BMC Genomics

RESEARCH ARTICLE

Open Access

Identification and population genetic analyses of copy number variations in six domestic goat breeds and Bezoar ibexes using next-generation sequencing

Jiazhong Guo[1†], Jie Zhong[1†], George E. Liu[2], Liu Yang[1], Li Li[1], Guangling Chen[3], Tianzeng Song[4] and Hongping Zhang[1*]

Abstract

Background: Copy number variations (CNVs) are a major form of genetic variations and are involved in animal domestication and genetic adaptation to local environments. We investigated CNVs in the domestic goat (*Capra hircus*) using Illumina short-read sequencing data, by comparing our lab data for 38 goats from three Chinese breeds (Chengdu Brown, Jintang Black, and Tibetan Cashmere) to public data for 26 individuals from three other breeds (two Moroccan and one Chinese) and 21samples from Bezoar ibexes.

Results: We obtained a total of 2394 CNV regions (CNVRs) by merging 208,649 high-confidence CNVs, which spanned ~ 267 Mb of total length and accounted for 10.80% of the goat autosomal genome. Functional analyses showed that 2322 genes overlapping with the CNVRs were significantly enriched in 57 functional GO terms and KEGG pathways, most related to the nervous system, metabolic process, and reproduction system. Clustering patterns of all 85 samples generated separately from duplications and deletions were generally consistent with the results from SNPs, agreeing with the geographical origins of these goats. Based on genome-wide F_{ST} at each CNV locus, some genes overlapping with the highly divergent CNVs between domestic and wild goats were mainly enriched for several immunity-related pathways, whereas the genes overlapping with the highly differentiated CNVs between highland and lowland goats were mainly related to vitamin and lipid metabolism. Remarkably, a 507-bp deletion at ~ 14 kb downstream of *FGF5* on chromosome 6 showed highly divergent ($F_{ST} = 0.973$) between the highland and lowland goats. Together with an enhancer activity of this sequence shown previously, the function of this duplication in regulating fiber growth deserved to be further investigated in detail.

Conclusion: We generated a comprehensive map of CNVs in goats. Many genetically differentiated CNVs among various goat populations might be associated with the population characteristics of domestic goat breeds.

Keywords: Goat, Copy number variations, Short-read sequencing, Comparative genomics

---

* Correspondence: zhp@sicau.edu.cn
†Jiazhong Guo and Jie Zhong contributed equally to this work.
¹College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China
Full list of author information is available at the end of the article

Background

As a major class of structural variations (SVs) that complement to single nucleotide variations (SNVs), copy number variations (CNVs) refer to duplications, deletions, and insertions of DNA sequences ≥50 bp in size between individuals within a species [1]. A growing body of work demonstrates important effects of CNVs on phenotypic variations of Mendelian and quantitative traits in domestic animals by various molecular mechanisms, such as gene dosage, gene disruption, and gene fusion. For example, a previous work demonstrated that a massive amplification of a duplicated sequence in intron 1 of *SOX5* causes Pea-comb in chickens [2]. An inverted duplication and junction of two distinct regions of the chicken genome disrupt long-range cis-regulatory elements of *EDN3* and thereby result in the Silkie phenotype (i.e., dermal hyperpigmentation) [3]. A genome-wide association study and whole-genome sequencing identified a significant correlation between the degree of white spotting and a 1 Mb copy number region harboring *EDNRA* in Boer goats [4]. A 12.1-kb large deletion at the *HMGA2* locus results in small size in dwarf rabbits [5]. A 660-Kb deletion harboring four genes affects fertility and milk production traits antagonistically in Nordic red cattle [6], illustrating the occurrence of balancing selection in livestock. However, there are few studies concerning genome-wide characteristics of CNVs in goats so far [7, 8].

Similar to SNPs, the population genetic properties of CNVs could reflect genomic changes driven by evolutionary factors (e.g., natural/artificial selection, demographic history, and genetic drift) [9–13]. A well-known example of positive selection is that the copy number of the salivary amylase (*AMY1*) gene highly differs between the human populations with high-starch diets and those with traditionally low-starch diets [9]. Similarly, a genomic region harboring *AMY2B* was identified as a signature of dog domestication and modern dogs show high *AMY2B* copy numbers compared to wolves [14], demonstrating important roles of CNVs during animal domestication. A recent study of genome-wide CNVs in eight cattle breeds detected many lineage-differentiated CNVs that might be due to the selection for different traits (e.g., parasite resistance, body size, and fertility) [15]. In domestic goats, a genomic comparison with Bezoar showed that many copy number variable genes were associated with coat color, immune response, and production traits, which provided a clue for goat domestication [16]. Based on the ADAPTmap data, Liu et al. reported a population differentiation in CNV across different geographical areas around the world, suggesting the population history of different domestic goat breeds [7]. However, it is still limited with respect to the population genetic analyses using large-scale CNV data in domestic goats.

Compared to SNP chips and CGH arrays, next-generation sequencing (NGS) or short-read sequencing technologies have been matured rapidly in the past years, allowing us to identify and genotype CNVs more comprehensively and precisely at the genome-wide level [17, 18]. In this study, we aimed to investigate the genetic diversity, population structure, and population differentiation in domestic goats, based on CNVs detected using the whole-genome sequencing data from our previous work as well as the public data.
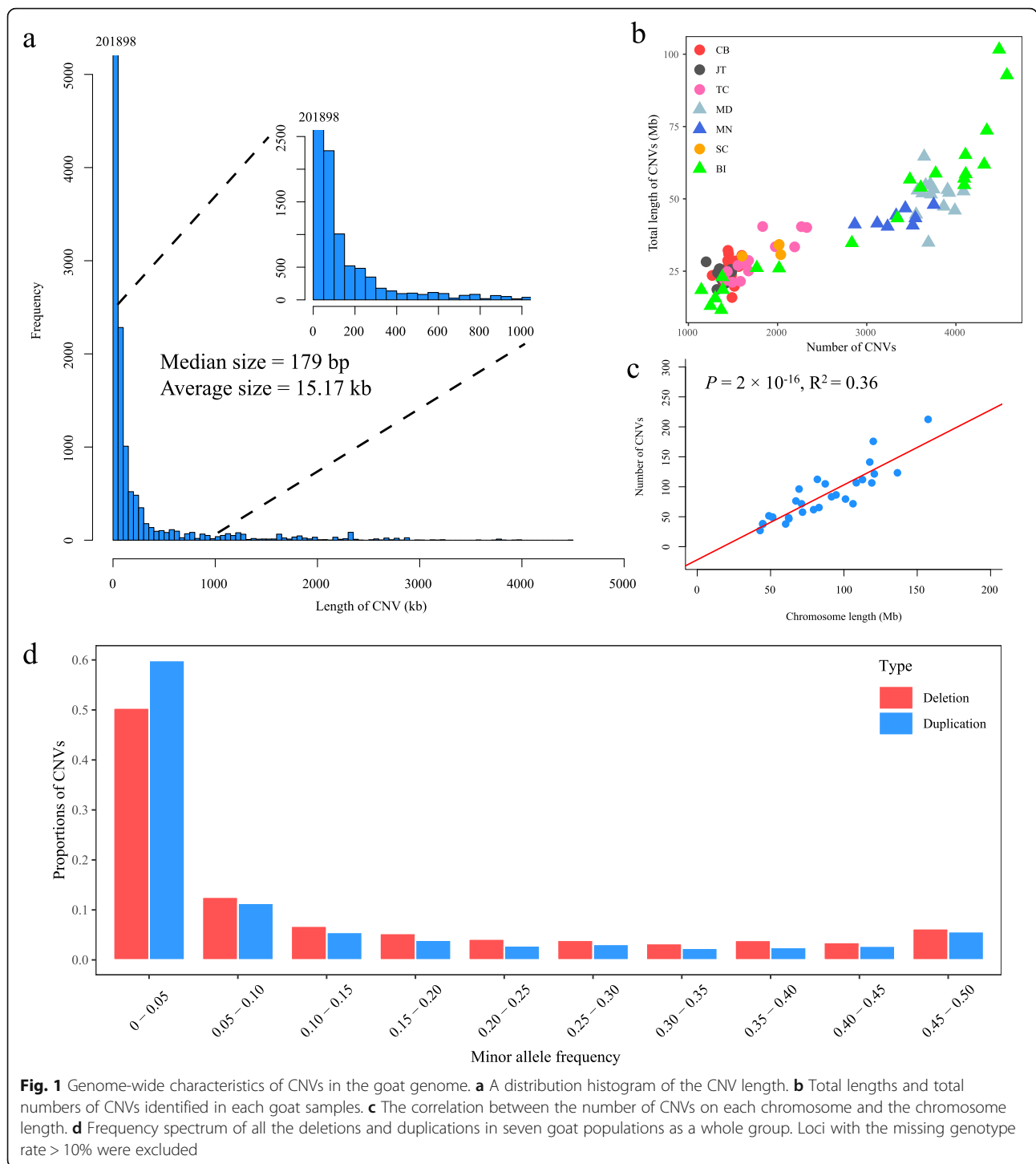
Results

