

Article

Not peer-reviewed version

---

# Genome-Wide Association Study of Copy Number Variation and Early Growth Traits in Inner Mongolian Cashmere Goats

---

[Yifan Liu](#) , [Haijiao Xi](#) , [Qi Xu](#) , Bohan Zhou , Jinqian Li , Rui Su , [Qi Lv](#) , Yanjun Zhang , Ruijun Wang , [Zhiying Wang](#) \*

Posted Date: 3 June 2025

doi: 10.20944/preprints202505.2485.v1

Keywords: copy number variation; CNV-based GWAS; early growth traits; inner Mongolia Cashmere Goat



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

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

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

Article

# Genome-Wide Association Study of Copy Number Variation and Early Growth Traits in Inner Mongolian Cashmere Goats

Yifan Liu <sup>1,2,3</sup>, Haijiao Xi <sup>1,2,3</sup>, Qi Xu <sup>1,2,3</sup>, Bohan Zhou <sup>1,2,3</sup>, Jinquan Li <sup>2,3</sup>, Rui Su <sup>1,2,3</sup>, Qi Lv <sup>1,2,3</sup>, Yanjun Zhang <sup>1,2,3</sup>, Ruijun Wang <sup>1,2,3</sup> and Zhiying Wang <sup>1,2,3,\*</sup>

<sup>1</sup> College of Animal Science, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia Autonomous Region, 010018, China

<sup>2</sup> Inner Mongolia Key Laboratory of Sheep & Goat Genetics Breeding and Reproduction, Hohhot, Inner Mongolia Autonomous Region, 010018, China

<sup>3</sup> Key Laboratory Of Mutton Sheep & Goat Genetics And Breeding, Ministry of Agriculture And Rural Affairs, Hohhot, Inner Mongolia Autonomous Region, 010018, China

\* Correspondence: wzhy0321@126.com; Tel.: +86 13474737060

**Simple Summary:** It is vitally important that scientists are able to describe their work simply and concisely to the public, especially in an open-access on-line journal. The simple summary consists of no more than 200 words in one paragraph and contains a clear statement of the problem addressed, the aims and objectives, pertinent results, conclusions from the study and how they will be valuable to society. This should be written for a lay audience, i.e., no technical terms without explanations. No references are cited and no abbreviations. Submissions without a simple summary will be returned directly. Example could be found at <http://www.mdpi.com/2076-2615/6/6/40/htm>.

**Abstract:** In goat breeding programs, early growth traits including birth weight (BW), weaning weight (WW), pre-weaning average daily gain (ADG), and yearling weight (YW) are crucial productivity indicators that directly influence growth rates and economic returns. However, the genetic mechanism of these traits in Inner Mongolia Cashmere goats (IMCGs) has not yet been elucidated. Copy number variation (CNV), as a prevalent form of genomic structural variation and a significant contributor to genetic diversity, has emerged as a valuable molecular marker for complex trait analysis. In this study, we used the genome-wide resequencing data of 461 Inner Mongolia cashmere goats for the first time to detect autosomal CNVs, and performed genome-wide association analysis (GWAS) based on CNVs for early growth traits (BW, WW, ADG, YW) of IMCGs. Results: We detected 26,003 non-redundant CNVs and 5,014 non-redundant CNVRs, covering a total of 1,015.4 Mb (38.97%) of the goat autosomal genome. GWAS analysis revealed 11 CNVs significantly associated with early growth traits, including two pleiotropic CNVs simultaneously influencing ADG and WW, suggesting their regulatory roles in growth and fat deposition processes. Through integrated bioinformatics analysis, we identified seven key candidate genes (ZN845, SOX15, FGF11, GPS2, DVL2, SPRY4, and STAT2) associated with early growth traits. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses demonstrated that these genes are primarily involved in biological pathways related to cell proliferation, differentiation, and protein phosphorylation. The identified CNVs were further validated through PCR verification. In addition, t-test was performed on the phenotypes of individuals with significant CNVs and all the IMCGs populations. Among the 11 significant CNVs, 9 were demonstrated to show significant associations with individual phenotypes, indicating the reliability and credibility of the study findings. This study significantly expands the genomic CNV map of IMCGs through large-scale genotyping. Our findings demonstrate the utility of CNV-based GWAS analysis in elucidating the genetic mechanisms underlying complex traits, providing valuable insights for molecular marker-assisted breeding and molecular genetic research of economically important traits in cashmere goats.

**Keywords:** copy number variation ; CNV-based GWAS ; early growth traits ; inner Mongolia Cashmere Goat

---

## 1. Introduction

Genome-wide association analysis (GWAS) was first put forward by Risch and Merikangas in 1996 during their research on the genetics of complex human diseases. This innovative approach allows for a more accurate genome-wide exploration and discovery of key genes linked to diseases or traits [1]. The superiority of GWAS is manifested in its capability to analyze the associations between genotypes and phenotypes. By conducting association analyses between a large number of genetic markers, such as single nucleotide polymorphisms (SNPs), copy number variations (CNVs), etc., and the phenotypic records of animals, relevant genes that affect complex traits or important quantitative economic traits can be identified across the entire animal genome [2–5]. Previously, the exploration of genes significantly associated with important economic traits in cashmere goats was mainly conducted through GWAS based on SNPs. This is the first time that GWAS analysis of the early growth traits of Inner Mongolia cashmere goats has been carried out based on CNVs.

Copy number variation generally refers to the gain or loss of genomic segments ranging from 50 base pairs to 5 megabases in length [6]. Multiple adjacent and overlapping CNVs can be combined into a copy number variation region (CNVR). Although both CNVs and SNPs can serve as genetic markers in GWAS analysis, compared with SNPs, CNVs can regulate more than five times the number of base pairs, and cover more functional genes [7].

Some studies have demonstrated that CNVs can exert significant impacts on the complex phenotypes and disease resistance of animal individuals by regulating the expression levels of functional genes related to relevant traits, and play a crucial role in the evolutionary process [8]. Xu et al. detected CNVs using the aCGH method and subsequently conducted GWAS. They discovered that the CNV of the *MYH3* gene in Nanyang cattle and Qinchuan cattle has a significant impact on skeletal muscle development [9]. Liang et al. found that the CNV of *SERPINA3-1* gene and *GAL3ST1* gene in Qinchuan cattle also affected bovine muscle development [10]. Ke et al. observed that when the CNV type of the *PLEC* gene on chromosome 14 of Leizhou black goats is in the amplified state, the chest circumference, body weight, carcass weight, cross-sectional area of the longissimus dorsi muscle, and shear force of individuals with the amplified CNV are superior to those of individuals with other CNV types [11]. Wang et al. identified a 2800-bp CNV in the *ORMDL1* gene of sheep, which has a significant effect on the body height, chest circumference, body weight, withers height, and cannon circumference of sheep ( $P < 0.05$ ) [12]. Feng et al. found that the CNV of *PIGY* gene on chromosome 6 of sheep had a significant effect on the body weight, chest circumference and cannon circumference of sheep ( $P < 0.05$ ), and the amplified type of CNV had a positive impact on the body weight of sheep [13]. Qiu et al. discovered that the CNV of *PDGFA*, *GPER1* and other genes in American Duroc pigs could affect the weight traits, such as daily weight gain and slaughter weight [14].

Inner Mongolia cashmere goat is a characteristic species of Inner Mongolia Autonomous Region in China, enjoying a high reputation worldwide. The cashmere they produce is hailed as "soft gold", and the mutton they yield is praised as "ginseng in meat", serving as an important source of economic income for local herders. Currently, there are numerous relevant studies on GWAS using CNV as a genetic marker. However, there has been no related research on conducting GWAS analysis of the early growth traits of Inner Mongolia cashmere goats based on CNV. In this study, based on the 10X re-sequencing data of 461 Inner Mongolia cashmere goats, the phenotypic data of their early growth traits, and the systematic environmental effect data, a mixed linear model was established to carry out GWAS analysis. The aim is to detect significant CNVs and related functional genes that affect the early growth traits of Inner Mongolia cashmere goats, providing a theoretical basis for the genetic improvement of the production performance of Inner Mongolia cashmere goats.

## 2. Materials and Methods

### 2.1. Ethical Approvale

In the experiment, the breeding environment complied with the relevant standards for general animal experimental facilities in the Chinese national standard "Laboratory Animal Environment and Facilities" (GB 14925 - 2023). The feeding and experimental operations of the animals met the requirements of animal welfare. In this study, no anesthesia or euthanasia was performed on experimental animals.

### 2.2. Phenotypic Measurement and Sample Preparation

The Inner Mongolia cashmere goats (n=461) used in this study were all obtained from Inner Mongolia Yiwei White Cashmere Goat Co., Ltd. All the goats were reared under consistent feeding environments and nutritional conditions, being provided with the same commercial diet and having unrestricted access to water. All Inner Mongolia cashmere goats used in this study were born from 2010 to 2018, and the birth weight and date of birth were recorded in detail. Before weaning, some nutritional supplements were provided in addition to the ewe's milk. The lambs were weaned when they were 4 months old. Birth weight (BW), weaning weight (WW) and weight at 12 months (YW) were measured using electronic scales in the same environment. Additionally, the weaning date was recorded. The average daily gain before weaning (ADG) is calculated by subtracting the birth weight from the weaning weight and then dividing the result by the number of days until weaning. BW was measured 0.5 hours after birth, and WW and YW were measured 12 hours after fasting [15]. Ear tissue samples were carefully collected using ear-notch forceps. Immediately after collection, they were swiftly transferred into pre-prepared cryotubes filled with 75% alcohol. Subsequently, these cryotubes were stored at -80 °C until DNA extraction was carried out.

### 2.3. Whole Genome Resequencing

Total Genomic DNA was extracted from ear tissue using the traditional phenol- chloroform method, and the DNA concentration was 50 ng /  $\mu$ L. The quality of DNA in all samples ( 461 DNA samples ) was evaluated based on light absorption ratios ( A260 / 280 and A260 / 230 ) and gel electrophoresis [16]. Following DNA extraction, qualified samples were fragmented using a Covaris crusher. The DNA fragments were then end-repaired, polyA-tailed, ligated with sequencing adapters, purified, and PCR-amplified to construct the sequencing library. The library was preliminarily quantified using Qubit 3.0, and the insert size distribution was verified using an Agilent 5300 system. After passing quality control, the library was subjected to paired-end sequencing (PE150) on the DNBSEQ-T7 platform.

### 2.4. Copy Number Variation Segmentation and Genotyping

The raw data were disconnected, filtered and subjected to various quality control steps to obtain clean data for subsequent bioinformatics analysis. In order to ensure the accuracy of the analysis, FastQC v0.11.5 was used to perform strict quality inspection on the original reading length [17]. The original readings were then compared with the *Capra hircus* reference genome ( Genome assembly ASM4082201v1, [https://www.ncbi.nlm.nih.gov/datasets/genome/GCA\\_040822015.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_040822015.1/)) using the Burrows-Wheeler Aligner ( BWA aln ) v0.7.8 with default parameters [18]. On average, 99% of the reads were successfully mapped to the reference genome, achieving a final average sequencing coverage of 9.46 $\times$  (ranging from 7 $\times$  to 12 $\times$ ) per individual. The initial BAM files, containing sequence alignment data for each individual, were generated using BWA and subsequently sorted and indexed using SAMtools v1.0 [19]. Using BAM files from all individuals, genome features such as GC content, repeat and gap content, read count and absolute copy number were calculated using the sliding window method ( 1 kb window, 800 bp step ) [20]. For the initial CNVnator output, quality control was performed by filtering copy number variations (CNVs) based on length (50 bp to 5 Mb). Plink

v1.90 beta was further used for filtering to remove CNVs with minor allele frequency (MAF < 0.01). For CNVs after quality control, the HandyCNV package in R v4.4.1 was used to merge CNVs with an overlap of more than 1 bp at the same position in all samples to determine CNV regions (CNVRs). HandyCNV defines CNVR as three types: amplification, deletion and mixing (when deletion and amplification occur in the same region) [21].

