicates, and the corresponding *β-actin* measurements were used for normalization. The reaction mixture and cycling conditions are provided in Supplemen-

tary Tables S2 and S3. No independently confirmed diploid calibrator samples were included; instead, an internal relative calibration approach was applied. For each technical replicate, ΔC_t was calculated by subtracting the C_t value of *β-actin* from that of the target amplicon. For each of the 10 target CNV loci, 60 ΔC_t values were obtained from 20 goats, with 3 technical replicates per goat. The minimum of these 60 values was defined as a single locus-specific internal calibration value and applied to all measurements at the corresponding locus. The $\Delta\Delta C_t$ value for each technical replicate was calculated by subtracting this calibration value from its ΔC_t, and relative copy number was estimated using $\mathrm{CN} = 2 \times 2^{-\Delta\Delta C_t}$ [38,39]. The three technical-replicate estimates for each sample were summarized as mean ± SD. Following the integer copy-number classification method described by Zhao et al., the mean relative copy number for each sample was rounded to the nearest integer: CN = 0 or 1 was classified as a deletion, CN = 2 as a relative non-deletion category, and CN ≥ 3 as a duplication. A sample–locus combination was considered concordant for the deletion classification when qPCR classified it as a deletion (CN = 0 or 1) and CNVpytor also predicted a deletion [16].

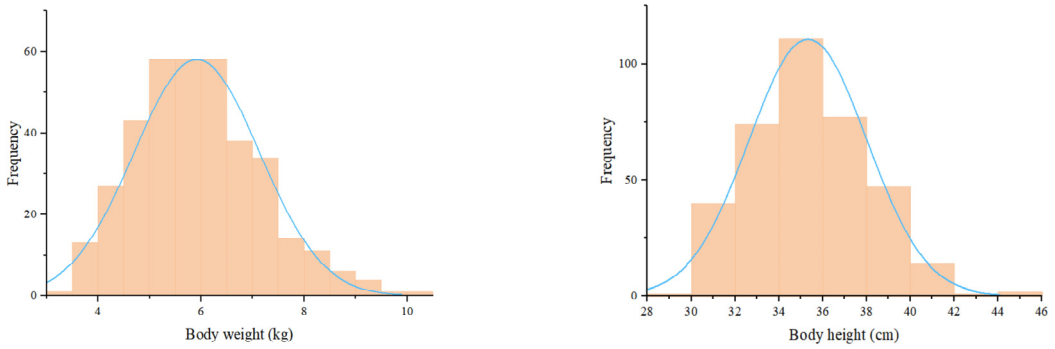
3. Results

3.1. Phenotypic Characteristics and Distributions

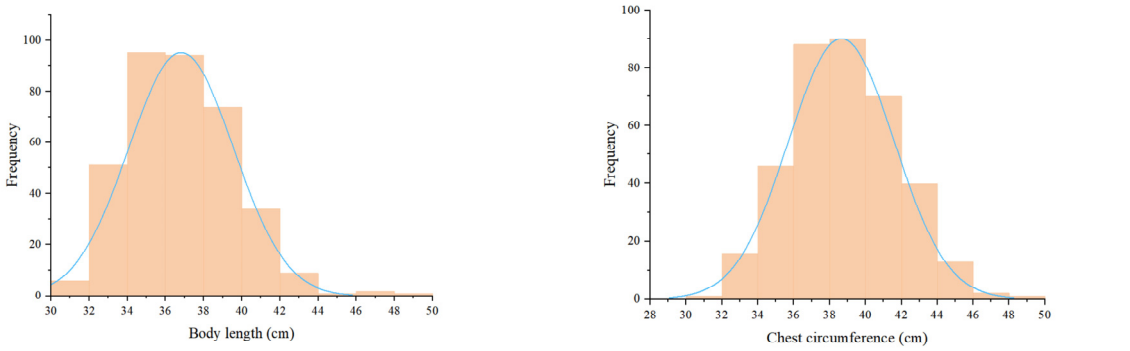
Descriptive statistics for body weight and body measurements in 367 Zhongwei goats are shown in Table 2. The mean ± SD values for BW, BH, BL, and CC were 5.92 ± 1.22 kg, 35.32 ± 2.69 cm, 36.81 ± 2.75 cm, and 38.65 ± 2.95 cm, respectively. BW had the highest coefficient of variation (20.67%), whereas the coefficients of variation for BH, BL, and CC were similar (7.48–7.63%). The phenotypic distributions of all four traits are presented in Figure 1, and Kolmogorov–Smirnov tests indicated no significant departure from normality ($p > 0.05$).

Table 2. Descriptive statistics of early body weight and body measurements.

Trait	N	Mean	SD	Min	Max	CV (%)
Body weight (BW), kg	367	5.92	1.22	3.40	10.00	20.67
Body height (BH), cm	367	35.32	2.69	29.50	45.50	7.62
Body length (BL), cm	367	36.81	2.75	30.30	49.80	7.48
Chest circumference (CC), cm	367	38.65	2.95	30.00	48.00	7.63



(**a**) (**b**)

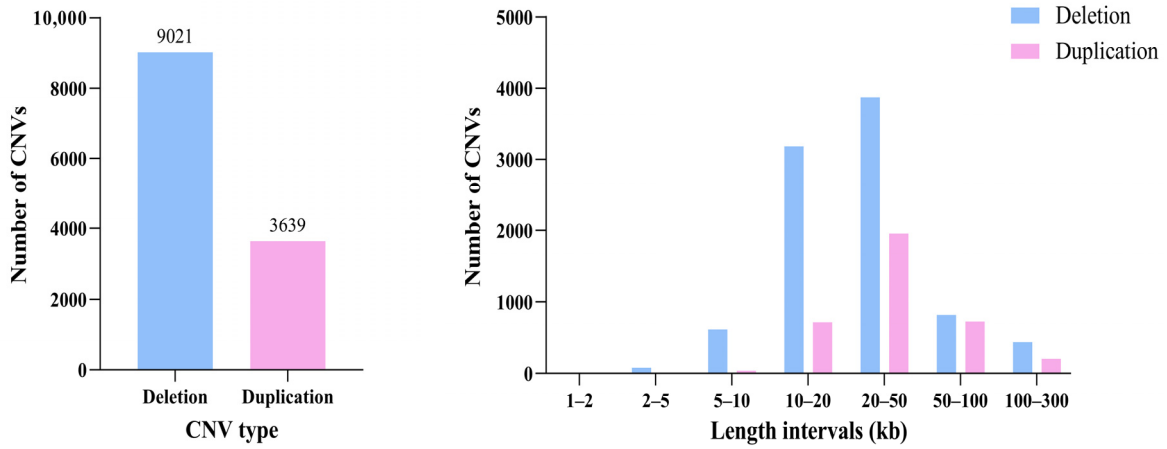


(**c**) (**d**)

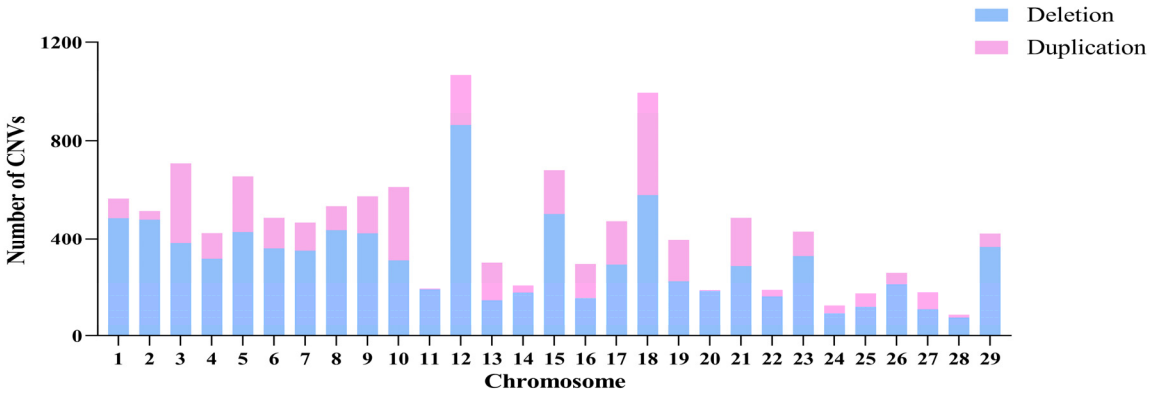
Figure 1. Phenotypic distributions of body weight and three body measurements. (**a**) Body weight (BW, kg). (**b**) Body height (BH, cm). (**c**) Body length (BL, cm). (**d**) Chest circumference (CC, cm). Histograms represent the frequency distributions of the observed phenotypic values, and the solid blue curves represent the corresponding fitted normal distributions.

3.2. CNV Detection and Distribution

After quality control of the whole-genome resequencing data from 367 Zhongwei goats, a total of 25.28 Tb of clean data were retained, averaging 68.89 Gb per individual. CNVpytor calling and quality filtering retained 162,378 individual-level CNV calls across the 367 goats. The merged multi-sample VCF contained 12,660 CNV variant records, including 9021 deletions (71.26%) and 3639 duplications (28.74%), indicating a predominance of deletion records (Figure 2a). The corresponding filtered multi-sample CNV call set is provided in Supplementary File S1. CNV lengths ranged from 3 to 799 kb, with an overall mean length of 37.04 kb; the mean lengths of deletion and duplication CNVs were 34.79 and 42.62 kb, respectively. Most CNVs were concentrated in the 10–50-kb range, with the largest number occurring in the 20–50-kb interval, followed by the 10–20-kb interval (Figure 2b). CNVs were distributed across all 29 autosomes, with marked differences in number and type composition among chromosomes (Figure 2c). Chr12 contained the largest number of CNVs (1067), followed by chr18 (995), whereas chr28 contained the fewest (88). Chr12 also contained the largest number of deletion CNVs, whereas chr18 contained the largest number of duplication CNVs. Detailed chromosome-level CNV counts and length-interval distributions are provided in Supplementary Tables S4 and S5, respectively.



(**a**) (**b**)



(**c**)

Figure 2. Genome-wide distribution and size characteristics of CNVs. (**a**) Composition of deletion and duplication CNVs. (**b**) Distribution of deletion and duplication CNVs across different length intervals. (**c**) Chromosomal distribution of deletion and duplication CNVs across the 29 autosomes. Blue and pink bars represent deletion and duplication CNVs, respectively. For clarity, panel (**b**) displays CNVs with lengths ≤300 kb.

3.3. Genome-Wide Distribution and Characteristics of CNVRs

Application of the predefined CNVR merging and filtering criteria yielded 1176 CNVRs with a cumulative length of 53.34 Mb, corresponding to approximately 2.16% of the total length of the 29 autosomes in the ARS1 reference genome. Among these, 980 were deletion CNVRs (83.33%), whereas 98 were duplication CNVRs and 98 were mixed CNVRs (8.33% each), indicating a predominance of deletion CNVRs (Figure 3a).