CNVs distribute ubiquitously in the goat genome

Based on the short-read sequencing data for 85 goats with the coverage depth of 5.25× ~ 15.90× (Additional file 1), we identified a total of 208,649 CNVs, including 17,876 duplications and 190,773 deletions across all autosomes (Additional file 2). As shown in Fig. 1a, the size of all the CNVs showed an L-shaped distribution (the median size = 179 bp, the average size = 15.17 kb), and the CNVs with ≤50 kb in size accounted for 96.76% (201,898) of the total CNV number. At the individual level (Fig. 1b), we discovered, on average, 2455 CNVs per goat genome with a range from 1145 to 4572, and the identified CNVs in each animal covered to 0.47 to 4.12% of the goat autosomal genome (2.47 Gb). It is noted that the number of identified CNVs in each goat appeared to increase with sequence depth in general. Based on a linear regression analysis, there was a strong positive relationship ($P < 2.0 \times 10^{-16}$, F-test) between the number of CNV events on each chromosome and the chromosome length (Fig. 1c). We also constructed site frequency spectra of duplications and deletions in seven goat populations as a whole (Fig. 1d). The loci with MAF ≤ 0.05 accounted for 59.88 and 50.38% of the total duplication and deletion loci, respectively, suggesting that many CNV loci were skewed towards rare CNVs.

The validation of CNVs by PCR and a second independent run of deep sequencing

We performed quantitative real-time PCR (qPCR) to test the accuracy for the CNVs detected by next-generation sequencing. The PCR results for all the seven deletions (~ 14 kb downstream of *FGF5*, ~ 470 kb downstream of *GABRB3*, *FOXO3*, *IGLON5*, *PRSS22*, *RARB*, and *KCNU1*) and one duplication harboring the complete *ASIP* and *AHCY* genes were all in agreement with our NGS prediction (Additional file 3).

To validate the detection accuracy of CNVs, we also carried out a second independent run of whole-genome deep sequencing for nine sampled goats (i.e., three individuals in each of the CB, JT, and TC goat breeds). After alignment against the reference genome

Fig. 1 Genome-wide characteristics of CNVs in the goat genome. **a** A distribution histogram of the CNV length. **b** Total lengths and total numbers of CNVs identified in each goat samples. **c** The correlation between the number of CNVs on each chromosome and the chromosome length. **d** Frequency spectrum of all the deletions and duplications in seven goat populations as a whole group. Loci with the missing genotype rate > 10% were excluded

(ARS1), the second runs of whole-genome sequencing data of the nine goats showed coverage depth of 25× at least (Additional file 1). The concordance rates for nine samples ranged from 68.23% (duplications in JT9 using Pindel) to 97.56% (duplications in CB10 using DELLY), suggesting the reliabilities of the CNVs identified above (Additional file 4).

CNVRs were enriched with genes mainly related to nervous system and immune system

By merging overlapping CNVs across all the 85 goat samples, we finally obtained 2394 CNVRs (1722 deletions, 148 duplications, and 524 complex CNVRs) in all seven goat populations as a whole (Additional file 5). These CNVRs covered 266.71 Mb (10.80%) of the goat

Guo *et al. BMC Genomics*   (2020) 21:840

Page 4 of 13

autosomal genome (2.47 Gb), and the median and average size of CNVRs was 2415 bp and 111.41 kb, respectively. The total number and length of the CNVRs identified in each population were provided in Additional file 5.