### 2.5. Genome-wide Association Study

The GWAS analysis was conducted using the Plink v1.90 software. A general linear model was used to perform the GWAS analysis based on CNV for early growth traits in IMCGs. The formula for this model is as follows:

$$y = X\alpha + Z\beta + e \quad (1)$$

Where  $y$  is the phenotype vector,  $\alpha$  is the fixed effect vector, the fixed effects included herds, sex, maternal age, and year of measurement.  $X$  is the structure matrix of fixed effect.  $\beta$  is CNV effect,  $Z$  is the structure matrix of CNV effect;  $e$  is the residual effect, and the distribution is  $e \sim N(0, \sigma^2)$ .

In the CNV-based GWAS, the Bonferroni method was used to determine the genome-wide significant (0.05/N) threshold, where  $N$  represents the number of CNVs. Given that is a stringent criterion, a more lenient threshold was also used for detecting the suggestive (1/N) CNVs [22,23]. The qqman package in R software was used to plot the Manhattan and Q-Q plots [24].

### 2.6. The Copy Number Variation and Phenotypic Comparison Between Individuals with Significant CNVs and IMCGs were Verified by qPCR

Eight candidate CNVs associated with early growth traits in IMCGs were selected for validation, including two for BW traits (CNV\_DEL\_17406 and CNV\_DEL\_18821), three for ADG and WW traits (CNV\_DEL\_11189, CNV\_DEL\_17895, and CNV\_DUP\_18956), and three for YW traits (CNV\_DEL\_4359, CNV\_DEL\_4552, and CNV\_DUP\_20170). These CNVs were subsequently verified using real-time quantitative polymerase chain reaction (RT-qPCR). The primers for these eight CNVs were designed using Primer Premier 5 software (Table S1). Following the methodology established by Sonika et al. [25], we selected the ACTB gene as the reference gene because the gene is highly conserved in goats and exists in the reference genome in the form of a single copy. A total of 64 DNA samples were selected for verification, including 32 samples containing the target CNV, and 32 normal samples without copy number variation in the test area were used as controls. The qPCR experiment was conducted using the 2×SG Green qPCR Mix (with ROX) (SINOGENE, Beijing, China). The PCR reaction was performed in a total volume of 15  $\mu$ L, containing 1  $\mu$ L DNA template (50 ng/ $\mu$ L), 0.25  $\mu$ L each of forward and reverse primers (10 pM/ $\mu$ L), 5  $\mu$ L of 2× Blue-SYBR Green mixture, and 6  $\mu$ L of nuclease-free water. The thermal cycling protocol consisted of an initial denaturation at 95°C for 10 min, followed by 40 cycles of amplification (95°C for 20 s, 60°C for 30 s), and a final dissociation curve analysis (95°C for 15 s, 60°C for 30 s, and 95°C for 15 s). All reactions were carried out in triplicate on a 96-well transparent reaction plate, and the average cycle threshold (Ct) values were calculated for subsequent copy number analysis. The relative copy number variation in the target region was determined using the  $2^{(-\Delta\Delta Ct)}$  method, where  $\Delta\Delta Ct$  was calculated as follows: [(mean Ct of target gene in test sample) - (mean Ct of reference gene (GCG) in test sample)] - [(mean Ct of target gene in reference sample) - (mean Ct of GCG in reference sample)]. Based on this calculation, a value approximating 2 was considered normal, while values  $\geq 3$  and  $\leq 1$  indicated copy number gain and loss, respectively.

After using qPCR to verify CNVs, the phenotypic data of individuals with significant CNVs in GWAS analysis and the phenotypic data of IMCGs were tested by t-test to verify the authenticity of GWAS results.

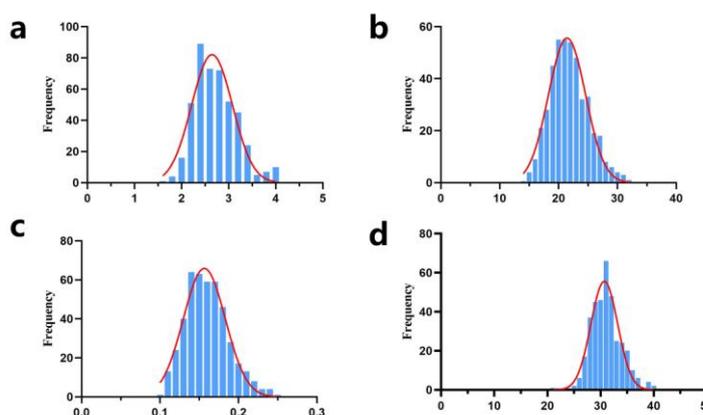
### 2.7. Candidate Gene Annotation and Functional Enrichment Analysis

The physical position information was retrieved from the *Capra hircus* reference genome (Genome assembly ASM4082201v1, [https://www.ncbi.nlm.nih.gov/datasets/genome/GCA\\_040822015.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_040822015.1/)). Candidate genes located within 300 kb upstream and downstream of the significant CNVs were identified using the bedtools software [26]. Subsequently, the overlapping genes were subjected to enrichment analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) tool, which included Gene Ontology (GO) functional annotation and KEGG pathway analysis. Enrichment terms with statistical significance ( $P < 0.05$ ), determined by Fisher's exact test, were selected to further investigate genes involved in relevant biological pathways and processes [14,27]. Gene functions were queried using Ensembl Biomart ([www.ensembl.org/biomart/martview](http://www.ensembl.org/biomart/martview)).

### 3. Results

#### 3.1. Phenotypic Statistics of Early Growth Traits in IMCGs

In this study, we analyzed four early growth traits: birth weight (BW), weaning weight (WW), average daily gain (ADG), and yearling weight (YW), with their descriptive statistics summarized in Table 1. The mean values ( $\pm$ SD) were determined as  $2.66 \pm 0.45$  kg for BW,  $21.90 \pm 3.28$  kg for WW,  $0.16 \pm 0.03$  kg for ADG, and  $31.21 \pm 3.60$  kg for YW. The coefficients of variation (CV) for these traits were calculated at 17.10%, 14.98%, 18.75%, and 11.53%, respectively. Prior to statistical analysis, data quality control measures were implemented, including the exclusion of phenotypic records with missing values and the removal of outliers exceeding the threshold of mean  $\pm$  3 standard deviations. Consequently, the sample sizes presented in Table 1 for each trait are slightly reduced from the initial cohort of 461 individuals. Furthermore, normality assessment through statistical tests confirmed that all traits followed a normal distribution pattern (Figure 1), validating the suitability of parametric statistical methods for subsequent analyses.



**Figure 1.** Normal distribution diagram of early growth traits of Inner Mongolia cashmere goats. (a) : Normal distribution map of birth weight traits (unit : kg) ; (b) : Normal distribution map of weaning weight traits (unit kg) ; (c) : Normal distribution map of average daily gain (unit : kg) ; (d) : Normal distribution of yearling weight traits (unit : kg); frequency= the frequency of different phenotypic values.

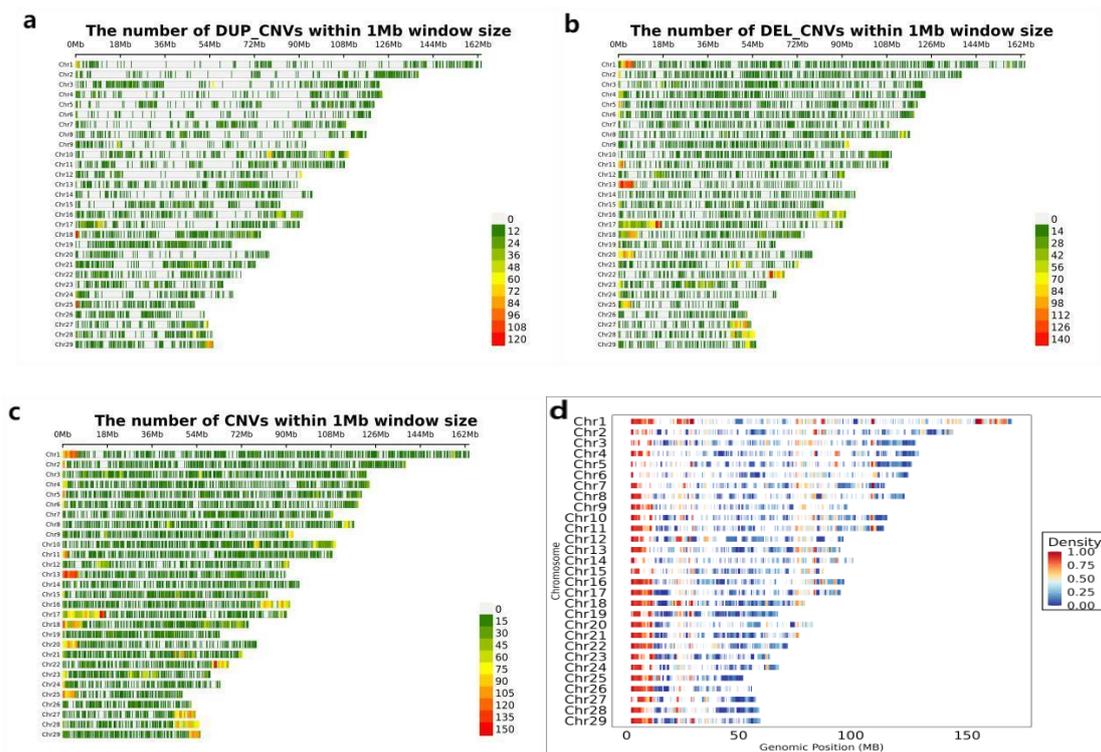
**Table 1.** Descriptive statistics of early growth traits of IMCGs (unit:kg) .

Trait	Number	Mean	SD	Max	Min	CV(%)
BW	448	2.69	0.46	4.00	1.70	17.10
WW	443	21.90	3.28	31.73	14.86	14.98
ADG	443	0.16	0.03	0.246	0.103	18.75
YW	359	31.21	3.60	39.5	21	11.53

Abbreviations: BW=birth weight; WW=weaning weight; ADG=average daily gain; YW=yearling weight; SD=standard deviation; Max=maximum; Min=minimum; CV=coefficient of variation.

### 3.2. Detection of Genome-Wide Copy Number Variation in IMCGs

Following quality control procedures, 11.87 TB of high-quality reads were retained from the initial 13.41 TB raw sequencing data obtained from 461 Inner Mongolian Cashmere Goats (IMCGs). Utilizing CNVnator software (v0.4.1), we performed genome-wide CNV detection across all 29 autosomes. The initial analysis identified 315,891 CNV events, categorized as either losses (copy number = 0 or 1) or gains (copy number  $\geq 2$ ). To ensure data reliability, we implemented stringent filtering criteria. CNVs with a population frequency below 0.01 (present in fewer than 5 individuals) were excluded as potential false positives. Furthermore, overlapping CNVs detected at identical genomic positions across multiple individuals were merged and systematically reannotated. This rigorous filtering process yielded a final set of 26,003 non-redundant CNVs, comprising 10,216 gain-type and 15,787 loss-type events (Figure 2a,b), each present in at least five individuals. The cumulative length of identified CNVs spanned 3,469.5 Mb, exceeding the total length of the goat autosomes (2,605.7 Mb). This apparent discrepancy arises from the overlapping nature of CNVs across individuals during the detection process, preventing direct calculation of chromosomal coverage. To address this, we generated two complementary visual representations: illustrates the density distribution of CNVs across chromosomes was shown in Figure 2c, while Figure 2d presents CNV coverage using a 1 Mb sliding window approach along the chromosomes.