CNVRs were distributed across all 29 autosomes (Figure 3b). Chr2 contained the largest number of CNVRs (71) and also the largest number of deletion CNVRs (64). Chr7 contained the largest number of duplication CNVRs (9), whereas chr18 contained the largest number of mixed CNVRs (14). CNVR lengths ranged from 4 to 1257 kb, with a mean length of 45.36 kb; the mean lengths of deletion, duplication, and mixed CNVRs were 30.02, 65.17, and 178.92 kb, respectively. CNVRs were concentrated mainly in the 10–20-kb and 20–50-kb intervals, which together contained 827 CNVRs (70.32% of all CNVRs; Figure 3c). Analysis using non-overlapping 1-Mb windows showed that only 14 windows contained ≥5 CNVRs, with a maximum of 10 CNVRs in a single window (Figure 3d), indicating variation in CNVR density across genomic regions. Chromosome-level CNVR counts and cumulative lengths are provided in Supplementary Table S6, whereas CNVR size distributions are provided in Supplementary Table S7. Overall descriptive statistics for CNVs and CNVRs are summarized in Supplementary Table S8.

3.4. Population Structure Assessment

Principal component analysis (PCA) showed that PC1 and PC2 explained 6.51% and 5.13% of the total variance represented by the PCA relationship matrix, respectively, accounting for a combined 11.65% (Figure 4). In the PC1–PC2 space, most individuals formed a relatively compact cluster; only a few showed modest deviations, and no clearly separated subpopulations were observed. Nevertheless, the first three principal components were retained as covariates in the association analyses to account for residual population structure.

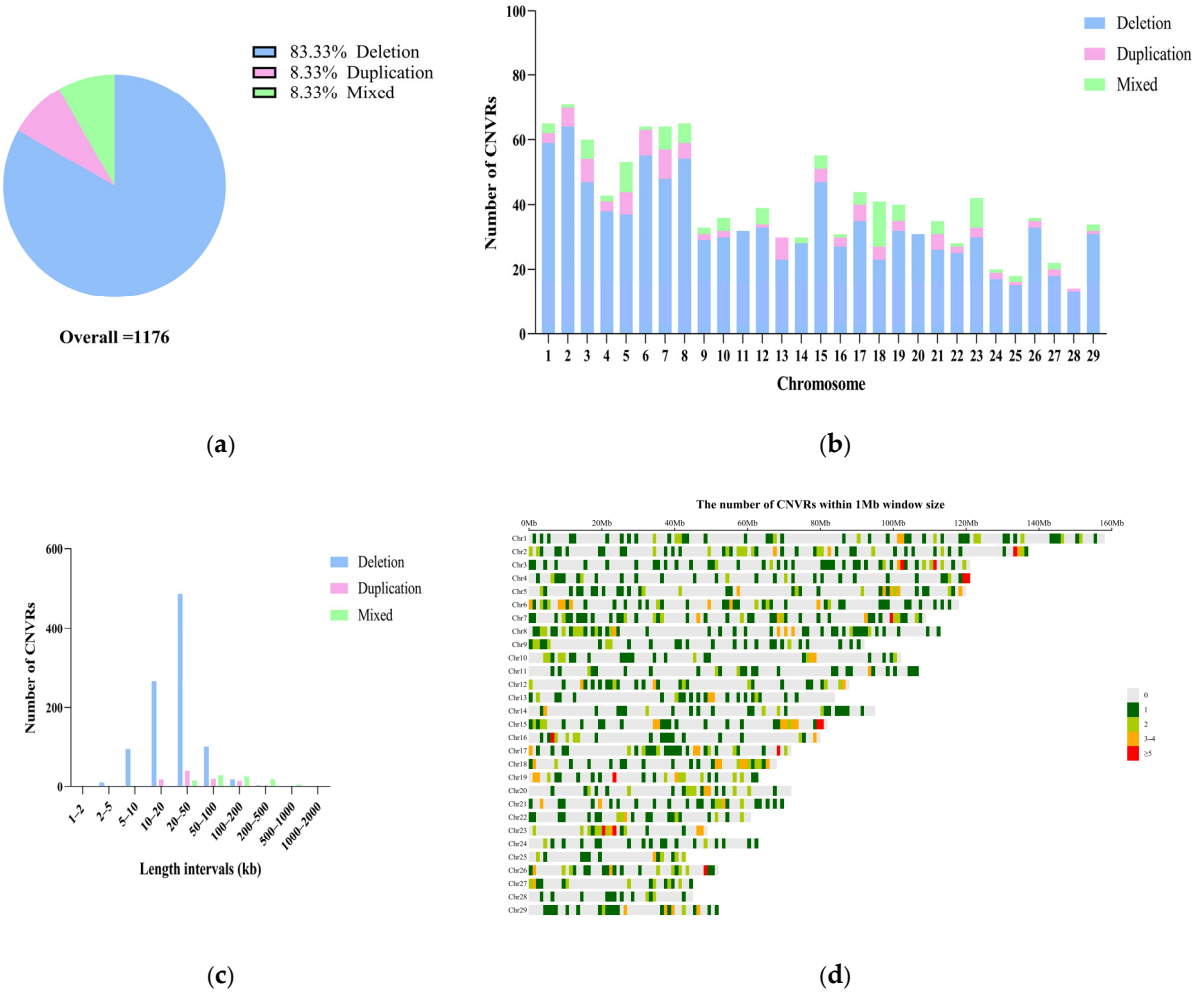


Figure 3. Genome-wide characteristics and distribution of CNVRs. (**a**) Proportions of deletion, duplication, and mixed CNVRs. (**b**) Chromosomal distribution of deletion, duplication, and mixed CNVRs across the 29 autosomes. (**c**) Length distribution of CNVRs by type across predefined size intervals. (**d**) Genome-wide CNVR density across the 29 autosomes in non-overlapping 1-Mb windows. Each CNVR was assigned to a window according to its midpoint, and colors indicate the number of CNVRs per window, categorized as 0, 1, 2, 3–4, and ≥5. Mixed CNVRs contain both deletion and duplication events.

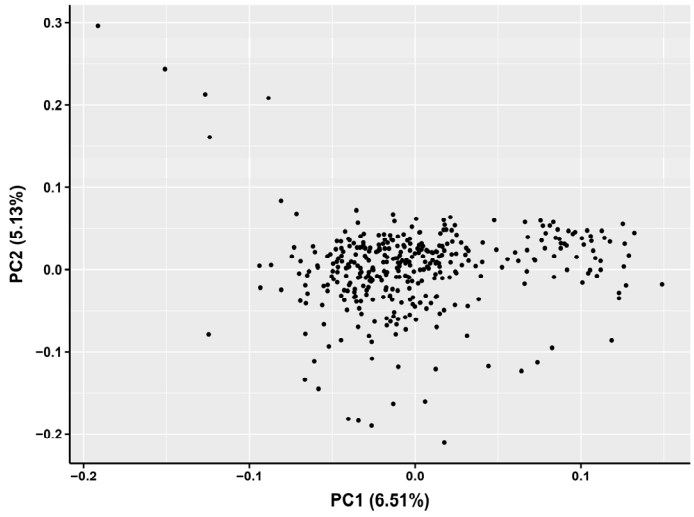


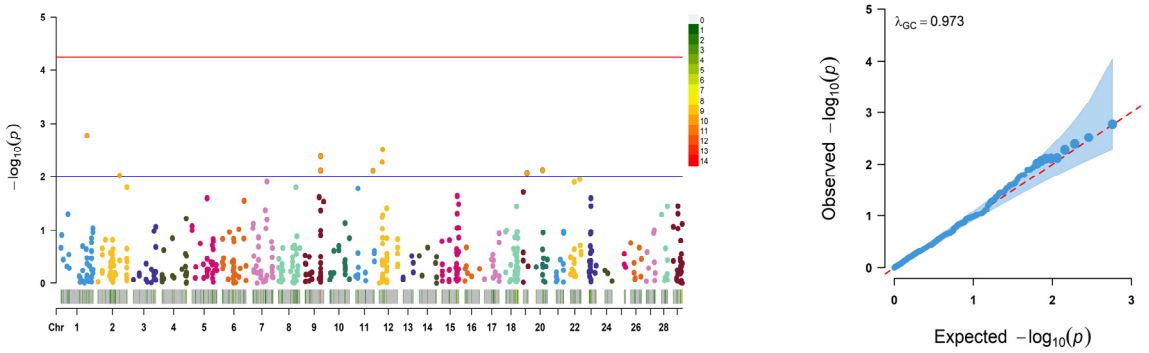
Figure 4. Principal component analysis based on CNV genotypes. Each point represents one individual Zhongwei goat.

3.5. CNV-GWAS of Early Body Weight and Body Measurements

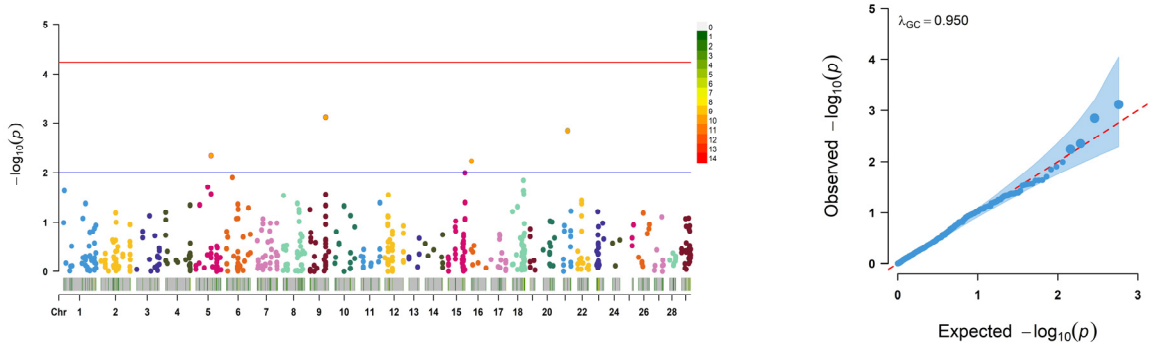
To identify candidate CNVs associated with body weight and body measurements, CNV-GWAS analyses were performed separately for body weight (BW), body height (BH), body length (BL), and chest circumference (CC). Following genotype quality control, 878 CNV markers were retained for each trait, of which 576 yielded valid association statistics per trait, whereas the remaining 302 markers had undefined association statistics. Among these 302 markers, 231 were deletions and 71 were duplications; all had only one observed non-missing genotype class and therefore lacked the genotype variation required for association testing. All 576 markers with valid association statistics were deletion-type CNVs. The genotype matrix comprising all 878 retained CNV markers is provided in Supplementary File S2. Although the fastGWA-mlm option was specified, the estimated genetic variance component was not statistically significant for any of the four traits; GCTA therefore automatically performed the association tests using linear regression. Manhattan and Q–Q plots for each trait are shown in Figure 5. The genomic inflation factors (λ_{GC}) for BW, BH, BL, and CC were 0.973, 0.950, 0.927, and 0.985, respectively, indicating no substantial overall inflation of the test statistics. Representative analysis scripts and key computational parameters used in this study are provided in Supplementary File S3.