From genome annotation (ARS1), we obtained a total of 2322 genes (gene symbols) overlapping with the CNVRs (hereafter referred to as CNV genes) in the seven goat populations (Fig. 2a and Additional file 5). The functional enrichment analysis showed that these CNV genes were significantly enriched in 57 GO terms and KEGG pathways ($P_{adj} < 0.05$), which were mainly related to sensory perception of smell system (e.g., olfactory receptor activity, detection of chemical stimulus involved in sensory perception of smell, and sensory perception of smell), nervous system (i.e., sensory perception, nervous system process) and immune system (e.g., defense response to virus, scavenger receptor activity, and natural killer cell mediated cytotoxicity) (Fig. 2b and Additional file 6). Particularly, the term of olfactory receptor activity included 86 OR genes (Ensemble ID) associated with CNVRs (Additional file 5).
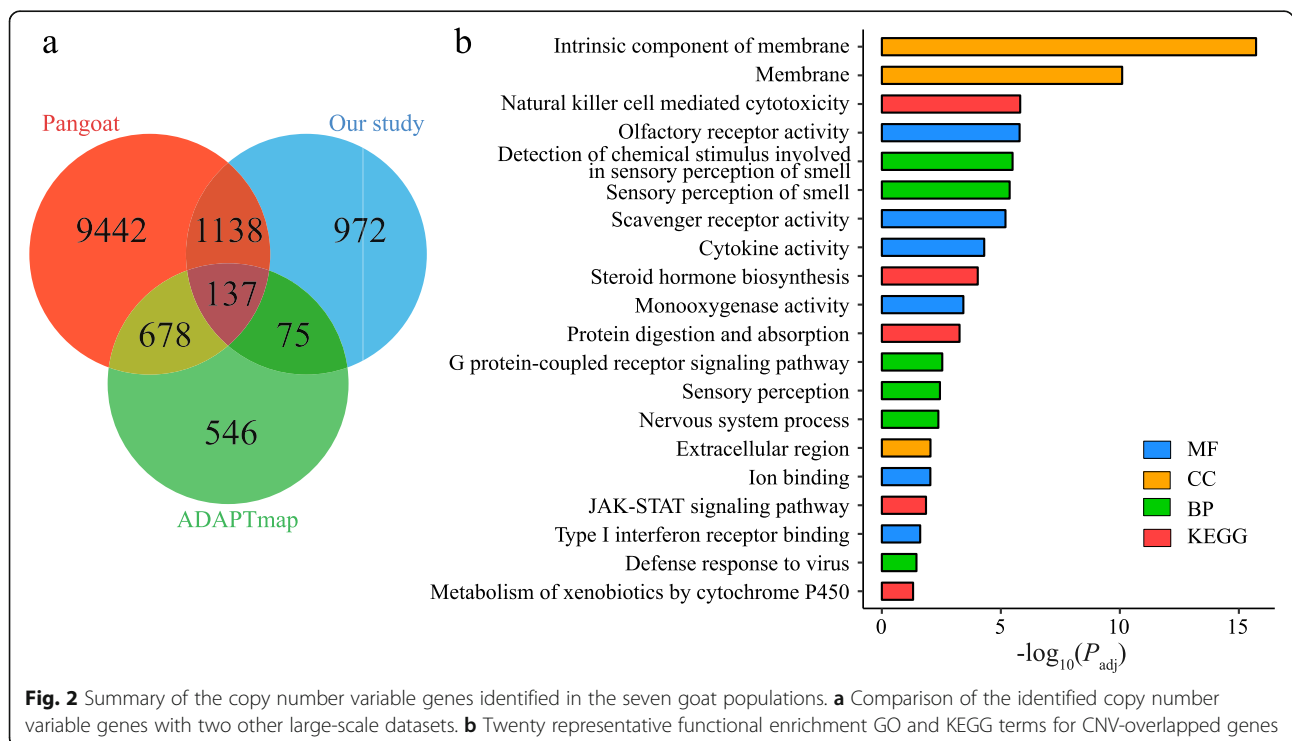
To further examine the credibility of the CNVRs detected in this study, we compared the CNV genes with those in the Goat Pan-genome database (hereafter referred to as Pangoat for simplicity) and a previous study [7] that used SNP data from the ADAPTmap project (see details in Methods). As shown in Fig. 2a, a total of 137 CNV genes (e.g., *ADAMTS20, ASIP, DGAT1*, and *AP2A2*) were shared among three datasets, suggesting that these CNVRs were common in a variety of goat populations. Furthermore, 54.91% of the total 2322 CNV genes (e.g., *ADCY5, CCND1, SOX13, SPATA5*, and *TLR6*) were present in the Goat Pan-genome database, whereas only 9.13% of the total CNV genes (e.g., *EXOSC4, IRF7*, and *MAPK15*) were shared with the CNV genes identified from the ADAPTmap project.

Considering the importance of coat color in goats, we attempted to examine whether our CNVRs overlap with pigmentation genes. Interestingly, we found a large duplication with 154,674 bp in length that encompasses the complete *ASIP* and *AHCY* genes and a partial sequence of *ITCH* at 63,226,821–63,381,495 bp on chromosome 13. This duplication was detected not only in white breeds (SC) but also in black goats (JT). Based on genotyping information from the NGS data, all the four SC goats (white coat color) were heterozygous for the two-copy *ASIP* allele (hereafter referred to as CN2-ASIP), while the CN2-ASIP allele showed a frequency of 11.11% in JT.

The population structure analysis of goat populations

The genome-wide deletion or duplication genotyping information allowed us to examine the population structure in seven goat populations. In terms of deletions, the PCA analysis using deletions (variance explained = 8.86 and 5.83% for PC1 and PC2, respectively) and duplications (variance explained = 7.41 and 4.38% for PC1 and

Fig. 2 Summary of the copy number variable genes identified in the seven goat populations. **a** Comparison of the identified copy number variable genes with two other large-scale datasets. **b** Twenty representative functional enrichment GO and KEGG terms for CNV-overlapped genes
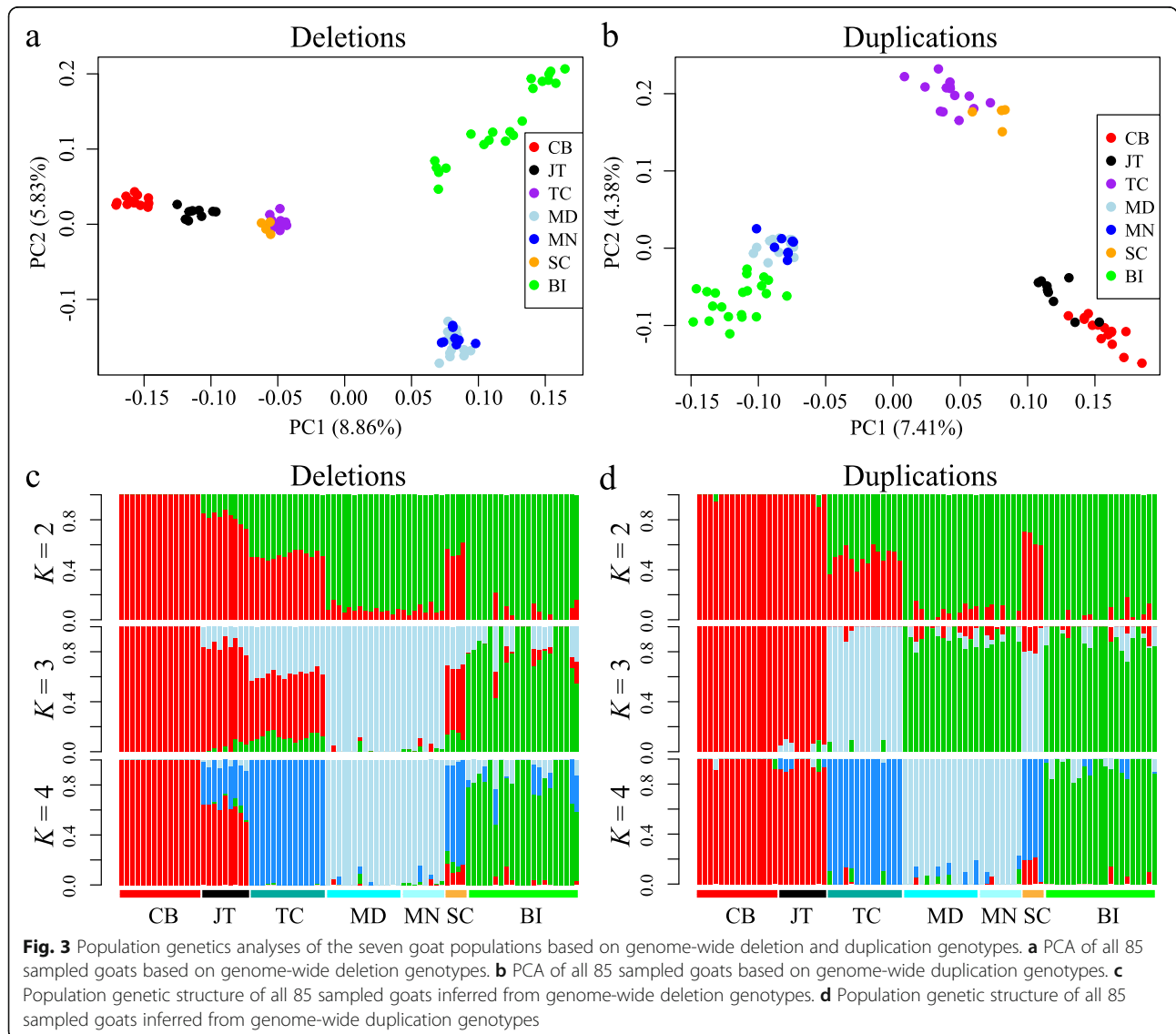
PC2, respectively) showed that the 85 goat samples mainly clustered into three large groups: Chinese breeds (i.e., CB, JT, TC, and SC), African breeds (i.e., MD and MN), and Bezoar ibexes (Fig. 3a). Based on duplications (Fig. 3b), the both Moroccan breeds and Bezoar ibexes clustered closely, whereas four Chinese breeds can be divided into two large groups: TC and SC (i.e., Cashmere goats), CB and JT (both breeds were from the Sichuan Basin). When the genetic admixture analysis was conducted by assuming different numbers of ancestral populations (K = 2 ∼ 7), the results generated from deletions showed that K = 3 was the most plausible number of genetically distinct groups (Fig. 3c). In this scenario, the ancestors of the 85 sampled goats can be resolved three populations: a wild goat population, a Moroccan goat population, and a Chinese goat population. According to duplications, all the sampled goats were thought to be

most likely (K = 2) from two ancestral populations: domestic population and a wild population (Fig. 3d).