**Figure 2.** The overall CNV maps for U.S. and IMCGs in the 29 autosomes. (a): The chromosome density distribution of the gains CNV ; (b): The chromosome density distribution of the losses CNV ; (c): The distribution of the number of CNVs on each chromosome ; (d): The distribution of the density of CNVs on each chromosome after dividing the area according to the size of 1Mb. Chr = chromosome; Genomic Position = the position of CNV on the chromosome.

The genomic distribution and characteristics of CNVs across the 29 autosomes are summarized in Table 2. On average; each chromosome contained 897 CNVs, consisting of 352 gain-type and 545 loss-type events. While CNV distribution generally correlated with chromosome length, we observed notable exceptions to this trend. Chromosome 17 exhibited the highest CNV count (1,638), followed

by chromosome 1 (1,503), despite chromosome 1 being the longest autosome. In contrast, chromosome 26 showed the lowest CNV frequency (359), with chromosome 24 displaying a similar low count (391)

The identified CNVs exhibited a size range from 0.8 kb to 4.99 Mb, with a mean length of 133 kb. Specifically, the largest CNV identified was CNV\_DUP\_13234 located on chromosome 16, spanning 4.99 Mb, while the smallest was CNV\_DEL\_7892 on chromosome 10, measuring 0.8 kb.

Of the total 26,003 CNVs detected, gain-type CNVs accounted for 39.3% (n = 10,216), ranging in size from 4.8 kb to 4.99 Mb. These gain events collectively spanned 2,285.9 Mb, with an average length of 224 kb. In contrast, loss-type CNVs represented 60.7% (n = 15,787) of the total, ranging from 0.8 kb to 4.85 Mb in length. The cumulative length of deletion CNVs was 1,183.5 Mb, with an average size of 75 kb.

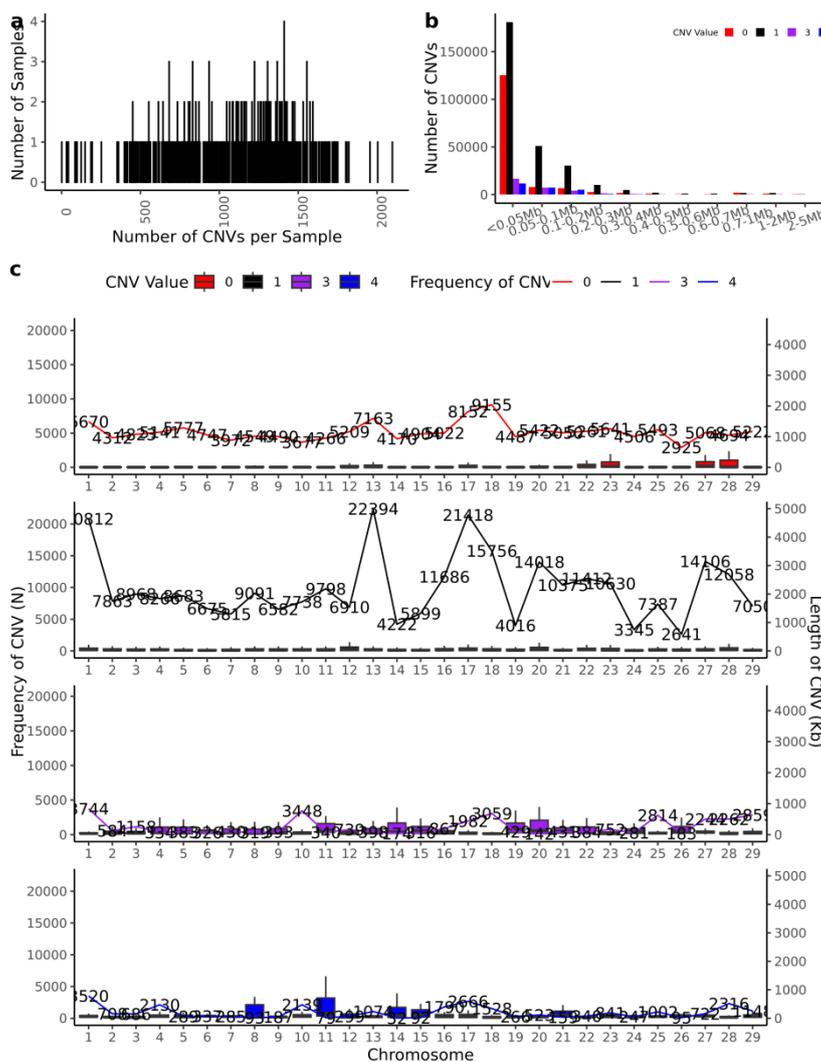
The CNV distribution patterns across chromosomes in 461 IMCG individuals are illustrated in Figure 2, where copy numbers of 0 or 1 represent losses and 3 or 4 represent gains. Figure 2a displays the individual CNV count distribution, revealing that the top three individuals contained 2,097, 2,004, and 1,955 CNVs respectively, while the bottom three individuals exhibited only 43, 34, and 31 CNVs. The average population was 685 CNVs per individual. Figure 2b presents the size distribution of CNVs across different copy number categories. The majority of CNVs (less than 0.05 Mb) were significantly more abundant than larger CNVs (2-5 Mb). Loss-type CNVs (copy number = 0 or 1) substantially outnumbered gain-type CNVs (copy number = 3 or 4), with most CNVs concentrated below 0.3 Mb in length.

Chromosomal distribution patterns are shown in Figure 2c. Loss-type CNVs (copy number = 0) were most prevalent on chromosome 18 (9,155 events) and least frequent on chromosome 26 (2,925 events). Similarly, CNVs with copy number = 1 showed maximal distribution on chromosome 13 (22,394 events) and minimal on chromosome 26 (2,641 events). For gain-type CNVs, those with copy number = 3 were most abundant on chromosome 10 (3,448 events) and least on chromosome 26 (183 events), while copy number = 4 CNVs peaked on chromosome 1 (3,520 events) and were rarest on chromosome 14 (32 events). This chromosomal distribution analysis confirms the relatively low CNV frequency on chromosome 26, consistent with the merged and reannotated CNV data presented in Table 2. The observed patterns suggest potential chromosome-specific mechanisms influencing CNV formation and maintenance in IMCGs.

Through integration of the 26,003 quality-controlled CNVs, we identified 5,014 CNV regions (CNVRs) (Supplementary Table 2) by merging adjacent CNVs with overlapping regions exceeding 1 bp. These CNVRs spanned all autosomes, ranging from 75 to 344 regions per chromosome, with a cumulative length of 1,015.43 Mb, representing 38.97% of the goat autosomal genome.

The CNVRs were classified into three categories: 1,085 gain-type CNVRs with a total length of 175 Mb (average length: 161.3 kb), 3,406 loss-type CNVRs spanning 328.76 Mb (average length: 96.5 kb), and 523 mixed-type CNVRs covering 511.67 Mb (average length: 978.3 kb). Chromosome 19 exhibited the highest CNVR coverage (65.58%), while chromosome 9 showed the lowest coverage (27.20%).

Analysis of the 20 longest CNVRs revealed that 50% were mixed-type, with 6 located in chromosomal telomeric regions. This distribution pattern suggests that CNVs frequently occur in highly repetitive telomeric regions, which are known hotspots for large-scale genomic recombination and replication events.



**Figure 3.** CNV mathematical statistics diagram. (a): The statistical map of the number distribution of CNVs in 461 individuals ; (b): the statistical map of the length of CNVs in 461 individuals ( where CNVs with copy number greater than 4 are counted as copy number 4 ) ; (c): The distribution map of CNVs on chromosomes of 461 individuals ( where CNVs with copy number greater than 4 are counted as copy number 4 ).

**Table 2.** The distribution of all CNVs in the 29 autosomes of IMCGs.

Chr	Length of CNVs DUP_C				Length of CNVs DEL_C					
	count	NVs counts	Max size (kb)	Average size (kb)	Min size (kb)	DEL_C NVs counts	Max size (kb)	Average size (kb)	Min size (kb)	
1	1503	415	73980	3422	178	1088	93020	4050	85	1.6
2	817	324	72388	4587	223	493	48960	2739	99	1.6
3	1036	502	88685	3294	177	534	33659	1426	63	1.6
4	827	328	63679	2690	194	499	40498	2582	81	1.6
5	846	338	96786	4485	286	508	26974	2286	53	1.6
6	633	174	33754	3142	194	459	40347	4634	88	1.6
7	687	330	84095	3285	255	357	18855	1778	53	1.6
8	880	352	64792	3549	184	528	12378	726	35	1.6
9	655	214	31962	2146	149	441	34011	1710	77	1.6
10	1203	705	97979	2422	139	498	33174	1249	67	0.8
11	909	390	117598	3729	302	519	27412	1414	53	1.6



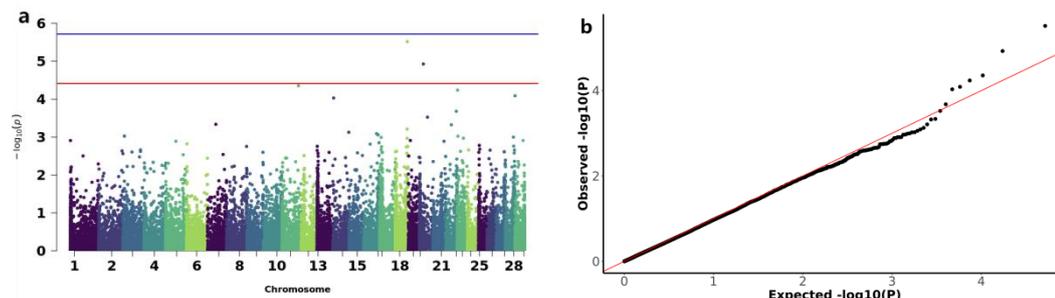
12	720	258	46924	4130	182	10.4	462	87199	4165	189	1.6
13	1310	378	125859	4086	333	9.6	932	58391	1208	63	1.6
14	491	209	47794	4354	229	8.8	282	24636	3558	87	1.6
15	715	362	70346	1770	194	8.8	353	22931	2049	65	1.6
16	1229	559	121854	4996	218	5.6	670	2281	4857	73	1.6
17	1638	500	106091	4653	212	8.0	1138	114083	2491	100	1.6
18	1493	629	137963	4555	219	7.2	864	55394	1189	64	1.6
19	609	374	113817	4564	304	11.2	235	10993	3342	47	0.8
20	811	157	40919	2086	261	8.0	654	79986	2543	122	1.6
21	924	405	93685	2100	231	9.6	519	24792	1257	48	0.8
22	905	268	70214	3579	262	13.6	637	63890	2682	100	1.6
23	880	281	71557	3488	255	9.6	599	55915	2194	93	1.6
24	391	192	45366	4572	236	9.6	199	11270	2479	57	1.6
25	755	326	91760	4768	281	7.2	429	27846	1980	65	1.6
26	359	194	56654	4493	292	9.6	165	6951	670	42	1.6
27	948	280	46586	2906	166	8.0	668	51005	1039	76	1.6
28	1031	328	74822	3056	228	7.2	703	64270	1950	91	1.6
29	798	444	98039	4518	221	8.8	354	12378	726	35	0.8
total	26003	10216	2285946	224			15787	1183502	75		

Abbreviations: Chr = chromosome; CNVs counts= number of gain-type CNVs + number of loss-type CNVs; DUP\_CNVs counts= number of gain-type CNVs; Length of DUP\_CNVs= the total length of gain-type CNV; DEL\_CNVs counts= number of loss-type CNVs; Length of DEL\_CNVs = the total length of loss-type CNVs; Max size = the maximum value of different types of CNV; Average size= the average value of different types of CNV; Min size= the minimum value of different types of CNVs.