In the primary CNV-GWAS analysis, the association between the CNV at chr26:1,003,001–1,019,000 and body length (BL) met the predefined per-trait Bonferroni significance threshold ($p = 3.30413 \times 10^{-5}$; threshold: $p < 0.05/878 = 5.69 \times 10^{-5}$). This CNV also remained significant after within-trait Benjamini–Hochberg false discovery rate (FDR) correction ($q = 0.0290$), but not after FDR correction across all four traits ($q = 0.1160$). In the sensitivity analysis restricted to markers with valid association p values, the corresponding q values were 0.0190 within BL and 0.0761 across all four traits, yielding the same overall conclusion.

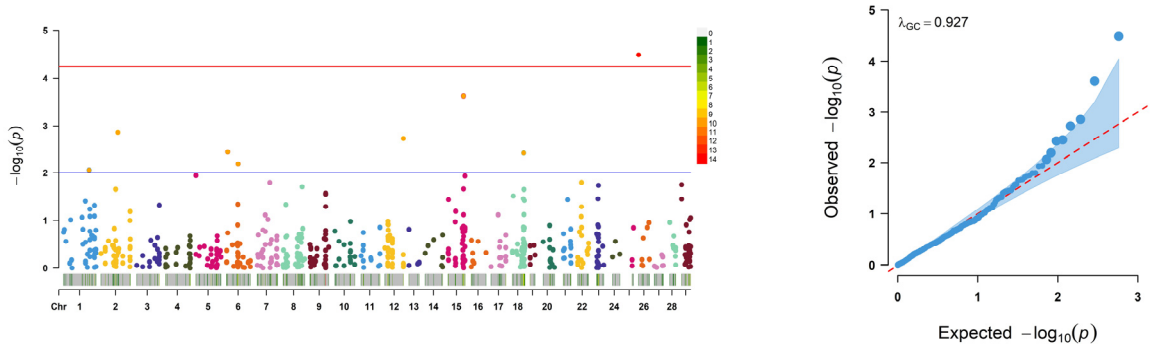
At the exploratory screening threshold of $p < 0.01$, 28 candidate CNV–trait association signals were identified, comprising 9 for BW, 4 for BH, 8 for BL, and 7 for CC. All of these signals were deletion CNVs (Table 3). The lowest association p value for each trait was observed for chr1:123,179,001–123,213,000 for BW ($p = 0.00169232$), chr9:74,505,001–74,548,000 for BH ($p = 0.000768266$), chr26:1,003,001–1,019,000 for BL ($p = 3.30413 \times 10^{-5}$), and chr29:26,217,001–26,251,000 for CC ($p = 0.000733386$). The estimated proportion of phenotypic variance explained (PVE) by the 28 candidate signals ranged from 4.59% to 23.08%. The marker-specific effective sample sizes (N) ranged from 38 to 153 animals. Among the 28 signals, 24 were trait-specific, whereas the remaining 4 signals formed 2 cross-trait overlapping regions. The chr9 candidate regions for BW and BH overlapped at chr9:74,505,001–74,548,000, whereas chr6:8,787,001–8,813,000 was identified as a candidate region for both BL and CC.



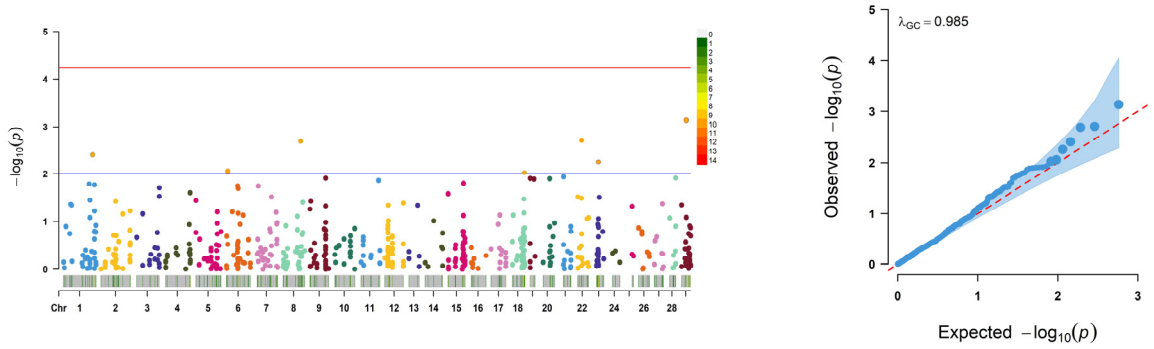
(**a**)



(**b**)



(**c**)



(**d**)

Figure 5. CNV-GWAS of body weight and three body measurements. (**a**) Body weight (BW). (**b**) Body height (BH). (**c**) Body length (BL). (**d**) Chest circumference (CC). In each panel, the Manhattan plot is shown on the left and the corresponding quantile–quantile (Q–Q) plot on the right. One BL-associated CNV (chr26:1,003,001–1,019,000) reached the per-trait Bonferroni-corrected significance (p = 3.30413 × 10^{-5}). In each Manhattan plot, the red solid line indicates the Bonferroni-corrected threshold (p = 0.05/878 = 5.69 × 10^{-5}), whereas the blue solid line indicates the exploratory candidate-association threshold (p = 0.01). In the Manhattan plots, point colors distinguish chromosomes, whereas the color scale in the upper-right corner indicates the number of CNV markers per genomic window in the density track below the x-axis, rather than association p values. Genomic inflation factors (λ_{GC}) are shown in the corresponding Q–Q plots. In the Q–Q plots, blue points represent observed association statistics, the red dashed line indicates the expected distribution under the null hypothesis, and the light-blue shaded area denotes the confidence interval.

3.6. Functional Annotation and Enrichment Analysis of Candidate Genes

To characterize the potential biological functions of the candidate CNVs, genes overlapping the candidate CNV intervals or located within 500 kb on either side of these intervals were annotated. A total of 322 gene–CNV interval association records were obtained. After gene-level deduplication, 287 nonredundant candidate genes were retained, distributed across 17 autosomes. According to the DAVID Gene List Report, 275 of the 287 submitted genes were successfully mapped, whereas the remaining 12 were not mapped. Based on their positional relationships with candidate CNVs, evidence from previous studies linking these genes to body-size-related phenotypes, and the functional enrichment results, 26 representative candidate genes were selected for presentation in Table 4.

For enrichment analysis, 3641 nonredundant genes annotated within 500 kb on either side of 576 CNV markers with valid association *p*-values were used as the primary background gene set. All 287 candidate genes were included in this background set. GO enrichment analysis identified 36 nominally significant terms ($p < 0.05$), of which 20 remained significant after Benjamini multiple-testing correction (adjusted $p < 0.05$), including 8 biological process, 6 cellular component, and 6 molecular function terms. KEGG enrichment analysis identified 11 nominally significant pathways ($p < 0.05$), of which 6 remained significant after correction (adjusted $p < 0.05$). The enriched GO terms were primarily associated with amino acid transport, transmembrane transport, lipid and lipoprotein metabolism, and ATP-dependent transport (Figure 6a). The significant KEGG pathways were mainly related to transport processes, lipid metabolism, immune regulation, and cell signaling (Figure 6b). Using the alternative background comprising 4799 genes, 25 of the 26 enrichment categories (GO terms and KEGG pathways) that were significant under the primary background remained significant after Benjamini correction (Supplementary File S4). Regional-exclusion sensitivity analyses showed that the enrichment results were sensitive to locally clustered gene families. After jointly excluding five selected genomic regions, none of the original 26 significant GO terms and KEGG pathways remained significant after Benjamini correction. Four additional GO terms and five additional KEGG pathways became significant in the combined-exclusion analysis; however, all nine categories included the same four MHC class II-related genes on chr23, indicating that the remaining enrichment results were also influenced by local gene clustering.

Table 3. Exploratory CNV association signals for early body weight and body measurements.

Trait	CNV region	Length (bp)	CNV type	N	CN0(*n*)	CN1(*n*)	Deletion Allele Frequency	β_{DEL}	SE	*p* Value	Within-Trait FDR *q* (878)	Four-Trait FDR *q* (3512)	Estimated PVE (%)
BW	chr1:123,179,001–123,213,000	34,000	Deletion	52	19	33	0.682692	−1.0171	0.323969	1.692×10^{-3}	0.925771	0.720996	15.93
	chr2:100,850,001–100,878,000	28,000	Deletion	83	3	80	0.518072	−1.71969	0.664362	9.640×10^{-3}	0.925771	1.000000	7.47
	chr9:74,254,001–74,391,000	137,000	Deletion	55	37	18	0.836364	−0.929154	0.323987	4.132×10^{-3}	0.925771	0.967514	13.01
	chr9:74,495,001–74,548,000	53,000	Deletion	43	18	25	0.709302	−0.929837	0.349034	7.721×10^{-3}	0.925771	1.000000	14.17
	chr11:87,186,001–87,208,000	22,000	Deletion	46	43	3	0.9673913	−1.79375	0.674273	7.808×10^{-3}	0.925771	1.000000	13.33
	chr12:14,253,001–14,587,000	334,000	Deletion	64	34	30	0.765625	0.78752	0.282579	5.321×10^{-3}	0.925771	1.000000	10.82
	chr12:15,813,001–16,209,000	396,000	Deletion	41	8	33	0.597561	1.31355	0.443982	3.091×10^{-3}	0.925771	0.967514	17.59
	chr19:18,897,001–18,978,000	81,000	Deletion	69	2	67	0.514493	2.12899	0.810451	8.616×10^{-3}	0.925771	1.000000	9.09
	chr20:44,983,001–45,000,000	17,000	Deletion	148	4	144	0.513514	1.52764	0.572326	7.604×10^{-3}	0.925771	1.000000	4.59
BH	chr5:73,056,001–73,084,000	28,000	Deletion	142	32	110	0.612676	−1.47054	0.517293	4.473×10^{-3}	1.000000	0.981736	5.38
	chr9:74,505,001–74,548,000	43,000	Deletion	43	30	13	0.848837	−2.86447	0.851513	7.683×10^{-4}	0.622467	0.674538	20.83
	chr16:6,931,001–6,943,000	12,000	Deletion	153	49	104	0.660131	1.2323	0.446546	5.787×10^{-3}	1.000000	1.000000	4.74
	chr21:38,061,001–38,090,000	29,000	Deletion	72	3	69	0.520833	4.83337	1.5147	1.418×10^{-3}	0.622467	0.720996	12.39
BL	chr1:119,231,001–119,251,000	20,000	Deletion	57	22	35	0.692982	−1.89361	0.72189	8.713×10^{-3}	0.956220	1.000000	10.77
	chr2:78,137,001–78,169,000	32,000	Deletion	48	2	46	0.520833	6.09617	1.90797	1.398×10^{-3}	0.409081	0.720996	17.54
	chr6:8,787,001–8,813,000	26,000	Deletion	50	6	44	0.56	3.35553	1.15220	3.588×10^{-3}	0.555658	0.967514	14.50
	chr6:56,594,001–56,635,000	41,000	Deletion	73	7	66	0.547945	2.87146	1.05391	6.438×10^{-3}	0.807542	1.000000	9.23
	chr12:85,969,001–86,000,000	31,000	Deletion	83	17	66	0.60241	2.23364	0.718968	1.892×10^{-3}	0.415230	0.720996	10.42