The differentiated CNVs between domestic goats and Bezoar ibexes were mainly related to immune system

To find the CNVs that were genetically differentiated between domestic goats (i.e., MD and MN) and Bezoar ibexes, we carried out a genome-wide scan by estimating the genome-wide fixation index (i.e., F_{ST}) at each CNV locus. To reduce the impact of the differences in sequencing depths, here we only selected 13 Bezoar ibexes that had similar sequencing depths with the goats from MD and MN breeds. The genome-wide average F_{ST} was 0.068 (median $F_{ST} = 0.009$) across a total of 8792 CNV loci (6544 deletions and 2248 duplications). Based on the 5% right tail (i.e., $F_{ST} \geq 0.318$) of the empirical F_{ST} distribution, 450 outlier loci that overlapped with 194

Fig. 3 Population genetics analyses of the seven goat populations based on genome-wide deletion and duplication genotypes. **a** PCA of all 85 sampled goats based on genome-wide deletion genotypes. **b** PCA of all 85 sampled goats based on genome-wide duplication genotypes. **c** Population genetic structure of all 85 sampled goats inferred from genome-wide deletion genotypes. **d** Population genetic structure of all 85 sampled goats inferred from genome-wide duplication genotypes
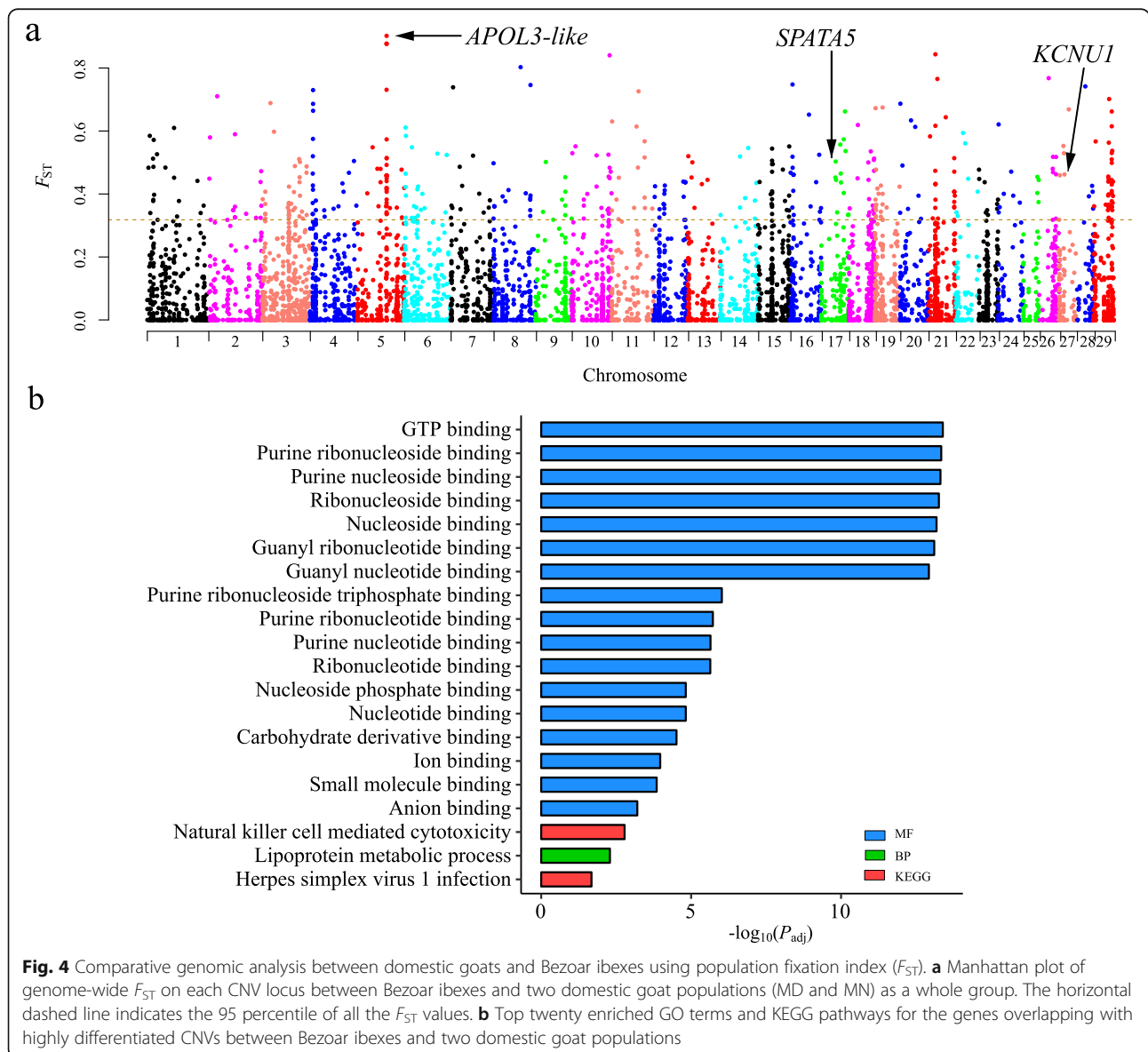
genes (Ensemble ID) were thought to be highly divergent between domestic goats and Bezoar ibex (Fig. 4a and Additional file 7). The functional analysis showed that the genes overlapping with the differentiated CNVs were significantly enriched ($P_{adj} < 0.05$) in the functional terms of immune system (e.g., natural killer cell mediated cytotoxicity, Herpes simplex virus 1 infection, and type I interferon receptor binding), metabolism (e.g., lipoprotein metabolic process) and digestion (carbohydrate derivative binding) (Fig. 4b and Additional file 8). The genome-wide distribution of global F_{ST} revealed the highest differentiated hit ($F_{ST} = 0.903$) was a 26,120-bp duplication (chromosome 5: 73,049,912–73,076,032 bp) that overlapped with the gene *APOL3-like* (Fig. 4a and Additional file 7). Additionally, we identified a 55-bp deletion (chromosome 17: 36,494,821–36,494,876 bp) in

SPATA5 and an 83-bp deletion (chromosome 27: 13,435, 433–13,435,516 bp) in *KCNU1*.