### 3.3. Genome-Wide Association Study of Copy Number Variations with Early Growth Traits in IMCGs

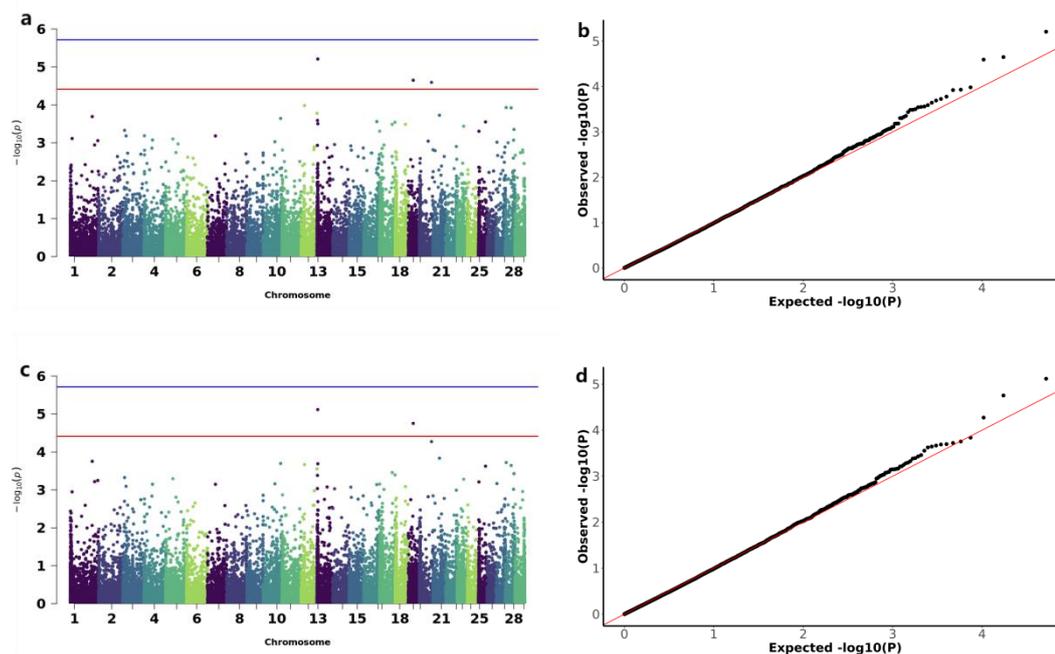
To investigate the functional significance of CNVs in goat growth development, we conducted genome-wide association studies (GWAS) for four early growth traits. The results are presented in Table 3. Using the Bonferroni correction method, we established a genome-wide significance threshold of  $3.85 \times 10^{-5}$  ( $1/N$ , where  $N = 26,003$ ) for identifying significant CNV-trait associations. The GWAS analysis revealed 11 CNVs that surpassed the significance threshold, demonstrating associations with early growth traits. Specifically, we identified two genome-wide significant CNVs associated with BW on chromosomes 18 and 20, three CNVs associated with both WW and ADG on chromosomes 13, 19, and 20, and six CNVs associated with YW distributed across chromosomes 3, 5, 21, 22 and 23. Consistent with the previously reported high genetic correlation between WW and ADG (GAO. et al., 2024), our analysis identified two overlapping CNVs (CNV\_DEL\_11189 and CNV\_DEL\_17895) that showed significant associations with both traits. These CNV regions, along with all other identified variants, were precisely mapped and validated using bedtools software. Functional annotation of the 11 significant CNVs revealed their association with 91 protein-coding genes, as detailed in Table 4.

The GWAS analysis for BW identified two significant loss CNVs: CNV\_DEL\_17406 on chromosome 18 and CNV\_DEL\_18821 on chromosome 20 (Figure 4a). These two CNVs were functionally annotated to four candidate genes, including Zinc finger protein 845 (*ZN845*), Endogenous retrovirus group K member 25 Env polyprotein (*ENK25*), Zinc finger protein 160 (*ZN160*), and Zinc finger protein 208 (*ZN208*) among others (Table 3).



**Figure 4.** GWAS and significant CNV analysis of BW (birth weight) traits of IMCGs. (a): Manhattan diagram of genome-wide association study of birth weight traits ; (b): Quantile-quantile ( Q-Q ) plots for genome-wide association studies of birth weight traits; GO=Gene ontology; KEGG=Kyoto Encyclopedia of Genes and Genomes.

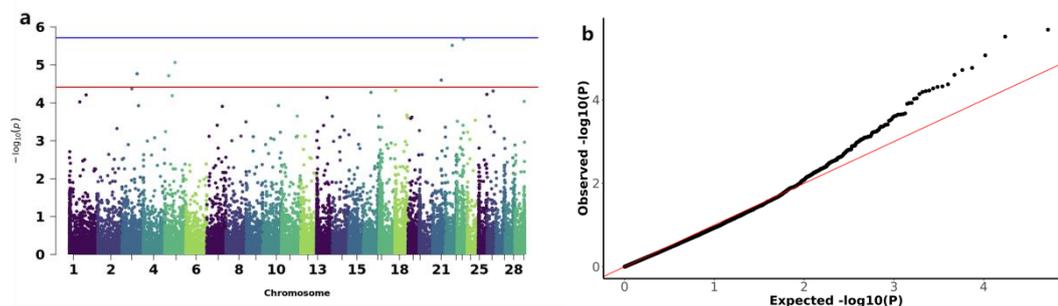
The GWAS analysis for ADG and WW identified three significant CNVs: two loss-type CNVs (CNV\_DEL\_11189 on chromosome 13 and CNV\_DEL\_17895 on chromosome 19) showing associations with both traits, and one gain-type CNV (CNV\_DUP\_18956 on chromosome 20) specifically associated with ADG (Figure 5). The overlapping associations of CNV\_DEL\_11189 and CNV\_DEL\_17895 with both ADG and WW support the previously reported high genetic correlation between these traits. Functional annotation revealed that these three CNVs encompassed 66 candidate genes, including Transcription factor SOX-15 (*SOX15*) and Thialysine N-epsilon-acetyltransferase (*SAT2*) (Table 3).



**Figure 5.** GWAS of ADG (average daily gain) and WW (weaning weight) traits of IMCGs. (a): Manhattan diagram of genome-wide association study of average daily gain traits ; (b): Quantile-quantile ( Q-Q ) plot of genome-wide association study of average daily gain traits ; (c): Manhattan diagram of genome-wide association study of weaning weight traits ; (d): Quantile-quantile ( Q-Q ) plot of genome-wide association study of weaning weight traits.

The GWAS analysis for YW identified six significant CNVs distributed across chromosomes 3,5,21,22 and 23: CNV\_DUP\_3067, CNV\_DEL\_4359, CNV\_DEL\_4552, CNV\_DUP\_19581, CNV\_DUP\_20170, and CNV\_DUP\_21558 (Figure 6a-b). Functional annotation of these six CNVs

revealed their association with 32 candidate genes, including Inositol 1,4,5-trisphosphate-gated calcium channel (*ITPR1*) and ADP-ribosylation factor-like protein 8B (*ARL8B*), among others (Table 3).



**Figure 6.** GWAS of YW(yearling weight) traits of IMCGs. (a): Manhattan diagram of genome-wide association study of yearling weight traits ; (b): Quantile-quantile (Q-Q) map of genome-wide association study for yearling weight trait.

**Table 3.** Significant CNV descriptive statistics.

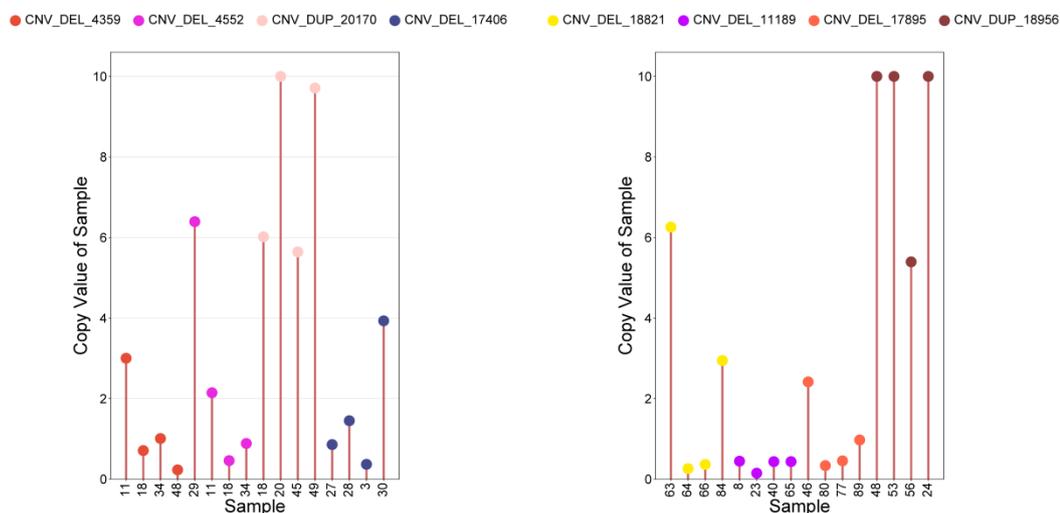
Trait	CNV ID	Type	Chr	Start (bp )	End (bp )	P-value	Candidate genes
BW	CNV_DEL_174_06	DEL	18	67870401	67903200	3.04E-06	<i>ENK25,ZN845,ZN160,ZN208</i>
	CNV_DEL_188_21	DEL	20	22544801	22552000	1.18E-05	
WW	CNV_DEL_111_89	DEL	13	3608001	4816000	7.61E-06	<i>TRA2B</i> <i>KCD11,SOX15,GBRAP,S35G3,TMM95,IPP2,SPEM1,CD68,SPEM2,SAT2,RNK,GPS2,CLD7,MPU1,TM102,BAP18,TNF13,TM256,MOT13,ZBTB4,ASGR1,PHF23,SHBG,CNEP</i>
	CNV_DEL_178_95	DEL	19	27445601	27462400	1.76E-05	<i>1,PLS3,YBOX2,ACADV,IF5A1,IF4A1,ELP5,EFNB3,GLUT4,FGF11,BCL6B,DVL2,TNF12,TNK1,ACHB,SEN3,NEUL4,ACAP1,TM102,BAP18,TNF13,TM256,MOT13,ZBTB4,ASGR1,PHF23,SHBG,CNEP1,PLS3,YBOX2,ACADV,I</i>
ADG	CNV_DEL_111_89	DEL	13	3608001	4816000	7.61E-06	<i>TRA2B</i> <i>KCD11,SOX15,GBRAP,S35G3,TMM95,IPP2,SPEM1,CD68,SPEM2,SAT2,RNK,GPS2,CLD7,MPU1,TM102,BAP18,TNF13,TM256,MOT13,ZBTB4,ASGR1,PHF23,SHBG,CNEP</i>
	CNV_DEL_178_95	DEL	19	27445601	27462400	1.76E-05	<i>1,PLS3,YBOX2,ACADV,I</i>