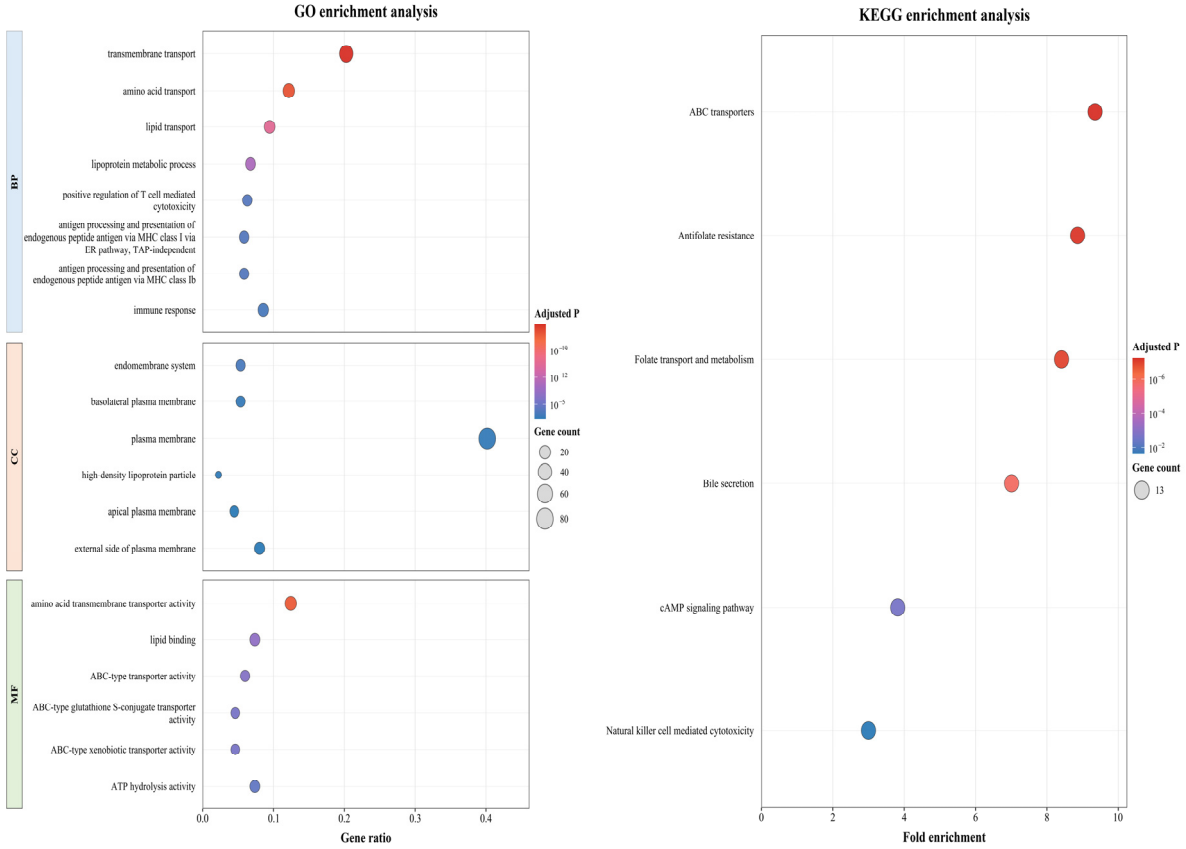
	chr15:73,301,001–73,324,000	23,000	Deletion	79	14	65	0.588608	2.84208	0.774942	2.450×10^{-4}	0.107537	0.430150	14.55
	chr18:59,411,001–59,427,000	16,000	Deletion	80	77	3	0.98125	4.51050	1.55828	3.797×10^{-3}	0.555658	0.967514	9.48
	chr26:1,003,001–1,019,000	16,000	Deletion	59	55	4	0.9661017	−5.62637	1.35529	**3.304×10^{-5}**	0.029010	0.116041	22.61
CC	chr1:135,463,001–135,475,000	12,000	Deletion	39	19	20	0.74359	2.57675	0.893417	3.925×10^{-3}	0.829724	0.967514	17.58
	chr6:8,787,001–8,813,000	26,000	Deletion	50	6	44	0.56	3.18592	1.21603	8.795×10^{-3}	0.829724	1.000000	12.07
	chr8:83,551,001–83,572,000	21,000	Deletion	56	10	46	0.589286	2.99461	0.971498	2.053×10^{-3}	0.600830	0.720996	14.51
	chr18:62,843,001–62,885,000	42,000	Deletion	84	26	58	0.654762	1.71186	0.659592	9.450×10^{-3}	0.829724	1.000000	7.42
	chr22:26,681,001–26,694,000	13,000	Deletion	77	71	6	0.961039	−3.66408	1.18364	1.964×10^{-3}	0.600830	0.720996	11.07
	chr23:23,178,001–23,213,000	35,000	Deletion	129	124	5	0.9806202	3.5334	1.27313	5.514×10^{-3}	0.829724	1.000000	5.63
	chr29:26,217,001–26,251,000	34,000	Deletion	38	18	20	0.736842	3.04699	0.902335	7.334×10^{-4}	0.600830	0.674538	23.08

Note: BW, body weight; BH, body height; BL, body length; CC, chest circumference; CNV, copy number variation; N, marker-specific effective sample size reported by fastGWA; CN0 and CN1, predicted zero-copy (GT 1/1) and one-copy (GT 0/1) states, respectively; β_{DEL}, deletion-allele effect estimate; SE, standard error; PVE, estimated proportion of phenotypic variance explained. The phenotypes were analyzed on their original scales without standardization; β_{DEL} is expressed in kg for BW and cm for BH, BL, and CC. Only candidate CNVs with $p < 0.01$ are shown. For loci with A1 = N and A2 = <DEL>, deletion-allele frequency and β_{DEL} were calculated as 1 − AF1 and −BETA, respectively. Deletion-allele frequencies and PVE estimates were based on marker-specific subsets of animals with non-missing genotypes rather than the entire cohort of 367 goats. Within-trait and four-trait FDR q-values were calculated using the Benjamini–Hochberg procedure with 878 markers per trait and 3512 marker–trait records, respectively; undefined p-values were assigned a value of 1 solely for FDR adjustment. CNV intervals represent the maximal observed span of individual calls sharing the same start coordinate. The bold p value indicates the BL-associated CNV at chr26:1,003,001–1,019,000 that reached the per-trait Bonferroni threshold ($p < 0.05/878 = 5.69 \times 10^{-5}$).

Table 4. Representative candidate genes within or near trait-specific and shared CNV regions associated with early body weight and body measurements.

Trait	Candidate CNV	Candidate Gene(s)	Position Relative to CNV
BW	chr2:100,850,001–100,878,000	*TANK*	Downstream (40,351 bp)
BW	chr9:74,254,001–74,391,000	*LATS1*	Upstream (404,605 bp)
BW	chr11:87,186,001–87,208,000	*KLF11*; *RRM2*; *ODC1*; *ADAM17*	*KLF11*: Upstream (29,988 bp); *RRM2*: Upstream (92,659 bp); *ODC1*: Upstream (387,150 bp); *ADAM17*: Downstream (318,592 bp)
BW	chr12:15,813,001–16,209,000	*ABCC4*	Downstream (188,252 bp)
BW	chr19:18,897,001–18,978,000	*WSB1*; *NF1*	*WSB1*: Upstream (245,490 bp); *NF1*: Upstream (370,310 bp)
BW/BH	BW: chr9:74,495,001–74,548,000; BH: chr9:74,505,001–74,548,000	*LOC106501757* (predicted NKG2D ligand 2-like); *IYD*	*LOC106501757*: Direct overlap with BW CNV; Upstream (5960 bp) of BH CNV; *IYD*: Downstream (261,112 bp) of both CNVs
BH	chr5:73,056,001–73,084,000	*RBFOX2*	Upstream (188,234 bp)
BL	chr1:119,231,001–119,251,000	*AGTR1*	Upstream (121,410 bp)
BL	chr18:59,411,001–59,427,000	*PPP2R1A*	Upstream (296,239 bp)
BL	chr26:1,003,001–1,019,000	*BNIP3*	Downstream (246,125 bp)
BL/CC	chr6:8,787,001–8,813,000	*TRAM1L1*; *LOC108636201*	*TRAM1L1*: Upstream (127,429 bp); *LOC108636201*: Downstream (260,833 bp)
CC	chr8:83,551,001–83,572,000	*IARS*; *OGN*; *OMD*; *ASPN*; *ECM2*	*IARS*: Downstream (29,059 bp); *OGN*: Downstream (217,288 bp); *OMD*: Downstream (242,404 bp); *ASPN*: Downstream (265,250 bp); *ECM2*: Downstream (303,095 bp)
CC	chr23:23,178,001–23,213,000	*ELOVL5*; *TNXB*; *AGPAT1*; *NOTCH4*	*ELOVL5*: Downstream (480,942 bp); *TNXB*: Upstream (489,323 bp); *AGPAT1*: Upstream (438,332 bp); *NOTCH4*: Upstream (397,250 bp)

Note: BW, body weight; BH, body height; BL, body length; CC, chest circumference; CNV, copy number variation. The table includes 26 representative candidate genes, comprising 22 genes from CNV regions identified for individual traits and 4 genes representing two cross-trait shared or overlapping regions. Upstream and downstream refer to the lower- and higher-coordinate sides of a CNV interval, respectively, irrespective of gene orientation. Distances indicate the number of intervening base pairs between CNV and gene boundaries, while direct overlap indicates intersecting genomic intervals. All listed CNV–trait association signals met the candidate-screening threshold of $p < 0.01$; one BL-associated CNV (chr26:1,003,001–1,019,000) additionally reached the per-trait Bonferroni-corrected threshold ($p < 0.05/878 = 5.69 \times 10^{-5}$).