The differentiated CNVs between highland goats and lowland goats included a 507-bp deletion at the ~ 14 kb downstream of *FGF5*

To identify the CNV loci that underlie genetic adaptation to high-altitude, we conducted genome-wide estimates of single-maker F_{ST} between a highland breed (i.e., TC) and two lowland breeds (i.e., CB and JT) as a whole control group. The genome-wide average F_{ST} was 0.064 (median $F_{ST} = 0.012$) across a total of 7896 CNV loci (5873 deletions and 2023 duplications). Based on the 5% right tail (i.e., $F_{ST} \geq 0.282$) of the empirical F_{ST} distribution, 398 outlier loci that overlapped with 317 genes (Ensemble ID) were thought to be highly divergent

Fig. 4 Comparative genomic analysis between domestic goats and Bezoar ibexes using population fixation index (F_{ST}). **a** Manhattan plot of genome-wide F_{ST} on each CNV locus between Bezoar ibexes and two domestic goat populations (MD and MN) as a whole group. The horizontal dashed line indicates the 95 percentile of all the F_{ST} values. **b** Top twenty enriched GO terms and KEGG pathways for the genes overlapping with highly differentiated CNVs between Bezoar ibexes and two domestic goat populations

between TC and the lowland breeds (Fig. 5a and Additional file 8). The functional terms significantly enriched for the differentiated CNV genes ($P_{adj} < 0.05$) were mainly related to vitamin and lipid metabolism (e.g., retinol metabolism, folate biosynthesis, and steroid hormone biosynthesis), immune system (e.g., scavenger receptor activity and natural killer cell mediated cytotoxicity), and reproduction (i.e., prolactin signaling pathway) (Fig. 5b and Additional file 9). For example, the CNV genes including eight cytochrome-P450 (*CYP450*) genes (e.g., *CYP2S1*, *CYP2C31*, and *CYP3A24*) were significantly enriched in monooxygenase activity. Additionally, six olfactory-related genes (e.g., *OLFM3*, *OR6C1*, and *OR6C3*) overlapped with the differentiated CNVs.

Among all the analyzed CNV loci, the highest differentiated hit ($F_{ST} = 0.973$) was a 507-bp deletion on chromosome 6 that located at 95,454,681–95,455,188 bp

and ~ 14 kb downstream of *FGF5* that is a major regulator affecting hair growth and development (Fig. 5a and Fig. 6a). Interestingly, an almost same deletion (a 504-bp deletion at 95,454,685–95,455,188 bp of chromosome 6) was recently identified in Chinese Cashmere goats with a high frequency [19]. We also examined the frequency patterns of this deletion among the analyzed goat population in this study. Based on the NGS data, this deletion was fixed (frequency = 100%) in TC ($n = 14$) and SC ($n = 4$), whereas the goat breeds with short hair do not carry this deletion except for one goat in JT (Fig. 6b).

Discussion

The characteristics of CNVs in the goat genome

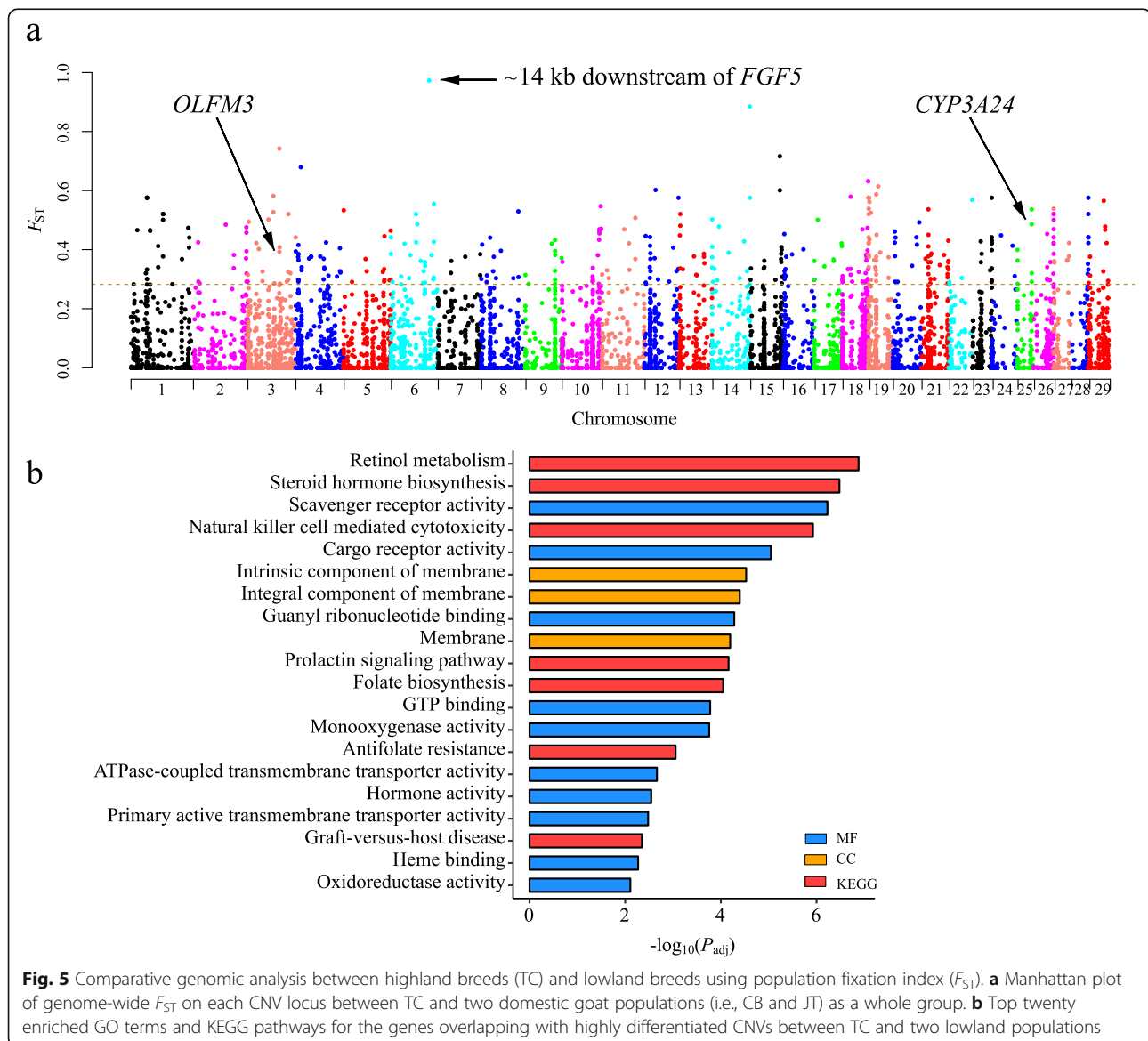In this study, we used the whole genome short-read sequencing data of 85 individuals from six domestic goat