						F5A1,IF4A1,ELP5,EFNB3 ,GLUT4, <b>FGF11</b> ,BCL6B, <b>D</b> <b>VL2</b> ,TNF12,TNK1,ACHB, SENP3,NEUL4,ACAP1,T CAB1,ASGR2,P53,FXR2, LOX12,NLGN2,DLG4,RP B1,DYH2
	CNV_DUP_189 56	DUP	20	69328801	69360800	2.55E-05 TCTP,TCPE,ACKMT,RO P1L,CMBL,MARH6
YW	CNV_DUP_306 7	DUP	3	85662401	86454400	1.71E-05 RL4,NFIA
	CNV_DEL_435 9	DEL	5	20603201	20628800	1.93E-05
	CNV_DEL_455 2	DEL	5	56564001	56606400	8.58E-06 ELOC, <b>SPRY4</b> ,APOF,IL23 A,RDH16,APOF,GP182,C NPY2,RDH16,ATPB,MIP ,RDH16,NACAM,RDH16 ,GLSL,PAN2, <b>STAT2</b> ,PRI1 ,H17B6,TIM,BAZ2A,RB MS2
	CNV_DUP_195 81	DUP	21	48555201	48634400	2.51E-05 TR10C,TR10D,LIPA1
	CNV_DUP_201 70	DUP	22	38884801	39041600	3.05E-06 BHE40,ARL8B,SUMF1,IT PR1
	CNV_DUP_215 58	DUP	23	37214401	37316800	2.08E-06

Abbreviations: BW=birth weight; WW=weaning weight; ADG=average daily gain; YW=yearling weight; CNV ID= Unique identifier for statistically significant CNVs; Type= the type of significant CNV; Chr= chromosome; Start= the starting position of significant CNV; End= the termination position of significant; P-value= the P-value of significant CNV; Candidate genes= candidate genes annotated within 300 kb upstream and downstream of significant CNV.

### 3.4. The Phenotypic Comparison Between the Identified CNVs and Individuals Containing Significant CNVs and IMCGs Populations was Verified by PCR analysis

The PCRs were performed to verify 8 CNVs in 32 samples. As shown in Figure 4, more than 78.0 % of the results are consistent with the type of CNV predicted using CNVnator. (Figure 7)



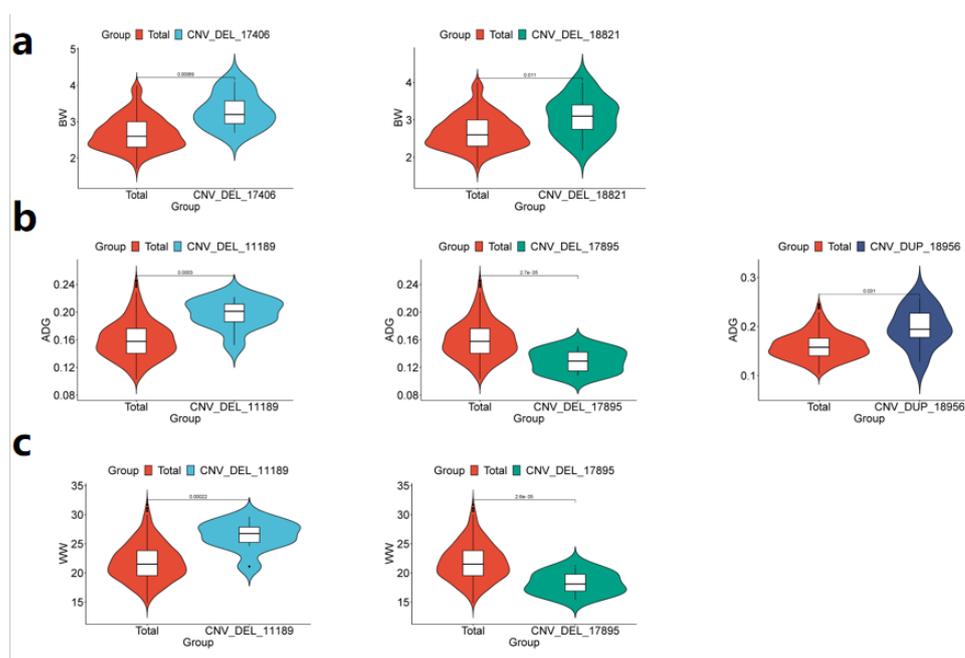
**Figure 7.** The CNV results were verified by qPCR. Sample = Individuals performing copy number detection; Copy Value of Sample = the copy number of the experimental individual; Copy Value: Experimental samples (values >10 truncated at 10); Control samples (default copy number = 2).

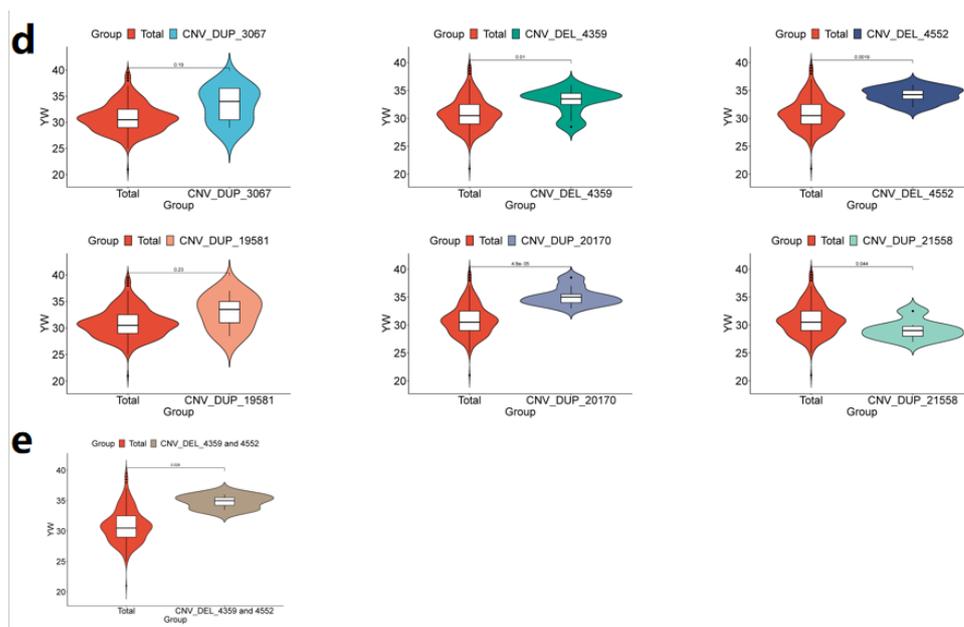
Comparative analysis revealed that individuals carrying either CNV\_DEL\_17406 or CNV\_DEL\_18821 exhibited significantly higher average BW ( $p < 0.05$ ) compared to the overall IMCG population mean of 2.69 kg (Figure 8a). This finding suggests a potential positive association between these CNVs and increased birth weight in the IMCG population.

Comparative analysis demonstrated significant phenotypic effects: individuals carrying CNV\_DEL\_11189 or CNV\_DUP\_18956 exhibited higher average ADG ( $p < 0.05$ ) than the population mean (0.16 kg), while those with CNV\_DEL\_17895 showed lower ADG ( $P < 0.05$ ) (Figure 8b). Similarly, for WW traits, CNV\_DEL\_11189 carriers displayed significantly higher WW ( $P < 0.05$ ) than the population mean (21.90 kg), whereas CNV\_DEL\_17895 carriers showed reduced WW ( $P < 0.05$ ) (Figure 8c).

Comparative phenotypic analysis demonstrated distinct effects of these CNVs on YW: individuals carrying CNV\_DEL\_4359, CNV\_DEL\_4552, or CNV\_DUP\_20170 exhibited significantly higher YW ( $P < 0.05$ ) than the population mean (31.21 kg). While CNV\_DUP\_3067 and CNV\_DUP\_19581 carriers showed elevated YW compared to the population mean, these differences did not reach statistical significance. In contrast, individuals with CNV\_DUP\_21558 displayed significantly lower YW ( $P < 0.05$ ) than the population average (Figure 8d). Notably, CNV\_DEL\_4359 and CNV\_DEL\_4552 showed particularly strong positive associations with YW, with carriers exhibiting significantly higher weights than the population mean (Figure 8e).

Comparative analysis using independent samples t-tests revealed significant phenotypic divergence ( $P < 0.05$ ) between CNV carriers and the Inner Mongolia Cashmere Goats (IMCGs) population baseline in 9 out of 11 candidate copy number variations (CNVs) associated with early growth traits. While two CNV loci (CNV\_DEL\_17406 and CNV\_DUP\_20170) showed non-significant P-values (0.19 and 0.23 respectively), their effect sizes (Cohen's  $d > 0.4$ ) demonstrated meaningful biological divergence from population means. This validation framework achieved 81.8% concordance between GWAS-identified loci and measurable phenotypic effects, with the remaining variants showing directional consistency in trait modulation, thereby providing orthogonal validation for the GWAS findings.





**Figure 8.** Phenotypic comparison between individuals with significant CNVs and IMCGs populations. (a): Phenotypic analysis of BW(birth weight) traits of individuals containing CNV\_DEL\_17406 or CNV\_DEL\_18821 ; (b): Phenotypic analysis of ADG(average daily gain) traits in individuals containing CNV\_DEL\_11189 or CNV\_DEL\_17895 or CNV\_DUP\_18956 ; (c): Phenotypic analysis of WW(weaning weight) trait in individuals with CNV\_DEL\_11189 or CNV\_DEL\_17895 ; (d): Phenotypic analysis of YW(yearling weight) traits of individuals containing CNV \_DUP\_3067 or CNV\_DEL\_4359 or CNV\_DEL\_4552 or CNV\_DUP\_19581 or CNV\_DUP\_20170 or CNV\_DUP\_21558 ; (e): Phenotypic analysis of YW(yearling weight) traits containing CNV\_DEL\_4359 and CNV\_DEL\_4552 individuals; Total = phenotypic value of Inner Mongolia cashmere goat population; unit : kg.