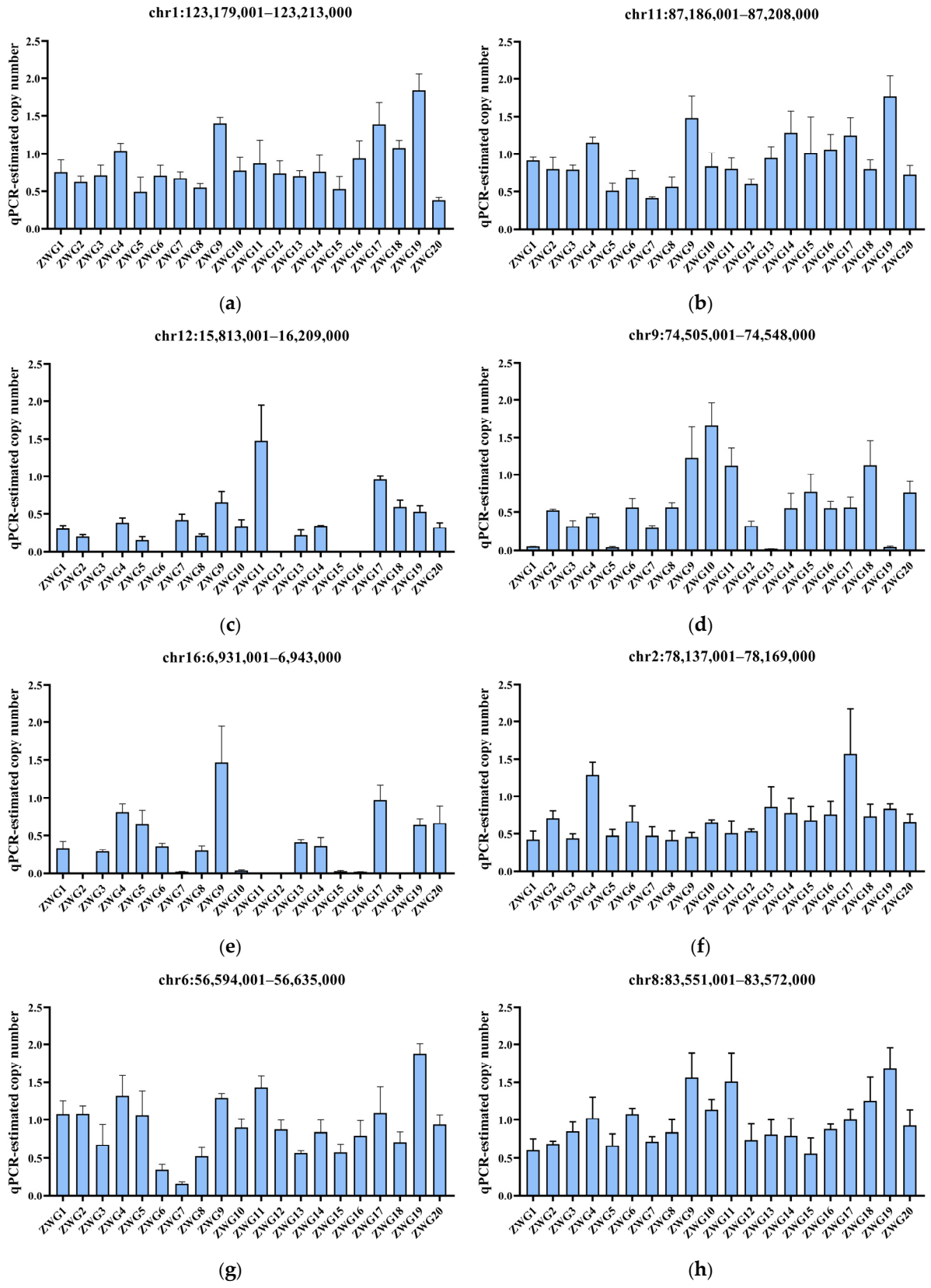
(**a**) (**b**)

Figure 6. Functional enrichment analysis of candidate genes within or near exploratory candidate CNV regions associated with body weight and body measurements in Zhongwei goats. (**a**) Gene Ontology (GO) enrichment analysis showing 20 significant terms, comprising 8 biological process (BP), 6 cellular component (CC), and 6 molecular function (MF) terms. (**b**) Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis showing 6 significant pathways. All displayed terms and pathways met the Benjamini-adjusted significance threshold (adjusted $p < 0.05$). The x-axis shows gene ratio (Count/List Total) in (**a**) and fold enrichment in (**b**). Bubble size represents the number of candidate genes, while bubble color indicates the Benjamini-adjusted p value, with warmer colors indicating lower values. In panel (**a**), CC refers to the GO cellular component category rather than chest circumference.

3.7. qPCR Assessment of Candidate CNVs

Relative copy numbers at 10 candidate deletion CNV loci (S1–S10) were assessed by qPCR in 20 Zhongwei goats predicted by CNVpytor to carry deletions at all 10 loci (Figure 7). Relative copy numbers were estimated using an internal calibration procedure, and the mean estimate from three technical replicates was rounded to the nearest integer for classification. Of the 200 sample–locus combinations, 190 (95.0%) were classified as deletions, including 66 assigned to CN0 and 124 to CN1. The remaining 10 combinations were classified as CN2. Locus-specific deletion classification rates were 100% for S3 and S5, 85% for S8, and 95% for the remaining seven loci. Of the 10 non-deletion classifications, five occurred in ZWG19 and two in ZWG9.

These findings indicate classification agreement with the selected CNVpytor-predicted deletions under the applied internal calibration procedure. However, they do not constitute independent confirmation of absolute copy-number states or establish the overall detection accuracy of CNVpytor (Supplementary File S5).


chr1:123,179,001–123,213,000
chr11:87,186,001–87,208,000
chr12:15,813,001–16,209,000
chr9:74,505,001–74,548,000
chr16:6,931,001–6,943,000
chr2:78,137,001–78,169,000
chr6:56,594,001–56,635,000
chr8:83,551,001–83,572,000
qPCR-estimated copy number
0.0
0.5
1.0
1.5
2.0
2.5
ZWG1
ZWG2
ZWG3
ZWG4
ZWG5
ZWG6
ZWG7
ZWG8
ZWG9
ZWG10
ZWG11
ZWG12
ZWG13
ZWG14
ZWG15
ZWG16
ZWG17
ZWG18
ZWG19
ZWG20

(**a**) (**b**)

(**c**) (**d**)

(**e**) (**f**)

(**g**) (**h**)

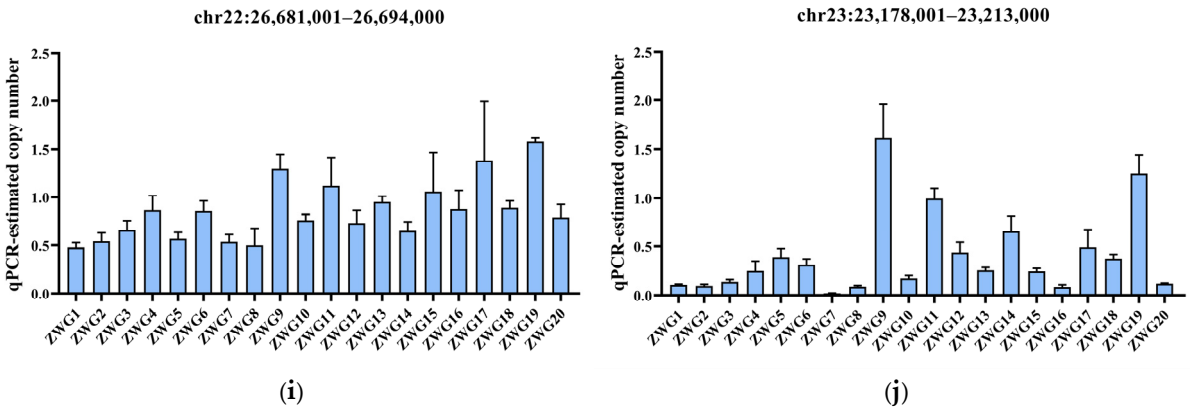


Figure 7. qPCR-based relative copy-number assessment of 10 selected candidate deletion CNVs. Panels (**a–j**) correspond to the ten candidate CNVs (S1–S10), respectively. The x-axis represents the 20 Zhongwei goat samples (ZWG1–ZWG20), and the y-axis represents the qPCR-estimated relative copy number. Bars and error bars represent the mean and standard deviation (SD) of three technical replicates, respectively. Relative copy numbers were estimated using an internal calibration procedure without independently confirmed diploid calibrator samples.

4. Discussion

The construction of CNV/CNVR maps is strongly influenced by the detection platform and analytical strategy used. Early goat CNV studies relied mainly on aCGH and SNP arrays. Fontanesi et al. identified 127 CNVRs in four goat breeds using aCGH [40], whereas Liu et al. identified 978 CNVRs across 50 goat breeds worldwide using the CaprineSNP50 array [41]. To improve the reliability of CNV detection from array data, Guan et al. used both PennCNV and QuantiSNP and integrated their results with EnsembleCNV, ultimately identifying 486 high-confidence CNVRs [42]. With the adoption of whole-genome sequencing (WGS), CNV detection is no longer constrained by predefined probe locations and can exploit more continuous genome-wide coverage. This broader genomic coverage has enabled the detection of numerous CNVs in goat populations, including both short CNVs and CNVs of intermediate length [8,43]. More recently, Bian et al. used long-read assemblies and a graph pangenome to characterize deletions, insertions, and complex structural variants in goats, demonstrating that detection technologies differ substantially in the classes of variants they capture and in their resolution [44].

The predominance of deletion events is a commonly observed pattern in goat CNV studies. Among the high-confidence CNVRs identified by Guan et al., loss events were much more frequent than gain events; Guo et al. likewise observed a high proportion of deletion CNVs, with 96.76% of CNVs ≤50 kb [42,43]. A similar pattern was observed in African goat populations, in which Nandolo et al. detected 253,553 CNVs from whole-genome resequencing data, 96.6% of which were deletions, with a mean length of only 3.3 kb [8]. Recent studies across multiple Chinese goat populations have further supported the predominance of deletion CNVRs [45]. In the present study, both CNVs and CNVRs were predominantly deletions, and most variants fell within shorter size classes, consistent with the general pattern reported in previous goat studies. Although all 28 CNV–trait association signals identified at $p < 0.01$ involved deletions, no duplication markers yielded valid association statistics in the present GWAS; therefore, this pattern does not establish deletion-specific biological effects on the measured traits. Nevertheless, studies differ in the numbers of CNVs/CNVRs, length distributions, type compositions, and genomic coverage, likely because of differences in population composition, detection platform, reference genome, analytical algorithm, filtering criteria, and CNVR merging rules. The criterion of at least 1 bp overlap used for CNVR construction may also have merged

distinct CNV events into larger regions, particularly mixed CNVRs; therefore, their boundaries should be interpreted cautiously. We used the read-depth (RD) workflow implemented in CNVpytor for CNV detection [22]. Although this approach can detect copy-number dosage changes from sequencing depth, short-read data provide limited resolution for precise breakpoint definition and for detecting some complex structural variants. Future studies could therefore combine multiple structural-variant signals or long-read sequencing to validate and resolve key candidate regions at higher resolution.