Fig. 5 Comparative genomic analysis between highland breeds (TC) and lowland breeds using population fixation index (F_{ST}). **a** Manhattan plot of genome-wide F_{ST} on each CNV locus between TC and two domestic goat populations (i.e., CB and JT) as a whole group. **b** Top twenty enriched GO terms and KEGG pathways for the genes overlapping with highly differentiated CNVs between TC and two lowland populations
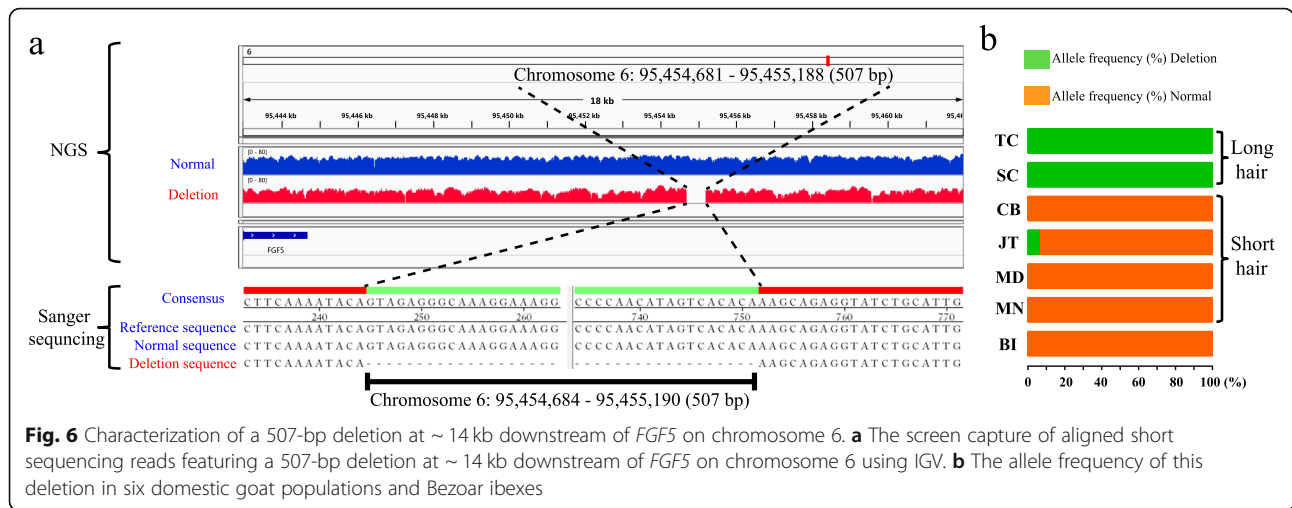
Fig. 6 Characterization of a 507-bp deletion at ~ 14 kb downstream of *FGF5* on chromosome 6. **a** The screen capture of aligned short sequencing reads featuring a 507-bp deletion at ~ 14 kb downstream of *FGF5* on chromosome 6 using IGV. **b** The allele frequency of this deletion in six domestic goat populations and Bezoar ibexes

populations and Bezoar ibexes to generate a detailed CNV map in the goat genome. In total, 2394 CNVRs were obtained by merging 208,649 high-confidence CNVs across all the goat samples, covering 10.80% of the goat autosomal genome. This is in line with the total coverage of CNVs in humans (12, 13%) [20, 21], cattle (6 and 8.9%) [22, 23], sheep (6.9%) [24], and yak (6.2%) [25]. For example, similar to the distribution in the cattle genome [22], the number of CNV events on each chromosome increased with the chromosome length. Compared to previous genome-wide studies [7, 8], we detected many more CNVs in the goat genome, which can be attributed to the differences in genetic data types (e.g., SNP chips and whole-genome sequencing) and the sampled populations. Notably, high amounts of CNVs with small sizes (~ 100 kb or less) resulted in an L-shaped distribution of the size of all CNVs identified here, supported by the results in other livestock breeds (e.g., cattle [22], pigs [26], and sheep [24]) and humans [27, 28]. Theoretically, CNVs in larger sizes (e.g., > 100 kb) change the imbalances or dosage effects of more genes in comparisons with the CNVs in smaller sizes [29], which results in that large CNVs are generally more deleterious at the phenotypic level and are thus relatively rare in a population.

Comparisons with previous large-scale detections of CNVRs in goats

Many genes (e.g., *ADAMTS20*, *ASIP*, *DGAT1*, and *AP2A2*) overlapping with the CNVs identified in this work are also present in the Pangoat database, supporting our results and these CNVs are likely common in domestic goats (Fig. 2a). However, there were fewer shared CNV genes (e.g., *ADAMTS20*, *ASIP*, and *DGAT1*) between our work and the result obtained from the ADAPTmap data [7], which can be explained by different sampled populations (i.e., the ADAPTmap data

included few samples from East Asia) and suggested that many novel CNVRs were identified in our study. For example, different members of the olfactory gene family that is characterized by frequent duplications and losses in humans [30] and vertebrates (e.g., chimpanzees [20], mouse [21], and fishes [23]) were identified in these goat datasets and this work. The Pangoat database not only collected the latest goat reference genome (ARS1), but also used nine de novo assemblies from seven *Caprini* species (e.g., sheep, argali sheep, mouflon sheep, and ibex) to construct the CNVR dataset. It additionally included multiple whole genome resequencing data, and thereby result in a larger total number of CNVR than ours.

As mentioned above, several widely distributed CNVRs in domestic goat breeds deserved more attentions, because they overlapped with the genes (e.g., *ASIP*, *DGAT1*, *PRP1*, and *PRP6*) known for their biological functions in livestock. For example, we found that one 154-kb CNV harboring *ASIP* on chromosome 13 was present in a white coated breed (i.e., SC) and two non-white coat breeds (i.e., JT and TC). It is generally recognized that *ASIP* is a key gene regulating the production of pheomelanin and eumelanin in animals [25], and thereby SNPs or short indels within this gene were primarily associated with yellow/red and black/brown coat color in animals including horses [31], pigs [32], and sheep [33]. A 190-kb duplication harboring the *ASIP* and *AHCY* coding regions and the *ITCH* promoter sequence was determined as the casual variants for the dominant white coat color in sheep [34]. However, the duplication harboring *ASIP* was present in five goat breeds with different coat colors (e.g., white, red, and black) from Italy [35], which was consistent with our finding. Furthermore, the copy number of a genomic region containing *ENDRA* was remarkably correlated with decreasing pigmentation in Boer goats [4]. Considering all together, there was no clear association of CNV of

the *ASIP* gene in this study with white coat color. Therefore, future analyses with a large sample size and additional computational pipelines will be required to obtain a definite conclusion.

The encoded protein of *DGAT1* catalyzes the conversion of diacylglycerol and fatty acyl CoA to triacylglycerol, and this gene has been identified as a major locus for milk production traits in Holstein [36]. Similar to the discovery (chromosome 14: 80,740,427−81,602,889 bp) using the ADAPTmap data, a CNVR (chromosome 14: 80,629,343−81,407,250 bp; Additional file 6) covering *DGAT1* was not only found in Moroccan breeds from Africa, but also was present in the Chinese goat breeds of this study. The other examples are described below, including prolactin (PRL), a key reproductive hormone, has the functions of regulating ovarian function, maintaining pregnant corpus luteum and promoting fetal development [37]. The proliferin-related protein (PRP), as a placental-specific hormone, is mainly involved in the PRL signaling pathway [38]. We also found *PRP1* and *PRP6* genes, whose duplications were recently reported to be associated with high-fecundity in Laoshan dairy goats [39].

Population genetic analyses based on CNVs

In this study, the PCA and population admixture analyses, based on the genome-wide duplication and deletion genotypes, display that all goat samples cluster into different groups, generally consistent with the geographical origins of the goat samples and the results generated from SNPs in our previous work [40]. Our population structure analyses demonstrated that partitioning of diversity among modern Chinese goat populations, African goat populations, and Bezoar ibexes, which was in line with the findings from the ADAPTmap data that include 1023 goats from 50 breeds around the world [7]. It is noteworthy that the results obtained from deletions were more similar to those from SNPs, which was also observed in humans [27] and monkeys [41]. Considering the limited locus numbers in our study, this difference might suggest that more genetic sites would result in higher resolutions in the investigations of population relationships.