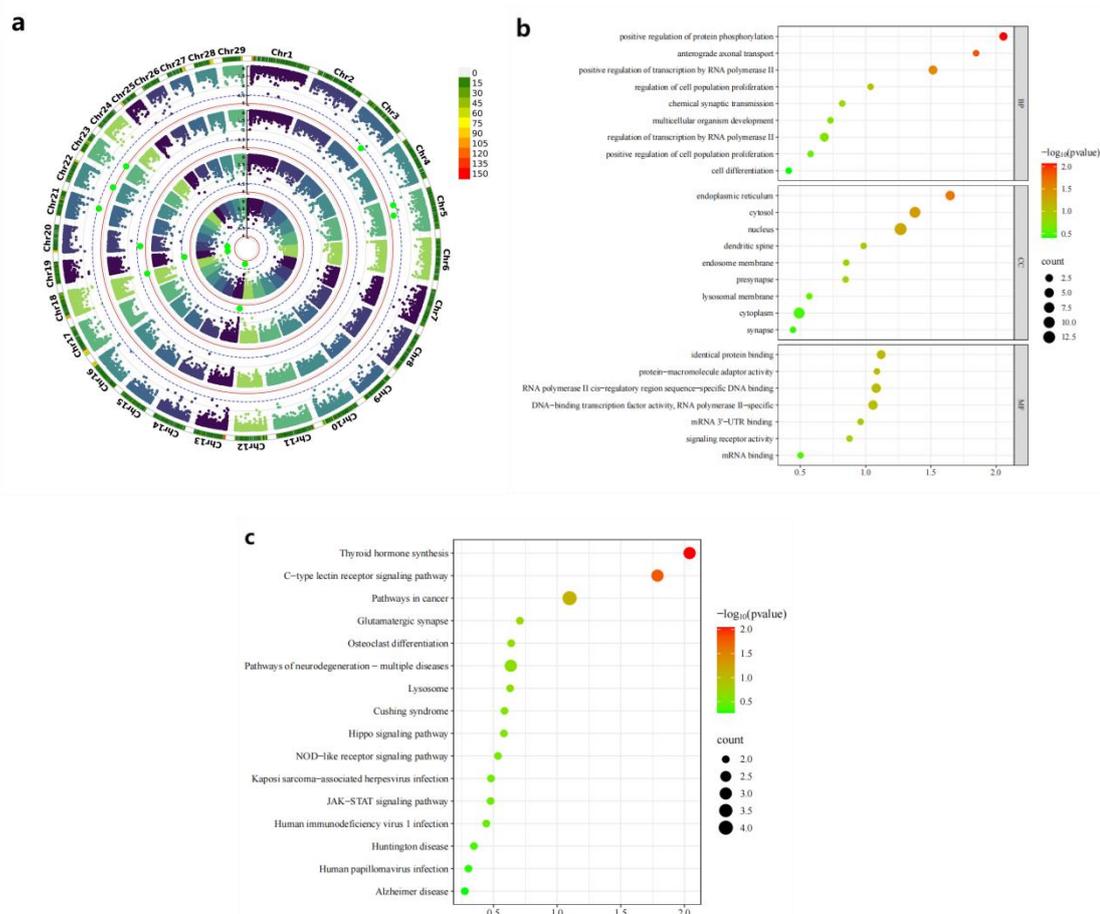
### 3.5. Functional Analysis of Genes Associated with Trait-related CNVs

Genomic annotation was performed on genes overlapping with the 11 significant CNVs, including their upstream and downstream 300 kb flanking regions. Based on the Ensembl annotation of the *Capra hircus* reference genome (Genome assembly ASM4082201v1), we identified a total of 91 genes. To elucidate the functional relevance of these genes to early growth traits in IMCGs, we conducted comprehensive enrichment analyses using both Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) terms. The gene set enrichment analysis revealed several functionally relevant terms associated with early growth traits in IMCGs. Specifically, the analysis identified 20 significant GO terms, comprising 6 cellular component terms, 7 biological process terms, and 7 molecular function terms. KEGG pathway analysis demonstrated that these genes are predominantly involved in crucial biological processes, including cellular development, hormonal biosynthesis, protein phosphorylation, and related metabolic pathways (Figure 9).

Through integrative analysis combining data from the GeneCards database and published literature, we identified several genes implicated in key biological pathways and processes. From this analysis, we selected seven candidate genes that not only overlapped with significant CNVs but also showed enrichment in our gene set analysis (Table 4). The candidate gene analysis revealed *ZN845* as the primary candidate for BW, while *SOX15*, *FGF11*, *GPS2*, and *DVL2* emerged as candidate genes for WW and ADG. For YW, *SPRY4* and *STAT2* were identified as potential candidate genes. These candidate genes and their associated traits are comprehensively summarized in Table 4.

Functional enrichment analysis revealed distinct biological roles for the identified candidate genes. For BW traits, *ZN845* did not exhibit significant enrichment in any specific pathways within our analysis framework. For ADG and WW traits, The *SOX15* gene demonstrated significant

associations with critical cellular processes, including cell differentiation, DNA-binding transcription factor activity, and RNA polymerase II-specific regulation. *FGF11* was prominently enriched in pathways governing protein phosphorylation modulation and cellular differentiation processes. *GPS2* showed specific involvement in transcriptional regulation mediated by RNA polymerase II, particularly in the context of human T-cell leukemia virus 1 infection. *DVL2* was significantly enriched in multiple pathways, including the regulation of cell population proliferation, signaling pathways controlling stem cell pluripotency, and positive regulation of protein phosphorylation. For YW traits, *STAT2* was associated with fundamental biological processes such as cell proliferation, same-protein binding interactions, and the regulation of osteoclast differentiation. *SPRY4* is significantly enriched in cell solutes (Table 4).



**Figure 9.** GO and KEGG functional enrichment maps of key candidate genes for early growth traits of Inner Mongolia cashmere goats. A: The Manhattan diagram of the GWAS results of the early growth traits of Inner Mongolia cashmere goats; B: the GO functional enrichment map of the key candidate genes; C: The KEGG pathway enrichment map of the key candidate genes; GO=Gene ontology; KEGG=Kyoto Encyclopedia of Genes and Genomes.

**Table 4.** GO and KEGG functional enrichment of 7 candidate genes for early growth traits of Inner Mongolia Cashmere Goats.

Trait	Chr	Gene	Significantly enriched pathways
BW	18	<i>ZN845</i>	cell differentiation, DNA-binding
ADG, WW	13	<i>SOX15</i>	transcription factor activity, RNA polymerase II-specific

	19	<i>FGF11</i>	positive regulation of protein phosphorylation, cell differentiation regulation of transcription by RNA
	13	<i>GPS2</i>	polymerase II, Human T-cell leukemia virus 1 infection
	19	<i>DVL2</i>	regulation of cell population proliferation, Signaling pathways regulating pluripotency of stem cells, positive regulation of protein phosphorylation
YW	5	<i>SPRY4</i>	cytosol
	5	<i>STAT2</i>	regulation of cell population proliferation, identical protein binding, Osteoclast differentiation

Abbreviations: BW=birth weight; WW=weaning weight; ADG=average daily gain; YW=yearling weight; Chr=chromosome; Gene= key candidate gene; Significantly enriched pathways= Functional enrichment of key candidate genes in GO terms and KEGG pathways.

#### 4. Discussion

With the rapid development of bioinformatics, the number of SNPs that significantly affect complex traits through GWAS analysis is increasing. However, many SNPs can only explain the heritability of some complex traits, which is called 'missing heritability' [28,29]. CNV is a widespread variation phenomenon in genome genetic variation and an important part of human, animal and plant genomes. GWAS analysis using CNV may explain the genetic variation of complex traits that cannot be explained by some SNPs [30]. Early growth traits ( BW, ADG, WW, YW ) are important traits of Inner Mongolia Cashmere Goats throughout the feeding cycle [15]. Therefore, it is of great significance for the breeding and feeding of cashmere goats to explore the molecular regulation mechanism of early growth traits of Inner Mongolia cashmere goats and the key candidate genes affecting early growth traits of cashmere goats. Although there have been many SNP-based GWAS studies to identify the key genes of weight traits in goats [5,31]. However, there are relatively few studies on GWAS analysis of body weight traits of cashmere goats based on CNV to determine the key genes affecting body weight traits of cashmere goats. Therefore, this study is the first to use the whole genome resequencing of Inner Mongolia Cashmere Goats to screen the CNV of Inner Mongolia Cashmere Goats population, and perform GWAS analysis of BW, ADG, WW, and YW to mine key candidate genes. In this study, CNVnator was used to detect CNV in high-depth whole-genome resequencing data ( 13.41Tb ) of 461 Inner Mongolia cashmere goats based on Read Depth method and 800 bp sliding window. The detection of CNV using resequencing data usually requires the sequencing depth of sequencing data to be at least 5× and above [32]. The average sequencing depth of resequencing data used in this study was 9.46×, which met the requirements of resequencing data to detect CNV. A total of 315,891 CNVs were detected in this study. After deleting CNVs with abnormal length and low frequency ( 1 % ), 26,003 CNVs were finally retained for CNV merging and GWAS analysis. Among them, 10,216 ( 39.3 % ) were gain-type CNVs and 15,787 ( 60.7 % ) were loss-type CNVs. The number of loss-type CNVs was greater than that of gain CNVs, which was consistent with the results of other studies on CNV detection in cattle, sheep and pigs [9,33,34]. Among the 32 experimental samples subjected to qPCR analysis, target CNVs were successfully detected in 25 samples, demonstrating a detection frequency of 78.0% (25/32). Notably, each target CNV was detected in at least two experimental individuals, ensuring the reliability of our findings. Furthermore, analysis of the seven experimental individuals that did not show target CNV detection revealed that five of these individuals exhibited copy number variations (CNVs) different from the normal diploid state (copy number ≠ 2). This observation suggests the presence of genomic structural

variations in the target CNV region, thereby providing additional evidence supporting the authenticity and reliability of the CNVs detected in this study.

Based on the GWAS results, we identified 11 CNVs significantly associated with BW, ADG, WW and YW. Comparative analysis between the phenotypes of individuals carrying significant CNVs and the overall population revealed distinct patterns of CNV effects. For the BW trait, two significant CNVs (CNV\_DEL\_17406 and CNV\_DEL\_18821) exhibited a gain-of-function effect in IMCG.

Furthermore, we identified two CNVs (CNV\_DEL\_11189 and CNV\_DEL\_17895) that were shared between ADG and WW traits, along with one ADG-specific CNV (CNV\_DUP\_18956). Notably, while CNV\_DUP\_18956 did not reach genome-wide significance for WW ( $P=5.34\times 10^{-5}$ ), its P-value was remarkably close to the significance threshold ( $P<3.85\times 10^{-5}$ ). Functional analysis showed that CNV\_DEL\_11189 exerted a gain effect on both ADG and WW traits in IMCG, whereas CNV\_DEL\_17895 displayed an inhibitory effect on these traits. The ADG-specific CNV\_DUP\_18956 demonstrated a gain effect on ADG traits in IMCG. Regarding YW traits, we detected six significant CNVs (CNV\_DUP\_3067, CNV\_DEL\_4359, CNV\_DEL\_4552, CNV\_DUP\_19581, CNV\_DUP\_20170, and CNV\_DUP\_21558). Among these, CNV\_DUP\_21558 showed an inhibitory effect on YW traits in IMCG, while the remaining five CNVs exhibited gain effects on YW traits.

Following the annotation methodology described by Xin et al. [35], we performed comprehensive gene annotation within 300 kb upstream and downstream regions of the 11 identified CNVs, identifying a total of 91 genes. Subsequent GO and KEGG pathway enrichment analyses revealed that 37 of these genes were significantly enriched in various biological functions and pathways. Through integration of previous research findings with our GO and KEGG enrichment results, we identified seven key candidate genes potentially influencing early growth traits in IMCGs: *ZN845*, *SOX15*, *FGF11*, *GPS2*, *DVL2*, *SPRY4*, and *STAT2*. Considering that weight gain in animals is closely associated with muscle development, fat deposition, and obesity [36], we conducted functional enrichment analysis of these seven candidate genes. The results demonstrated their predominant involvement in crucial biological pathways, particularly cell proliferation and energy metabolism.