In CNV-GWAS, stringent multiple-testing correction is used to determine genome-wide significance, whereas a less stringent exploratory threshold can be used to prioritize potential candidate signals. Stylianou et al. retained the Bonferroni genome-wide significance criterion while also using $p < 0.01$ to screen additional candidate CNV loci [30]. Similarly, in a bovine CNV-GWAS, Bernini et al. used $p < 0.01$ to screen potential association signals given the low frequencies and complex genetic architecture of CNVs [29]. Accordingly, we used Bonferroni correction as the formal criterion for genome-wide significance and $p < 0.01$ as an exploratory candidate-screening threshold. The association findings included one BL-associated CNV reaching the per-trait Bonferroni-corrected significance threshold, alongside 27 additional exploratory candidate signals. Across the four traits, 2304 marker–trait tests yielded valid association p values. Under the global null hypothesis with appropriately calibrated p values, approximately 23 associations would be expected to reach the nominal threshold of $p < 0.01$ by chance, compared with the 28 observed signals. This observed count was relatively close to the null expectation, suggesting that the number of nominal associations alone should not be taken as evidence of a substantial enrichment of associations. Nevertheless, these candidate signals may provide valuable exploratory leads for independent validation and further functional investigation. Because animals without valid CNV genotype calls were treated as missing, the association tests primarily compared CN0 and CN1 states among animals with available calls, rather than deletion carriers and confirmed diploid non-carriers across the entire cohort of 367 goats. Given the incomplete CNV genotype data and small genotype groups at some loci, the estimated effects and proportions of phenotypic variance explained (PVE) should be interpreted cautiously, as they may be unstable or overestimated [46]. These estimates were based on marker-specific subsets of animals with valid genotypes and should not be directly extrapolated to the entire cohort. The chr26 BL-associated signal, based on only four CN1 animals, warrants particular caution because its effect estimate may be sensitive to individual observations. Although PC1–PC3 were included as covariates, PC2 and PC3 were associated with sequencing batch, and batch-adjusted GWAS sensitivity analyses were not performed; therefore, residual batch-related confounding cannot be excluded. Of the 28 CNV–trait association signals, 24 reached the exploratory threshold for only one trait, whereas the remaining 4 corresponded to two shared or overlapping CNV regions associated with BW/BH and BL/CC, respectively. Similar cross-trait associations have been reported in goats. Liu et al. found that CNV27 was associated with both chest width and ischial width in African indigenous goats [14], and two CNVs associated with both preweaning average daily gain and weaning weight were identified in Inner Mongolian cashmere goats [15]. Thus, the BW/BH overlapping region on chr9 and the BL/CC shared region on chr6 identified here warrant further validation. However, without replication in an independent population or functional evidence, these regions should not yet be considered confirmed pleiotropic loci.

A total of 287 nonredundant candidate genes were annotated within or near the candidate CNV regions. Considering genomic location, functional annotations, enrichment results, and previous studies, *ODC1*, *RBFOX2*, *ECM2*, *OGN*, *OMD*, *ASPN*, *TNXB*, and *ELOVL5* were prioritized as biologically plausible candidate genes. Other candidates, including *PPP2R1A*, were also considered based on their reported biological functions.

These genes may provide biological context for the candidate CNV regions associated with the measured traits. Because the ±500 kb annotation window was not defined using population-specific linkage disequilibrium decay, physical proximity alone does not establish regulatory relationships between candidate CNVs and nearby genes.

Among the candidate genes with reported functions relevant to muscle development, *ODC1* encodes ornithine decarboxylase, which catalyzes the rate-limiting step in polyamine biosynthesis. Its expression in skeletal muscle is regulated by the androgen receptor, and *ODC1* has been implicated in myoblast proliferation [47]. *RBFOX2* regulates alternative splicing during myogenesis. By modulating the splicing of genes such as *MEF2D* and *ROCK2*, *RBFOX2* influences myoblast fusion and terminal differentiation, supporting a role in muscle fiber formation [48]. *ECM2* has also been implicated in muscle development and participates in the differentiation of bovine skeletal muscle-derived satellite cells [49]. In the present study, *ODC1*, *RBFOX2*, and *ECM2* were located near candidate CNVs associated with BW, BH, and CC, respectively. Their reported functions provide a biological rationale for investigating these candidate CNV regions in relation to early muscle development and body measurements in Zhongwei goats, although their contributions to the observed associations remain to be established.

Several other candidate genes identified in this study have reported functions related to extracellular matrix organization and skeletal development. Near the CC-associated candidate CNV on chr8, *OGN*, *OMD*, and *ASPN* have functions relevant to these processes. *OGN* participates in collagen fibrillogenesis and extracellular matrix organization [50], whereas ASPN can inhibit TGF-β-mediated chondrogenic differentiation [51,52]. Of particular interest, *OMD* was previously identified as a candidate gene for body weight at different ages in dairy cattle [53]. Functional studies further showed that *OMD* interacts with BMP2 to enhance BMP/SMAD signaling and osteogenic differentiation [54], while animal models have implicated it in bone and cartilage remodeling [55]. These findings suggest that extracellular matrix organization and skeletal development warrant further investigation in relation to the CC-associated candidate region. Another CC-associated candidate CNV on chr23 was located near *TNXB*, which encodes an extracellular matrix protein. An amino acid substitution in *TNXB* has been associated with mature body weight in several sheep breeds [56]. For BL, *PPP2R1A* was located near a candidate CNV on chr18. In mice, conditional deletion of *PPP2R1A* in leptin receptor-positive mesenchymal stem cells was shown to impair endochondral ossification involving RUNX2 signaling [57]. These findings provide functional evidence supporting the further investigation of *PPP2R1A* in the BL-associated genomic region.

The CC-associated candidate CNV on chr23 was also located near genes involved in lipid metabolism and adipocyte development, including *ELOVL5*, *AGPAT1*, and *NOTCH4*. Among these, *ELOVL5* is particularly noteworthy because of functional and genetic evidence from small ruminants. *ELOVL5* encodes a fatty acid elongase involved in long-chain unsaturated fatty acid synthesis. Shi et al. showed that altered *ELOVL5* expression affected the fatty acid composition, particularly C16:1n-7 and C18:1n-7, in goat mammary epithelial cells [58]. In sheep, an *ELOVL5* polymorphism was associated with tail width, tail fat weight, and relative tail fat weight, and its expression was higher in the tail adipose tissue of large-tailed than small-tailed sheep [59]. Together, these findings support a potential role for *ELOVL5* in lipid metabolism and fat deposition in small ruminants. The nearby gene *AGPAT1* participates in glycerolipid biosynthesis, and its expression was positively correlated with intramuscular fat content in bovine longissimus dorsi muscle [60]. *NOTCH4* has also been implicated in adipocyte development. Lai et al. demonstrated that activated NOTCH4 promoted proliferation and adipogenic differentiation in 3T3-L1 preadipocytes, accompanied by changes in ERK signaling and adipogenic regulators, including PPARγ and C/EBPα [61]. In addition, *KLF11* was identified near a

BW-associated candidate CNV on chr11. In goat intramuscular preadipocytes, *KLF11* knockdown increased lipid droplet accumulation and the expression of PPARγ and C/EBPβ while reducing Pref-1 expression, suggesting a negative regulatory role in adipogenic differentiation [62]. Collectively, these studies provide a functional basis for investigating lipid metabolism and adipocyte development in relation to the corresponding candidate CNV regions, without establishing direct effects of the CNVs on fat deposition or the measured body traits.

The two cross-trait CNV association regions identified in this study also contained genes of potential interest. In the overlapping BW/BH region on chr9, *IYD* encodes iodotyrosine deiodinase, an enzyme involved in iodide recycling during thyroid hormone synthesis. Because thyroid hormones are important for longitudinal skeletal growth, endochondral ossification, and skeletal maturation [63], *IYD* represents a biologically plausible candidate for further investigation. Fonseca et al. identified *IYD* within a candidate regulatory network potentially involved in the pleiotropic regulation of production and fertility traits in beef cattle [64], providing additional comparative genetic evidence. The shared BL/CC region on chr6 was located near *TRAM1L1*. In a genome-wide association study of 304,782 Nellore cattle, Ogunbawo et al. identified a genomic window containing *TRAM1L1* among candidate regions associated with expected progeny differences for body weight at 120 days of age [65]. These findings support further examination of the two cross-trait candidate regions in Zhongwei goats. However, the overlapping associations observed across traits do not establish pleiotropic effects of the nearby genes. Independent replication and functional studies are needed to determine whether these genes contribute to associations with more than one of the measured traits.

qPCR has been used to assess computationally predicted CNVs in sheep and goats [11,13,43]. In the present study, 10 candidate deletion CNVs were examined in 20 Zhongwei goats selected on the basis of CNVpytor-predicted deletions at the corresponding loci. Using an internal relative calibration and integer-based classification procedure, 190 of the 200 sample–locus combinations (95.0%) were classified as deletions, whereas the remaining 10 were assigned to CN2. Discrepancies between methods may arise because CNVpytor infers interval-level copy-number changes from WGS read-depth signals, whereas qPCR estimates the local relative copy number of specific amplicons. Differences in CNV boundaries, local sequence characteristics, read-depth variability, and qPCR measurement error may also contribute to discordant classifications [22,41]. Importantly, no independently confirmed diploid calibrator samples were available, and the relative copy-number classifications may therefore depend on the internal calibration procedure. Accordingly, the observed 95.0% agreement describes the classification outcomes of the preselected deletion-predicted sample–locus combinations under the applied procedure. It should not be interpreted as an estimate of overall genome-wide CNV detection accuracy or as independent validation of the corresponding CNV–trait associations.

Given the limited marker-specific effective sample sizes and uncertainties in CNV genotyping, the candidate CNV–trait associations identified in this study, including the BL-associated signal that met the per-trait Bonferroni threshold, require replication in larger, independent Zhongwei goat populations. Although PC1–PC3 were included as covariates, PC2 and PC3 were associated with sequencing batch, and batch-adjusted GWAS sensitivity analyses were not performed; therefore, residual batch-related confounding cannot be excluded. In addition, incomplete individual-level records of dam parity, family relationships, and management groups limited our ability to account for potential maternal, familial, and management-related confounding. Future studies should prioritize independent verification of copy-number states and more reliable CNV genotyping, particularly for the BL-associated CNV on chr26 and the overlapping BW/BH and

BL/CC regions on chr9 and chr6, respectively. Gene-expression analyses across independently verified copy-number states, together with functional experiments, may help clarify the potential contributions of these candidate CNVs and nearby genes to early-life body weight and body measurements.

5. Conclusions

Genome-wide CNV profiling and CNV-GWAS in Zhongwei goats identified 28 candidate CNV–trait association signals across four measured traits. The BL-associated chr26 signal showed sensitivity to individual observations in auxiliary leave-one-out analyses and is therefore interpreted as an exploratory candidate association. No signal remained significant after FDR correction across all four traits. Given substantial CNV genotype missingness and small effective sample sizes at some loci, these associations warrant cautious interpretation. Based on genomic locations, functional annotations, enrichment results, and previous studies, *ODC1*, *RBFOX2*, *ECM2*, *OGN*, *OMD*, *ASPN*, *TNXB*, and *ELOVL5* were prioritized as biologically plausible candidate genes with reported functions related to muscle and skeletal development, extracellular matrix organization, and lipid metabolism. These findings provide a genome-wide characterization of CNVs in Zhongwei goats and identify candidate genomic regions and genes for independent replication and further functional investigation.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/vetsci13111264/s1. Table S1: Primer sequences used for qPCR assessment of candidate CNVs; Table S2: Reaction mixture for qPCR assessment of candidate CNVs; Table S3: Thermal cycling conditions for qPCR assessment of candidate CNVs; Table S4: Distribution of CNVs across the 29 autosomes in Zhongwei goats; Table S5: Distribution of CNVs across different length intervals in Zhongwei goats; Table S6: Chromosomal distribution and length of CNVRs in Zhongwei goats; Table S7: Distribution of CNVRs across different length intervals in Zhongwei goats; Table S8: Descriptive statistics of CNVs and CNVRs in Zhongwei goats; Supplementary File S1: Filtered multi-sample CNV call set comprising 12,660 CNV variant records from 367 Zhongwei goats; Supplementary File S2: Genotype matrix comprising 878 CNV markers used for CNV-GWAS in 367 Zhongwei goats; Supplementary File S3: Analysis scripts and computational parameters for CNV-GWAS in Zhongwei goats; Supplementary File S4: Detailed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis results for candidate genes, including regional-exclusion sensitivity analyses; Supplementary File S5: qPCR data analysis results for the assessment of candidate deletion CNVs.