Domestication-associated CNVs

During domestication mainly driven by artificial selection, livestock species have evolved different phenotypic and genetic characteristics as compared to their wild progenitors, including morphological, behavioral, and physiological traits [42]. In this study, the most differentiated CNV loci between domestic goats and BI overlapped with *APOL3-like*, which might be a paralog of *APOL3* and is adjacent to *APOL3* in the goat genome. CNVs at the *APOL3* locus have been reported in several

cattle breeds [43, 44] and found to be associated with adult cattle stature [45]. Furthermore, some genes overlapping with the highly divergent CNVs between domestic and wild goats were enriched in the immune system, which was consistent with the reports that some immunity-related CNV genes evolved rapidly between domestic and wild goats (*Capra aegagrus*) [16], and between domestic pigs and wild boars [46]. These findings collectively suggested the biological processes underlying the domestication were similar in different livestock species.

Notably, some genes play roles in spermatogenesis and sperm capacitation. For example, SPATA5 is a spermatogenesis-associated factor and is expressed specifically in the spermatogonia and spermatocytes [47]. A large deletion harboring *SPATA5* was associated with multiple malformations and hearing loss in humans [48]. The *KCNU1* gene, also known as *SLO3*, is expressed only in mammalian testis and plays a vital role in male fertility [49]. The CNV overlapping with *KCNU1* was recently reported in pigs [50].

Several interesting CNVs in Tibetan cashmere goats

Although the relevant issues have been extensively investigated using SNPs, there were a few reports on the selection signals that underlie genetic adaptations to high altitude in humans and animals based on CNVs [51–53]. In this study, some genes overlapping with the CNVs, which were highly differentiated between Tibetan goats and lowland goat breeds, deserved more attention because of their known functions. As a superfamily of 56 functional genes, the *CYP450* gene family is mainly involved in drug metabolism and bioactivation in humans [54, 55]. The characteristics of several *CYP450* gene copy number have been investigated in different human populations [54, 56, 57], particularly for *CYP2D6* [58–60]. Interestingly, some *CYP450* genes were found to be related to genetic adaptations for the high altitude in horses [61] and frogs [62], which can be explained by their function as monooxygenases. Based on an CGH array, CNVs harboring *CYP2C* were also identified in a highland sheep breed (i.e., Mongolian sheep) [63], which was consistent with our findings.

Cashmere-related traits (e.g., fiber length) shaped by artificial selection and adaptation to low temperatures are the most important economic traits in Cashmere goat populations on plateaus of Asia. It is widely recognized that *FGF5* is a key regulator for hair length in animals [64–66], and its disruption via the CRISPR/Cas9 system leads to longer fibers in genome-edited goats [67, 68] and sheep [69]. However, the natural causal mutations affecting cashmere-related traits in goats remain elusive. We recently identified one SNP (c.-253G > A) in the 5′-UTR of *FGF5* that could introduce a start codon during translation, and a large allele frequency difference at this site was observed between long hair and short hair goats and goats

Guo *et al. BMC Genomics*   (2020) 21:840

Page 10 of 13

[40]. In this study, we discovered a 507-bp deletion downstream of *FGF5* fixed in Tibetan Cashmere goats (TC) population. Remarkably, a previous study revealed that an almost same deletion (95,454,685–95,455,188 bp on chromosome 6) was present in all Chinese Cashmere goats at a high frequency, and this deletion can function as a enhancer [19], which was validated by a recent study in Chinese Cashmere goats [70]. Therefore, we proposed that both SNP and CNV mutations could contribute to cashmere-related traits in goats by regulating the expression of *FGF5*. However, more functional experiments are needed to fully uncover their biological functions in detail.

Conclusion

We obtained a comprehensive map of CNVs in goats. Many genetically differentiated CNVs among various goat populations might be associated with the population characteristics of domestic goat breeds. Taken together, our findings provide a better understanding of the evolutionary history of domestic goats, and represent a valuable molecular variation resource for future work on examining genetic mechanisms of phenotypic variations and genetic improvement in goats.

Methods

Whole-genome sequencing data analysis

The information of sampled goats and short read alignment against the goat reference genome (ARS1) have been described in our previous work [40] but are included here for completeness. In brief, whole-genome sequencing data of four Chinese breeds (Chengdu Brown (CB, $n = 15$), Jintang Black (JT, $n = 9$), Tibetan Cashmere (TC, $n = 14$), and Shaanbei Cashmere (SC, $n = 4$)), two Moroccan breeds (Draa - MD, $n = 14$ and Northern - MN, $n = 8$), and 21 Bezoar ibexes were analyzed in this study. The goat samples from the CB, JT, and TC populations were collected by ourselves and were from a breeding farm in Dayi County, a breeding farm in Jintang County, and Coqen County in Tibet, respectively, and the animals were released after the sample collection. In contrast, the whole-genome sequencing data of the goat samples from three other domestic goat breeds (i.e., SC, MD, and MN) and Bezoar ibexes were downloaded from NCBI. According to the alignment against the goat reference genome (ARS1) using BWA (v0.7.12) [71], the coverage depth of whole genome sequencing data of 85 goats varied between 5.25× and 15.90× (Additional file 1). To improve accuracy of the identification of CNVs, duplicated reads were removed from raw alignments with the Picard software (v2.10.6) (http://broadinstitute.github.io/picard/), followed by local realignment around existing indels with GATK (v3.8–0) [72].

CNV discovery and genotyping

Three software with different approaches (i.e., split-read, paired-end mapping, and read depth) were applied to detect duplications and deletions (Additional file 9). We firstly employed DELLY [73] and Pindel [74] with default parameters to detect raw CNVs in each goat genome, respectively. Raw CNVs were then filtered by applying the following thresholds: length > 50 bp and "PASS" loci (DELLY); length > 50 bp, variant allele frequency > 0.2, and supported read number ≥ 3 (Pindel). We also used CNVcaller [75] with default parameters to detect raw copy number variation regions (CNVRs) at the population level.

To further reduce redundancy and obtain high-confidence CNVs and CNVRs, we conducted the following quality control pipeline (Additional file 9): (1) After applying the command 'intersect' of BEDTools [76], we only retained the CNVs in each goat genome, if the CNVs detected with DELLY and Pindel overlap in the genomic coordinates and the overlapping sequence length accounted for ≥50% of a CNV size. We further redefined the length and location of these CNVs based on the overlapped sequence. (2) We then merged multiple adjacent CNVs into a CNVR across individuals within a population and at the meta-population level using the command 'merge' of BEDTools if the overlapping sequence was at least > 1 bp along their genomic coordinates [77], and discarded the CNVRs only including one CNV in each population or the meta-population. (3) The CNVRs derived from the above step and the CNVRs generated with CNVcaller were combined into a unique CNVR, if multiple CNVRs overlapped with ≥50% in size of the shortest CNVR. A raw CNVR was moved if it was only generated from one approach. Here, the CNVRs that exclusively include deletions or duplications were defined as deletion CNVRs and duplication CNVRs, respectively, whereas the CNVRs that comprised both deletions and duplications, we defined them as complex CNVRs.