Among these genes, zinc finger protein 845 (*ZN845*), located within the CNV\_DEL\_17406 region on chromosome 18, emerged as particularly relevant to body weight. As a member of the zinc finger protein family, *ZN845* represents a crucial class of transcription factors that play pivotal roles in various biological processes, including DNA recognition, RNA packaging, transcriptional activation, apoptosis regulation, protein folding and assembly, and lipid binding. Previous studies have established its critical functions in plant stress resistance and abiotic stress responses [27,37], as well as its regulatory role in animal muscle development. Our findings strongly suggest that *ZN845* serves as a key candidate gene influencing early growth traits in cashmere goats.

The transcription factor *SOX-15* (*SOX15*), located within the CNV\_DEL\_17895 region on chromosome 19, represents the sole member of group G in the Sry-related high-mobility-group (HMG) box (SOX) gene family. This gene demonstrates significant associations with average daily gain (ADG) and weaning weight (WW), playing crucial roles in multiple biological processes, including male gonad development, striated muscle tissue formation, myoblast proliferation, and skeletal muscle regeneration. Experimental evidence from Ito et al. [38] demonstrated that mice with defective *SOX15* genes exhibited significantly delayed skeletal muscle regeneration, highlighting the gene's critical role in murine muscle development. Furthermore, Kayo Yamada et al. [39] identified substantial *SOX15* expression in murine placental tissue, suggesting its significant impact on placental development in mammals. These collective findings from both muscle and placental development studies strongly indicate that the *SOX15* gene serves as a key candidate gene influencing early growth traits in Inner Mongolia cashmere goats.

Fibroblast growth factor 11 (*FGF11*), co-localized with *SOX15* in the CNV\_DEL\_17895 region of chromosome 19, represents a crucial member of the intracellular fibroblast growth factor family, primarily involved in nervous system development and function in animals [40]. Experimental evidence from Zhao et al. [41] demonstrated that hypothalamic injection of RNA virus to inactivate

*FGF11* in mice fed a high-fat diet resulted in significant body weight reduction, decreased fat synthesis rates, and increased brown fat thermogenesis, highlighting the gene's substantial impact on adipocyte synthesis. Further supporting this, Li et al. [42] observed that *FGF11* knockout mice exhibited reduced expression of peroxisome proliferator-activated receptor gamma (*PPAR $\gamma$* ) gene, accompanied by decreased rates of adipocyte proliferation and differentiation. Notably, restoration of *FGF11* expression normalized *PPAR $\gamma$*  levels and accelerated adipocyte proliferation and differentiation. Recent findings by Jiang et al. [43] in goat models revealed that *FGF11* specifically regulates brown adipocyte differentiation and thermogenesis, while showing no significant effect on white adipocyte proliferation and differentiation. These collective findings underscore the dual role of *FGF11* in both adipocyte proliferation/differentiation and thermoregulation in goats. Based on its demonstrated functions in adipocyte regulation and cold resistance, we conclude that *FGF11* serves as a key candidate gene influencing early growth traits in Inner Mongolia cashmere goats.

The G protein pathway suppressor 2 (*GPS2*) gene, co-localized within the CNV\_DEL\_17895 region on chromosome 19, plays a crucial role in multiple biological processes including inflammation regulation, adipocyte differentiation, and lipid metabolism. Experimental studies by Justin et al. [44] demonstrated that *GPS2* knockout (*GPS2* AKO) mice fed a high-fat diet exhibited significantly increased susceptibility to inflammation and disrupted fat synthesis processes, accompanied by marked adipocyte proliferation. Further supporting these findings, Drareni et al. [45] established a connection between *GPS2* and the expansion of hypertrophic white adipose tissue in humans. Their observations of *GPS2* AKO mice revealed adipocyte hypertrophy, inflammation, and mitochondrial dysfunction under conditions of energy excess, providing compelling evidence for *GPS2*'s significant role in adipocyte differentiation and inflammatory responses. These collective findings from both murine and human studies strongly suggest that the *GPS2* gene serves as a key candidate gene influencing early growth traits in Inner Mongolia cashmere goats, particularly through its regulatory effects on adipocyte differentiation and metabolic processes.

Segment polarity protein disheveled homolog DVL-2 (*DVL2*) gene is also located in the CNV\_DEL\_17895 region of Chr19, which is involved in cell proliferation, protein phosphorylation and osteoblast differentiation [46]. Yamaguchi et al. found that the body's Stau1 protein negatively regulates the myogenesis of *C2C12* myoblasts by binding to the mRNA 3'untranslated region (UTR) of the *DVL2* gene. It shows that *DVL2* gene has an inhibitory effect on myogenesis [47]. The results of this study indicate that *DVL2* gene is a key candidate gene for early growth traits of Inner Mongolia cashmere goats.

SPRY domain-containing protein 4 (*SPRY4*) gene located in the CNV\_DEL\_4552 region of Chr5 is a protein-coding gene of the Spry family, which is related to YW and participates in biological processes such as cell proliferation, migration, inflammation, oxidative stress, apoptosis and organ development [48]. Li et al. [49] found that *SPRY4* gene was positively correlated with adipogenic differentiation of human mesenchymal stem cells (MSC). In vivo and in vitro experiments confirmed that *SPRY4* gene promoted hAMSC adipogenesis through MEK-ERK1 / 2 pathway, indicating that *SPRY4* gene was related to adipogenesis. The results of this study indicate that *SPRY4* gene is a key candidate gene for early growth traits of Inner Mongolia cashmere goats.

Signal transducer and activator of transcription 2 (*STAT2*) gene, also located in CNV\_DEL\_4552 region of Chr5, is involved in cell proliferation, protein phosphorylation and type I interferon mediated signaling pathway [50]. Yang et al. found that circCAPRIN1 promotes lipid synthesis by enhancing Acetyl-CoA carboxylase 1 (*ACC1*), and further analysis found that circCAPRIN1 directly binds to signal transduction and *STAT2* gene to activate *ACC1* transcription, thereby regulating lipid metabolism, indicating that *STAT2* gene is related to lipid synthesis in organisms [51]. The results of this study indicate that *STAT2* gene is a key candidate gene for early growth traits of Inner Mongolia cashmere goats.

The growth of body weight is related to animal muscle development, fat deposition and obesity. The key candidate genes of early growth traits of Inner Mongolia cashmere goats identified in this study are related to animal fat differentiation and muscle development.

## 5. Conclusions

In this study, we conducted a comprehensive genome-wide CNV detection analysis in the Inner Mongolia Cashmere Goat population, followed by GWAS analysis of early growth traits based on the CNV profiles. Our analysis identified a total of 26,003 CNVs, among which 11 demonstrated genome-wide significance. Through systematic characterization, we identified ZN845 as the candidate gene for body weight (BW) traits, while SOX15, FGF11, GPS2, and DVL2 emerged as candidate genes associated with average daily gain (ADG) and weaning weight (WW) traits. For yearling weight (YW) traits, SPRY4 and STAT2 were identified as potential candidate genes. Notably, while the FGF11 gene has been previously established as a significant regulator of adipocyte differentiation in goats, our study represents the first identification of the remaining genes in relation to muscle development and fat deposition in cashmere goats. These novel findings provide valuable targets for future functional studies and genetic investigations. The results of this study establish a crucial genetic foundation for elucidating the molecular mechanisms underlying early growth traits in Inner Mongolia Cashmere Goats. Furthermore, these findings offer significant potential for advancing genetic improvement strategies aimed at enhancing production performance in cashmere goat populations.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Table S1: qPCR primer and probe sequence information; Table S2: CNVR descriptive statistics.

**Author Contributions:** “Conceptualization, Yifan Liu and Haijiao Xi.; methodology, Bohan Zhou.; software, Qi Xu and Ruijun Wang.; validation, Yifan Liu.; formal analysis, Jinquan Li.; investigation, Yanjun Zhang.; resources, Rui Su.; data curation, Qi Lv.; writing—original draft preparation, Yifan Liu.; writing—review and editing, Zhiying Wang.; funding acquisition, Zhiying Wang. All authors have read and agreed to the published version of the manuscript.”

**Funding:** “This research was funded by Project for Natural Science Foundation of Inner Mongolia Autonomous Region, grant number 2023LHMS03003”; High level Achievement Cultivation Special Project of the Department of Animal Science, Inner Mongolia Agricultural University, grant number GZL202204”; Iconic Innovation Team of Ordos City, grant number TD20240001”; Key Project of First Class Discipline Scientific Research by the Education Department of Inner Mongolia Autonomous Region, grant number YLXKZX-NND-007”; China Agriculture Research System of MOF and MARA, grant number CARS-39”**Institutional Review Board Statement:** Not applicable. None of the tests described in the manuscript required the Ethics Committee or Institutional Review Board approval. There were no experimental animals involved in the study. Informed Consent Statement: Not applicable.

**Data Availability Statement:** No new data has been created yet.

**Acknowledgments:** The authors have reviewed and edited the output and take full responsibility for the content of this publication.”

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Appendix A

### Appendix A.1

**Table A1.** qPCR primer and probe sequence information.

CNVs	Chr	Forward (5'-3')	Reverse (5'-3')
------	-----	-----------------	-----------------

CNV_DEL_4359	5	ACTGCCTGATACTGAGTTTC CA	CACAAATGTTTACAGCAGC GT
CNV_DEL_4552	5	TGCAAGGAGGAAGCTAGA CC	TGAGCAGAGTTCCATGTCCT
CNV_DUP_20170	22	ATCTGGTGGGAGAGTTTGC A	GAGGGGAGGGGACAGTTAT G
CNV_DEL_17406	18	TCCTCCGAGTGTTCAAACC A	TCCCTATGTGTTTACGGTCT
CNV_DEL_18821	20	AAGGAGACAAAGGAGGTG GG	TCAGCAGGGAGAGTGTTTCC
CNV_DEL_11189	13	GACAGTGCTGCTACAACCTC G	TCCTCTCCAGTGCTGACATG
CNV_DEL_17895	19	GCCAGGCTGTGTAAGAAGT G	GGTGTCTTGCGTTGCTTAGG
CNV_DUP_18956	20	TGCATGTCTGTGTGTGAC C	TCCAACCCAGGGATTGAAC
<i>ACTB</i>	1	CCCTGGAGAAGAGCTACGA G	TAGTTTCGTGAATGCCGCAG

Abbreviations: CNVs= verified CNV; Chr= chromosome; Forward (5'-3') = forward primer sequence Reverse (5'-3') = reverse primer sequence.