Author Contributions: Conceptualization, L.Q., Y.Z. and R.W.; methodology, L.Q., M.H., Y.R. and R.W.; formal analysis, L.Q., M.H. and Y.R.; investigation, Z.Z., H.D., D.L., Y.Y. and B.Z.; data curation, L.Q. and M.H.; writing—original draft preparation, L.Q.; writing—review and editing, L.Q., M.H., Y.R., X.W., X.A., F.S., Y.Z. and R.W.; visualization, L.Q., X.W. and X.A.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Research and Development Program of Ningxia Hui Autonomous Region (grant number 2026BBF02019) and the Conservation and Innovative Utilization of Sheep and Goat Genetic Resources (grant number BR261201).

Institutional Review Board Statement: All experimental procedures, including sample collection, were conducted in accordance with the International Guiding Principles for Biomedical Research Involving Animals and the biomedical research ethics regulations of Inner Mongolia Agricultural University. The study was approved by the Special Committee on Scientific Research and Academic Ethics of Inner Mongolia Agricultural University (Approval No. (2020) 056; approval date: 6 May 2020). No additional specific permits were required for sample collection, and no endangered or protected species were involved.

Informed Consent Statement: Informed consent for animal sampling was obtained from the management of the Zhongwei Goat Breeding Farm.

Data Availability Statement: The whole-genome resequencing data from all 367 Zhongwei goats analyzed in this study are publicly available in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1121220 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1121220; accessed on 20 April 2026). Summary results are presented in the article and its Supplementary Materials. Representative analysis scripts and key computational parameters are provided in Supplementary File S3. The filtered multi-sample CNV call set containing 12,660 CNV variant records and the genotype matrix comprising 878 CNV markers used for association analysis are provided in Supplementary Files S1 and S2, respectively. Individual-level phenotypic and associated covariate data are not publicly available due to ethical and informed-consent considerations. These data may be made available by the corresponding author upon reasonable request, subject to applicable ethical and informed-consent requirements.

Acknowledgments: The authors thank the staff of the Zhongwei Goat Breeding Farm (Zhongwei, China) for their assistance with sample collection.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

BW	Body weight
BH	Body height
BL	Body length
CC	Chest circumference
CNV	Copy number variation
CNVR	Copy number variation region
SNP	Single nucleotide polymorphism
GWAS	Genome-wide association study
WGS	Whole-genome sequencing
qPCR	Quantitative PCR
aCGH	Array comparative genomic hybridization
FDR	False discovery rate
PCA	Principal component analysis
GRM	Genomic relationship matrix
PVE	Phenotypic variance explained
ZWG	Zhongwei goats

References

1. Lu, C.D. The role of goats in the world: Society, science, and sustainability. *Small Rumin. Res.* **2023**, *227*, 107056. https://doi.org/10.1016/j.smallrumres.2023.107056.
2. Yang, R.; Zhou, D.; Tan, X.; Zhao, Z.; Lv, Y.; Tian, X.; Ren, L.; Wang, Y.; Li, J.; Zhao, Y.; et al. Genome-Wide Association Study of Body Conformation Traits in Tashi Goats (*Capra hircus*). *Animals* **2024**, *14*, 1145. https://doi.org/10.3390/ani14081145.
3. Zhang, L.; Wang, F.; Gao, G.; Yan, X.; Liu, H.; Liu, Z.; Wang, Z.; He, L.; Lv, Q.; Wang, Z.; et al. Genome-Wide Association Study of Body Weight Traits in Inner Mongolia Cashmere Goats. *Front. Vet. Sci.* **2021**, *8*, 752746. https://doi.org/10.3389/fvets.2021.752746.
4. Liu, Y.; Ding, Y.; Liu, Z.; Chen, Q.; Li, X.; Xue, X.; Pu, Y.; Ma, Y.; Zhao, Q. Integration Analysis of Transcriptome and Proteome Reveal the Mechanisms of Goat Wool Bending. *Front. Cell Dev. Biol.* **2022**, *10*, 836913. https://doi.org/10.3389/fcell.2022.836913.
5. Xiao, P.; Zhong, T.; Liu, Z.; Ding, Y.; Guan, W.; He, X.; Pu, Y.; Jiang, L.; Ma, Y.; Zhao, Q. Integrated Analysis of Methylome and Transcriptome Changes Reveals the Underlying Regulatory Signatures Driving Curly Wool Transformation in Chinese Zhongwei Goats. *Front. Genet.* **2020**, *10*, 1263. https://doi.org/10.3389/fgene.2019.01263.

6. Han, M.; Wang, X.; Du, H.; Cao, Y.; Zhao, Z.; Niu, S.; Bao, X.; Rong, Y.; Ao, X.; Guo, F.; et al. Genome-wide association study identifies candidate genes affecting body conformation traits of Zhongwei goat. *BMC Genom.* **2025**, *26*, 37. https://doi.org/10.1186/s12864-024-11097-1.
7. Chen, Y.; Khan, M.Z.; Wang, X.; Liang, H.; Ren, W.; Kou, X.; Liu, X.; Chen, W.; Peng, Y.; Wang, C. Structural variations in livestock genomes and their associations with phenotypic traits: A review. *Front. Vet. Sci.* **2024**, *11*, 1416220. https://doi.org/10.3389/fvets.2024.1416220.
8. Nandolo, W.; Meszaros, G.; Wurzinger, M.; Banda, L.J.; Gondwe, T.N.; Mulindwa, H.A.; Nakimbugwe, H.N.; Clark, E.L.; Woodward-Greene, M.J.; Liu, M.; et al. Detection of copy number variants in African goats using whole genome sequence data. *BMC Genom.* **2021**, *22*, 398. https://doi.org/10.1186/s12864-021-07703-1.
9. McCarroll, S.A.; Altshuler, D.M. Copy-number variation and association studies of human disease. *Nat. Genet.* **2007**, *39*, S37–42. https://doi.org/10.1038/ng2080.
10. Rafter, P.; Gormley, I.C.; Purfield, D.; Parnell, A.C.; Naderi, S.; Berry, D.P. Genome-wide association analyses of carcass traits using copy number variants and raw intensity values of single nucleotide polymorphisms in cattle. *BMC Genom.* **2021**, *22*, 757. https://doi.org/10.1186/s12864-021-08075-2.
11. Yuan, C.; Lu, Z.; Guo, T.; Yue, Y.; Wang, X.; Wang, T.; Zhang, Y.; Hou, F.; Niu, C.; Sun, X.; et al. A global analysis of CNVs in Chinese indigenous fine-wool sheep populations using whole-genome resequencing. *BMC Genom.* **2021**, *22*, 78. https://doi.org/10.1186/s12864-021-07387-7.
12. Gordeeva, V.; Sharova, E.; Arapidi, G. Progress in Methods for Copy Number Variation Profiling. *Int. J. Mol. Sci.* **2022**, *23*, 2143. https://doi.org/10.3390/ijms23042143.
13. Wang, Z.; Guo, J.; Guo, Y.; Yang, Y.; Teng, T.; Yu, Q.; Wang, T.; Zhou, M.; Zhu, Q.; Wang, W.; et al. Genome-Wide Detection of CNVs and Association With Body Weight in Sheep Based on 600K SNP Arrays. *Front. Genet.* **2020**, *11*, 558. https://doi.org/10.3389/fgene.2020.00558.
14. Liu, M.; Woodward-Greene, J.; Kang, X.; Pan, M.G.; Rosen, B.; Van Tassell, C.P.; Chen, H.; Liu, G.E. Genome-wide CNV analysis revealed variants associated with growth traits in African indigenous goats. *Genomics* **2020**, *112*, 1477–1480. https://doi.org/10.1016/j.ygeno.2019.08.018.
15. Liu, Y.; Xi, H.; Xu, Q.; Zhou, B.; Li, J.; Su, R.; Lv, Q.; Zhang, Y.; Wang, R.; Wang, Z. Genome-wide association study of copy number variation and early growth traits in inner Mongolian cashmere goats. *Front. Vet. Sci.* **2025**, *12*, 1651622. https://doi.org/10.3389/fvets.2025.1651622.
16. Zhao, J.; Liu, Z.; Wang, X.; Xin, X.; Du, L.; Zhao, H.; An, Q.; Ding, X.; Zhang, Z.; Wang, E.; et al. The Identification of Goat *KCNJ15* Gene Copy Number Variation and Its Association with Growth Traits. *Genes* **2024**, *15*, 250. https://doi.org/10.3390/genes15020250.
17. Qiu, Y.; Ding, R.; Zhuang, Z.; Wu, J.; Yang, M.; Zhou, S.; Ye, Y.; Geng, Q.; Xu, Z.; Huang, S.; et al. Genome-wide detection of CNV regions and their potential association with growth and fatness traits in Duroc pigs. *BMC Genom.* **2021**, *22*, 332. https://doi.org/10.1186/s12864-021-07654-7.
18. Zhou, Y.; Utsunomiya, Y.T.; Xu, L.; Hay el, H.A.; Bickhart, D.M.; Alexandre, P.A.; Rosen, B.D.; Schroeder, S.G.; Carvalheiro, R.; de Rezende Neves, H.H.; et al. Genome-wide CNV analysis reveals variants associated with growth traits in *Bos indicus*. *BMC Genom.* **2016**, *17*, 419. https://doi.org/10.1186/s12864-016-2461-4.
19. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
20. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
21. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
22. Suvakov, M.; Panda, A.; Diesh, C.; Holmes, I.; Abyzov, A. CNVpytor: A tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. *GigaScience* **2021**, *10*, giab074. https://doi.org/10.1093/gigascience/giab074.
23. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
24. Davoudi, P.; Do, D.N.; Rathgeber, B.; Colombo, S.M.; Sargolzaei, M.; Plastow, G.; Wang, Z.; Karimi, K.; Hu, G.; Valipour, S.; et al. Genome-wide detection of copy number variation in American mink using whole-genome sequencing. *BMC Genom.* **2022**, *23*, 649. https://doi.org/10.1186/s12864-022-08874-1.

25. 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. https://doi.org/10.1186/s13742-015-0047-8.
26. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
27. Jiang, L.; Zheng, Z.; Qi, T.; Kemper, K.E.; Wray, N.R.; Visscher, P.M.; Yang, J. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **2019**, *51*, 1749–1755. https://doi.org/10.1038/s41588-019-0530-8.
28. Shim, H.; Chasman, D.I.; Smith, J.D.; Mora, S.; Ridker, P.M.; Nickerson, D.A.; Krauss, R.M.; Stephens, M. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE* **2015**, *10*, e0120758. https://doi.org/10.1371/journal.pone.0120758.
29. Bernini, F.; Boccardo, A.; Delledonne, A.; Bronzo, V.; Lanfredi, G.; Bagnato, A.; Strillacci, M.G. Bovine respiratory disease-associated ultrasonographic lung lesions in Holstein calves: A genomic perspective on copy number variants and homozygosity. *Front. Anim. Sci.* **2025**, *6*, 1700819. https://doi.org/10.3389/fanim.2025.1700819.
30. Stylianou, C.E.; Wiggins, G.A.R.; Lau, V.L.; Dennis, J.; Shelling, A.N.; Wilson, M.; Sykes, P.; Amant, F.; Annibali, D.; De Wispelaere, W.; et al. Germline copy number variants and endometrial cancer risk. *Hum. Genet.* **2024**, *143*, 1481–1498. https://doi.org/10.1007/s00439-024-02707-9.
31. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
32. Devlin, B.; Roeder, K. Genomic Control for Association Studies. *Biometrics* **1999**, *55*, 997–1004. https://doi.org/10.1111/j.0006-341X.1999.00997.x.
33. Bickhart, D.M.; Rosen, B.D.; Koren, S.; Sayre, B.L.; Hastie, A.R.; Chan, S.; Lee, J.; Lam, E.T.; Liachko, I.; Sullivan, S.T.; et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **2017**, *49*, 643–650. https://doi.org/10.1038/ng.3802.
34. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
35. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. https://doi.org/10.1038/nprot.2008.211.
36. 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. https://doi.org/10.1093/nar/gkac194.
37. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T.L. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **2012**, *13*, 134. https://doi.org/10.1186/1471-2105-13-134.
38. D'Haene, B.; Vandesompele, J.; Hellemans, J. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* **2010**, *50*, 262–270. https://doi.org/10.1016/j.ymeth.2009.12.007.
39. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* **2001**, *25*, 402–408. https://doi.org/10.1006/meth.2001.1262.
40. Fontanesi, L.; Martelli, P.L.; Beretti, F.; Riggio, V.; Dall'Olio, S.; Colombo, M.; Casadio, R.; Russo, V.; Portolano, B. An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genom.* **2010**, *11*, 639. https://doi.org/10.1186/1471-2164-11-639.
41. Liu, M.; Zhou, Y.; Rosen, B.D.; Van Tassell, C.P.; Stella, A.; Tosser-Klopp, G.; Rupp, R.; Palhiere, I.; Colli, L.; Sayre, B.; et al. Diversity of copy number variation in the worldwide goat population. *Heredity* **2019**, *122*, 636–646. https://doi.org/10.1038/s41437-018-0150-6.
42. Guan, D.; Martinez, A.; Castello, A.; Landi, V.; Luigi-Sierra, M.G.; Fernandez-Alvarez, J.; Cabrera, B.; Delgado, J.V.; Such, X.; Jordana, J.; et al. A genome-wide analysis of copy number variation in Murciano-Granadina goats. *Genet. Sel. Evol.* **2020**, *52*, 44. https://doi.org/10.1186/s12711-020-00564-4.
43. Guo, J.; Zhong, J.; Liu, G.E.; Yang, L.; Li, L.; Chen, G.; Song, T.; Zhang, H. Identification and population genetic analyses of copy number variations in six domestic goat breeds and Bezoar ibexes using next-generation sequencing. *BMC Genom.* **2020**, *21*, 840. https://doi.org/10.1186/s12864-020-07267-6.
44. Bian, P.; Li, J.; Zhou, S.; Wang, X.; Gong, M.; Guo, X.; Cai, Y.; Yang, Q.; Fu, J.; Li, R.; et al. A Graph-based Goat Pangenome Reveals Structural Variations Involved in Domestication and Adaptation. *Mol. Biol. Evol.* **2024**, *41*, msae251. https://doi.org/10.1093/molbev/msae251.

45. Li, W.; Su, Y.; Liu, C.; Yan, X.; Lv, Q.; Su, R. Identification of the Genetic Characteristics of Copy Number Variation Regions in Diverse Goat Populations. *Genes* **2026**, *17*, 627. https://doi.org/10.3390/genes17060627.
46. Forde, A.; Hemani, G.; Ferguson, J. Review and further developments in statistical corrections for Winner's Curse in genetic association studies. *PLoS Genet.* **2023**, *19*, e1010546. https://doi.org/10.1371/journal.pgen.1010546.
47. Lee, N.K.; Skinner, J.P.; Zajac, J.D.; MacLean, H.E. Ornithine decarboxylase is upregulated by the androgen receptor in skeletal muscle and regulates myoblast proliferation. *Am. J. Physiol. Endocrinol. Metab.* **2011**, *301*, E172–179. https://doi.org/10.1152/ajpendo.00094.2011.
48. Singh, R.K.; Xia, Z.; Bland, C.S.; Kalsotra, A.; Scavuzzo, M.A.; Curk, T.; Ule, J.; Li, W.; Cooper, T.A. Rbfox2-coordinated alternative splicing of *Mef2d* and *Rock2* controls myoblast fusion during myogenesis. *Mol. Cell* **2014**, *55*, 592–603. https://doi.org/10.1016/j.molcel.2014.06.035.
49. Liu, C.; Tong, H.; Li, S.; Yan, Y. Effect of *ECM2* expression on bovine skeletal muscle-derived satellite cell differentiation. *Cell Biol. Int.* **2018**, *42*, 525–532. https://doi.org/10.1002/cbin.10927.
50. Ge, G.; Seo, N.S.; Liang, X.; Hopkins, D.R.; Hook, M.; Greenspan, D.S. Bone morphogenetic protein-1/tolloid-related metalloproteinases process osteoglycin and enhance its ability to regulate collagen fibrillogenesis. *J. Biol. Chem.* **2004**, *279*, 41626–41633. https://doi.org/10.1074/jbc.M406630200.
51. Kizawa, H.; Kou, I.; Iida, A.; Sudo, A.; Miyamoto, Y.; Fukuda, A.; Mabuchi, A.; Kotani, A.; Kawakami, A.; Yamamoto, S.; et al. An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis. *Nat. Genet.* **2005**, *37*, 138–144. https://doi.org/10.1038/ng1496.
52. Kou, I.; Nakajima, M.; Ikegawa, S. Expression and regulation of the osteoarthritis-associated protein asporin. *J. Biol. Chem.* **2007**, *282*, 32193–32199. https://doi.org/10.1074/jbc.M706262200.
53. Yin, T.; Konig, S. Genome-wide associations and detection of potential candidate genes for direct genetic and maternal genetic effects influencing dairy cattle body weight at different ages. *Genet. Sel. Evol.* **2019**, *51*, 4. https://doi.org/10.1186/s12711-018-0444-4.
54. Lin, W.; Zhu, X.; Gao, L.; Mao, M.; Gao, D.; Huang, Z. Osteomodulin positively regulates osteogenesis through interaction with BMP2. *Cell Death Dis.* **2021**, *12*, 147. https://doi.org/10.1038/s41419-021-03404-5.
55. Zappia, J.; Tong, Q.; Van der Cruyssen, R.; Cornelis, F.M.F.; Lambert, C.; Pinto Coelho, T.; Grisart, J.; Kague, E.; Lories, R.J.; Muller, M.; et al. Osteomodulin downregulation is associated with osteoarthritis development. *Bone Res.* **2023**, *11*, 49. https://doi.org/10.1038/s41413-023-00286-5.
56. Cinar, M.U.; Mousel, M.R.; Herndon, M.K.; Taylor, J.B.; White, S.N. Tenascin-XB (TNXB) amino acid substitution E2004G is associated with mature weight and milk score in American Rambouillet, Targhee, Polypay, and Suffolk sheep. *Small Rumin. Res.* **2018**, *166*, 129–133. https://doi.org/10.1016/j.smallrumres.2018.06.013.
57. Yen, Y.T.; Chien, M.; Wu, P.Y.; Hung, S.C. PP2A in LepR+ mesenchymal stem cells contributes to embryonic and postnatal endochondral ossification through Runx2 dephosphorylation. *Commun. Biol.* **2021**, *4*, 658. https://doi.org/10.1038/s42003-021-02175-1.
58. Shi, H.B.; Du, Y.; Zhang, C.H.; Sun, C.; He, Y.L.; Wu, Y.H.; Liu, J.X.; Luo, J.; Loor, J.J. Fatty acid elongase 5 (ELOVL5) alters the synthesis of long-chain unsaturated fatty acids in goat mammary epithelial cells. *J. Dairy. Sci.* **2018**, *101*, 4586–4594. https://doi.org/10.3168/jds.2017-14061.
59. Zhao, L.; Li, F.; Liu, T.; Yuan, L.; Zhang, X.; Zhang, D.; Li, X.; Zhang, Y.; Zhao, Y.; Song, Q.; et al. Ovine *ELOVL5* and *FASN* genes polymorphisms and their correlations with sheep tail fat deposition. *Gene* **2022**, *807*, 145954. https://doi.org/10.1016/j.gene.2021.145954.
60. Jeong, J.; Kwon, E.G.; Im, S.K.; Seo, K.S.; Baik, M. Expression of fat deposition and fat removal genes is associated with intramuscular fat content in longissimus dorsi muscle of Korean cattle steers. *J. Anim. Sci.* **2012**, *90*, 2044–2053. https://doi.org/10.2527/jas.2011-4753.
61. Lai, P.Y.; Tsai, C.B.; Tseng, M.J. Active form Notch4 promotes the proliferation and differentiation of 3T3-L1 preadipocytes. *Biochem. Biophys. Res. Commun.* **2013**, *430*, 1132–1139. https://doi.org/10.1016/j.bbrc.2012.12.024.
62. Wang, J.; Wang, Y.; Chi, Y.; Zhu, J.; Lin, Y. KLF11 Inhibits the Differentiation of Goat Intramuscular Preadipocyte. *Acta Vet. Zootech. Sin.* **2020**, *51*, 976–986. https://doi.org/10.11843/j.issn.0366-6964.2020.05.009.
63. Williams, G.R. Thyroid hormone actions in cartilage and bone. *Eur. Thyroid. J.* **2013**, *2*, 3–13. https://doi.org/10.1159/000345548.

64. Fonseca, P.A.S.; Id-Lahoucine, S.; Reverter, A.; Medrano, J.F.; Fortes, M.S.; Casellas, J.; Miglior, F.; Brito, L.; Carvalho, M.R.S.; Schenkel, F.S.; et al. Combining multi-OMICs information to identify key-regulator genes for pleiotropic effect on fertility and production traits in beef cattle. *PLoS ONE* **2018**, *13*, e0205295. https://doi.org/10.1371/journal.pone.0205295.
65. Ogunbawo, A.R.; Hidalgo, J.; Mulim, H.A.; Carrara, E.R.; Ventura, H.T.; Souza, N.O.; Lourenco, D.; Oliveira, H.R. Applying the algorithm for Proven and young in GWAS Reveals high polygenicity for key traits in Nellore cattle. *Front. Genet.* **2025**, *16*, 1549284. https://doi.org/10.3389/fgene.2025.1549284.

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.