**Table A2.** CNVR descriptive statistics.

Chr	length (kb)	CNVR counts	Length of CNVR (kb)	Coverage (%)	Max size (kb)	Average size (kb)	Min size (kb)
1	16359965 5	344	66064800	40.38	5278400	192048.84	1600
2	13802417 1	267	50842400	36.84	4587200	190420.97	1600
3	12226996 4	276	44209600	36.16	4652000	160179.71	1600
4	12350029 7	247	39647200	32.10	2583200	160514.98	1600
5	12035340 2	254	36261600	30.13	4760800	142762.20	1600
6	11890757 8	217	35036800	29.47	5974400	161459.91	1600
7	10885720 6	245	36660800	33.68	3284000	149635.92	1600
8	11732430 0	216	37060000	31.59	4484000	171574.07	1600
9	92782956 10989002	213	25237600	27.20	1817600	118486.38	1600
10	10850326 9	266	41044000	37.35	1899200	154300.75	1600
11	91085198	225	43721600	40.30	5149600	194318.22	1600
12	89647956	123	42136000	46.26	4960000	342569.11	1600
13	95224248	139	40302400	44.96	5753600	289945.32	1600
14	82415014	181	27188000	28.55	3641600	150209.94	1600
15	91491535	193	29837600	36.20	2164000	154598.96	1600
16	90066741	147	38597600	42.19	6874400	262568.71	1600
17	74692390	158	44683200	49.61	6545600	282805.06	1600
18	63059135	139	45996000	61.58	4545600	330906.47	1600
19		111	41355200	65.58	9339200	372569.37	1600

20	78028267	118	25727200	32.97	5857600	218027.12	1600
21	72404257	144	27694400	38.25	2244000	192322.22	1600
22	67135385	109	29719200	44.27	6012800	272653.21	1600
23	59408587	102	30975200	52.14	5171200	303678.43	1600
24	63378952	125	20003200	31.56	2479200	160025.60	1600
25	48079060	75	26138400	54.37	6067200	348512.00	1600
26	51813540	104	17853600	34.46	6302400	171669.23	1600
27	53382197	88	18271200	34.23	3536000	207627.27	1600
28	54948720	91	26006400	47.33	4981600	285784.62	1600
29	55425393	97	27160800	49.00	5890400	280008.25	1600

Abbreviations: Chr = chromosome; Length: the length of the chromosome; CNVR counts= the number of CNVR; Length of CNVR= the length of CNVR; Coverage= CNVR length / chromosome length\*100%; Max size = the maximum value of CNVR; Average size= the average value of CNVR; Min size= the minimum value of CNVR.

## References

- GAUTIER M, LALOE D, MOAZAMI-GOUDARZI K. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds [J]. *PLoS One*, 2010, 5(9).
- LOZANO A C, DING H, ABE N, et al. Regularized multi-trait multi-locus linear mixed models for genome-wide association studies and genomic selection in crops [J]. *BMC Bioinformatics*, 2023, 24(1): 399.
- ZHUANG Z, XU L, YANG J, et al. Weighted Single-Step Genome-Wide Association Study for Growth Traits in Chinese Simmental Beef Cattle [J]. *Genes (Basel)*, 2020, 11(2).
- ZHANG L, LIU J, ZHAO F, et al. Genome-wide association studies for growth and meat production traits in sheep [J]. *PLoS One*, 2013, 8(6): e66569.
- AL-MAMUN H A, KWAN P, CLARK S A, et al. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight [J]. *Genet Sel Evol*, 2015, 47(1): 66.
- ZHANG Z, PENG M, WEN Y, et al. Copy number variation of EIF4A2 loci related to phenotypic traits in Chinese cattle [J]. *Vet Med Sci*, 2022, 8(5): 2147-2156.
- LIU Y, WANG J, LI S, et al. Research progress of genomic copy number variation in livestock genetics and breeding [ J ]. *China Livestock and Poultry Breeding Industry*, 2024, 20 ( 11 ) : 35-50.
- POS O, RADVANSZKY J, BUGLYO G, et al. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects [J]. *Biomed J*, 2021, 44(5): 548-559.
- XU Y, JIANG Y, SHI T, et al. Whole-genome sequencing reveals mutational landscape underlying phenotypic differences between two widespread Chinese cattle breeds [J]. *PLoS One*, 2017, 12(8): e0183921.
- BENFICA L F, BRITO L F, DO BEM R D, et al. Genome-wide association study between copy number variation and feeding behavior, feed efficiency, and growth traits in Nelore cattle [J]. *BMC Genomics*, 2024, 25(1): 54.
- WANG K, ZHANG Y, HAN X, et al. Effects of Copy Number Variations in the Plectin (PLEC) Gene on the Growth Traits and Meat Quality of Leizhou Black Goats [J]. *Animals (Basel)*, 2023, 13(23).
- WANG X, CAO X, WEN Y, et al. Associations of ORMDL1 gene copy number variations with growth traits in four Chinese sheep breeds [J]. *Arch Anim Breed*, 2019, 62(2): 571-578.
- FENG Z, LI X, CHENG J, et al. Copy Number Variation of the PIGY Gene in Sheep and Its Association Analysis with Growth Traits [J]. *Animals (Basel)*, 2020, 10(4).
- QIU Y, DING R, ZHUANG Z, et al. Genome-wide detection of CNV regions and their potential association with growth and fatness traits in Duroc pigs [J]. *BMC Genomics*, 2021, 22(1): 332.
- XU Q. Genome-wide association analysis of important traits in Inner Mongolia cashmere goats and design of low-density SNP chip [ D ], 2024.
- WANG Y, DING X, TAN Z, et al. Genome-Wide Association Study of Piglet Uniformity and Farrowing Interval [J]. *Front Genet*, 2017, 8: 194.

17. CHEN S, ZHOU Y, CHEN Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, 34(17): i884-i890.
18. LI H, DURBIN R. Fast and accurate short read alignment with Burrows–Wheeler transform [J]. *Bioinformatics*, 2009, 25(14): 1754-1760.
19. LI H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data [J]. *Bioinformatics*, 2011, 27(21): 2987-2993.
20. GAO Y, JIANG J, YANG S, et al. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing [J]. *BMC Genomics*, 2017, 18(1): 265.
21. ZHOU J, LIU L, LOPDELL T J, et al. HandyCNV: Standardized Summary, Annotation, Comparison, and Visualization of Copy Number Variant, Copy Number Variation Region, and Runs of Homozygosity [J]. *Front Genet*, 2021, 12: 731355.
22. DING R, YANG M, QUAN J, et al. Single-Locus and Multi-Locus Genome-Wide Association Studies for Intramuscular Fat in Duroc Pigs [J]. *Front Genet*, 2019, 10: 619.
23. YANG Q, CUI J, CHAZARO I, et al. Power and type I error rate of false discovery rate approaches in genome-wide association studies [J]. *BMC Genet*, 2005, 6 Suppl 1(Suppl 1): S134.
24. D. TURNER S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots [J]. *Journal of Open Source Software*, 2018, 3(25).
25. AHLAWAT S, VASU M, CHOUDHARY V, et al. Comprehensive evaluation and validation of optimal reference genes for normalization of qPCR data in different caprine tissues [J]. *Mol Biol Rep*, 2024, 51(1): 268.
26. QUINLAN A R, HALL I M. BEDTools: a flexible suite of utilities for comparing genomic features [J]. *Bioinformatics*, 2010, 26(6): 841-842.
27. RIVALS I, PERSONNAZ L, TAING L, et al. Enrichment or depletion of a GO category within a class of genes: which test? [J]. *Bioinformatics*, 2007, 23(4): 401-407.
28. BUNIELLO A, MACARTHUR J A L, CERESO M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019 [J]. *Nucleic Acids Res*, 2019, 47(D1): D1005-d1012.
29. MANOLIO T A, COLLINS F S, COX N J, et al. Finding the missing heritability of complex diseases [J]. *Nature*, 2009, 461(7265): 747-753.
30. GRAHAM S E, CLARKE S L, WU K H, et al. The power of genetic diversity in genome-wide association studies of lipids [J]. *Nature*, 2021, 600(7890): 675-679.
31. GHASEMI M, ZAMANI P, VATANKHAH M, et al. Genome-wide association study of birth weight in sheep [J]. *Animal*, 2019, 13(9): 1797-1803.
32. WANG X, ZHENG Z, CAI Y, et al. CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations [J]. *Gigascience*, 2017, 6(12): 1-12.
33. KANG X, LI M, LIU M, et al. Copy number variation analysis reveals variants associated with milk production traits in dairy goats [J]. *Genomics*, 2020, 112(6): 4934-4937.
34. CHEN C, LIU C, XIONG X, et al. Copy number variation in the MSR3 gene enlarges porcine ear size through a mechanism involving miR-584-5p [J]. *Genet Sel Evol*, 2018, 50(1): 72.
35. GUANG-XIN E, YANG B G, ZHU Y B, et al. Genome-wide selective sweep analysis of the high-altitude adaptability of yaks by using the copy number variant [J]. *3 Biotech*, 2020, 10(6): 259.
36. ZHANG L, WANG F, GAO G, et al. Genome-Wide Association Study of Body Weight Traits in Inner Mongolia Cashmere Goats [J]. *Front Vet Sci*, 2021, 8: 752746.
37. LIU Y, KHAN A R, GAN Y. C2H2 Zinc Finger Proteins Response to Abiotic Stress in Plants [J]. *Int J Mol Sci*, 2022, 23(5).
38. ITO M. Function and molecular evolution of mammalian Sox15, a singleton in the SoxG group of transcription factors [J]. *Int J Biochem Cell Biol*, 2010, 42(3): 449-452.
39. YAMADA K, KANDA H, TANAKA S, et al. Sox15 enhances trophoblast giant cell differentiation induced by Hand1 in mouse placenta [J]. *Differentiation*, 2006, 74(5): 212-221.
40. WU X, LIM, LI Y, et al. Fibroblast growth factor 11 (FGF11) promotes non-small cell lung cancer (NSCLC) progression by regulating hypoxia signaling pathway [J]. *J Transl Med*, 2021, 19(1): 353.

41. CHO J H, KIM K, CHO H C, et al. Silencing of hypothalamic FGF11 prevents diet-induced obesity [J]. *Mol Brain*, 2022, 15(1): 75.
42. LEE K W, JEONG J Y, AN Y J, et al. FGF11 influences 3T3-L1 preadipocyte differentiation by modulating the expression of PPARgamma regulators [J]. *FEBS Open Bio*, 2019, 9(4): 769-780.
43. JIANG T, SU D, LIU X, et al. Transcriptomic Analysis Reveals Fibroblast Growth Factor 11 (FGF11) Role in Brown Adipocytes in Thermogenic Regulation of Goats [J]. *Int J Mol Sci*, 2023, 24(13).
44. ENGLISH J, OROFINO J, CEDERQUIST C T, et al. GPS2-mediated regulation of the adipocyte secretome modulates adipose tissue remodeling at the onset of diet-induced obesity [J]. *Mol Metab*, 2023, 69: 101682.
45. DRARENI K, BALLAIRE R, BARILLA S, et al. GPS2 Deficiency Triggers Maladaptive White Adipose Tissue Expansion in Obesity via HIF1A Activation [J]. *Cell Rep*, 2018, 24(11): 2957-2971 e2956.
46. HUANG X, WANG Z, LI D, et al. Study of microRNAs targeted Dvl2 on the osteoblasts differentiation of rat BMSCs in hyperlipidemia environment [J]. *Journal of cellular physiology*, 2018, 233(9): 6758-6766.
47. ZHOU Y, SHANG H, ZHANG C, et al. The E3 ligase RNF185 negatively regulates osteogenic differentiation by targeting Dvl2 for degradation [J]. *Biochem Biophys Res Commun*, 2014, 447(3): 431-436.
48. PAN H, XU R, ZHANG Y. Role of SPRY4 in health and disease [J]. *Front Oncol*, 2024, 14: 1376873.
49. LI N, CHEN Y, WANG H, et al. SPRY4 promotes adipogenic differentiation of human mesenchymal stem cells through the MEK-ERK1/2 signaling pathway [J]. *Adipocyte*, 2022, 11(1): 588-600.
50. STEEN H C, GAMERO A M. STAT2 phosphorylation and signaling [J]. *Jakstat*, 2013, 2(4): e25790.
51. YANG Y, LUO D, SHAO Y, et al. circCAPRIN1 interacts with STAT2 to promote tumor progression and lipid synthesis via upregulating ACC1 expression in colorectal cancer [J]. *Cancer Commun (Lond)*, 2023, 43(1): 100-122.

**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.