Genome-wide deletion and duplication genotypes were obtained using DELLY [73]. First, all deletion/duplication sites across all samples were merged into a unified site list using the command 'merge' in DELLY. This merged CNV site list was then genotyped with the command 'call' in DELLY. All genotyped samples were merged into a single BCF using the command 'merge' in bcftools [78], and the BCF was converted to VCF with the command 'view' in bcftools. Finally, the genotypes of CNVs located in the unique CNVRs were extracted for downstream analyses with the command 'intersect' in BEDTools.

The validation of CNVs by PCR and a second independent run of deep sequencing

PCR and quantitative PCR (qPCR) have been performed to validate the accuracy for CNVs detected by using

next-generation sequencing. Seven deletions and one duplication were randomly selected for PCR validation using genomic DNA of the same samples. The primer information for PCR validation was provided in Additional file 10. The *MC1R* gene was used as the reference gene for qPCR experiments [8], and we calculated the relative copy number for each selected site using the $2^{-\Delta\Delta CT}$ method.

We also conducted a second run of whole-genome deep sequencing data ($> \sim 25\times$) for nine sampled goats (i.e., three individuals in each of the CB, JT, and TC goat breeds). The alignment of reads and identification of CNVs for these datasets were carried out using the workflow as described above. In addition, the alignments of animals were visualized with Integrative Genomics Viewer (IGV) [79] to confirm the existence of several CNV loci.

CNV-based population genetic analyses

To explore the utility of CNVs in population genetic analyses, the genetic structure analysis was carried out with the ADMIXTURE [80] using genome-wide duplication and deletion genotypes in all sampled goats. Principal component analysis (PCA) implemented in GCTA [81] was also performed to examine population stratification at the individual level. Notably, we employed VCFtools [82] to remove the CNV loci with a minor allele frequency (MAF) lower than 0.05 and the loci with more than 10% missing genotypes at the meta-population level, before conducting the population structure analyses and other downstream population genetic analyses.

Comparative genome analyses of different goat groups were conducted by calculating genome-wide F_{ST} (i.e. Weir and Cockerham's estimator [83]). To find the domestication-related CNVs, the two domestic goat breeds (i.e., MD and MN) were merged to be a whole group. We calculated genome-wide F_{ST} values at each CNV loci (MAF ≥ 0.05) between domestic goats and Bezoar ibexes. To detect the CNVs that are associated with adaptation to high-altitude, we also determined the pairwise F_{ST} at each CNV loci between the highland breed (TC) and the lowland control group of breeds (CB and JT). Based on the empirical F_{ST} distribution, we focused on the top 5% of CNV loci showing extremely high F_{ST} value, and tested whether these "outlier" loci were related to important traits in goats.

Functional enrichment analysis of genes overlapping with CNVs

The genes that were located within or overlapped with CNVs were extracted based on genome coordinates. Functional enrichment analyses for these genes were then carried out using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost) [84].

Comparison of CNVR-overlapped genes with known large-scale CNV datasets

Our CNVRs were compared with two large-scale CNVR datasets: the CNVRs from the Goat Pan-genome database (http://animal.nwsuaf.edu.cn/code/index.php/pan-Goat) and the CNVRs from a previous study that used CaprineSNP50 genotyping data generated by the ADAP Tmap project [7]. The CNVRs in the Pangoat database were obtained from the goat ARS1 genome assembly and other nine de novo assemblies from seven *Caprini* species (including sheep, argali sheep, mouflon sheep, wild goat, ibex, Barbary sheep, and blue sheep) as well as multiple resequencing data. Given the differences between platforms and pipelines, the genes overlapped with CNVRs in each dataset were compared.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-020-07267-6.

Additional file 1: Table S1. Summary of mapping statistics for 85 goat samples and nine high-depth resequenced goats included in this study.

Additional file 2: Table S2. List of CNVs identified in 85 goat genomes included in this study.

Additional file 3: Figure S1. The PCR results of validation experiments for eight CNV loci.

Additional file 4: Figure S2. Results of validation for the identification of CNVs using an independent run of deep sequencing data.

Additional file 5: Table S3. List of CNVRs and genes overlapping with CNVRs identified in this study.

Additional file 6: Table S4. Significantly enriched terms for the genes overlapped with 2394 CNVRs identified in seven goat populations.

Additional file 7: Table S5. List of highly differentiated CNVs between BI and domestic goats and between TC and lowland goats.

Additional file 8: Table S6. Significantly enriched terms for the differentiation CNV genes identified in this study.

Additional file 9: Figure S3. The pipeline used to identify CNVs and CNVRs. Each step was described in detail in the Methods section.

Additional file 10: Table S7. List of the primer information for the validation of CNVs by PCR and qPCR.

Guo *et al. BMC Genomics*   (2020) 21:840

Page 12 of 13

Availability of data and materials

The raw sequencing data of 85 animals for the identification of CNVs in this study are available from the NCBI SRA database (dataset numbers: PRJNA548681, PRJNA422206, PRJEB3136, PRJEB5900, and PRJEB3134). The deep sequencing data of nine goats for the validation in this study are also deposited at NCBI under accession number PRJNA642185.

Ethics approval and consent to participate

In this study, all experimental protocols were approved by the Institutional Animal Care and Use Committee of Sichuan Agricultural University, Sichuan, China (No. DKY-S20123122–2).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China. [2]Animal Genomics and Improvement Laboratory, BARC, Agricultural Research Service, USDA, Beltsville, MD 20705, USA. [3]Sichuan Technical Exchange Center, Science & Technology Department of Sichuan Province, Chengdu 610016, China. [4]Institute of Animal Science, Tibet Academy of Agricultural and Animal Husbandry Sciences, Lhasa 850009, China.

References

1. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. Nature. 2011;470(7332):59–65.
2. Wright D, Boije H, Meadows JRS, Bed'hom B, Gourichon D, Vieaud A, et al. Copy number variation in intron 1 of SOX5 causes the pea-comb phenotype in chickens. PLoS Genet. 2009;5(6):e1000512.
3. Dorshorst B, Molin A-M, Rubin C-J, Johansson AM, Strömstedt L, Pham M-H, et al. A complex genomic rearrangement involving the Endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet. 2011;7(12): e1002412.
4. Menzi F, Keller I, Reber I, Beck J, Brenig B, Schütz E, et al. Genomic amplification of the caprine EDNRA locus might lead to a dose dependent loss of pigmentation. Sci Rep. 2016;6(1):28438.
5. Carneiro M, Hu D, Archer J, Feng C, Afonso S, Chen C, et al. Dwarfism and altered craniofacial development in rabbits is caused by a 12.1 kb deletion at the HMGA2 locus. Genetics. 2017;205(2):955–65.
6. Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbrandtsen B, Karim L, et al. A 660-kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic red cattle: additional evidence for the common occurrence of balancing selection in livestock. PLoS Genet. 2014;10(1):e1004049.
7. Liu M, Zhou Y, Rosen BD, Van Tassell CP, Stella A, Tosser-Klopp G, et al. Diversity of copy number variation in the worldwide goat population. Heredity. 2019;122(5):636–46.
8. Fontanesi L, Martelli PL, Beretti F, Riggio V, Dall'Olio S, Colombo M, et al. An initial comparative map of copy number variations in the goat (Capra hircus) genome. BMC Genomics. 2010;11(1):639.
9. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 2007;39(10):1256–60.
10. Xue Y, Sun D, Daly A, Yang F, Zhou X, Zhao M, et al. Adaptive evolution of UGT2B17 copy-number variation. Am J Hum Genet. 2008;83(3):337–46.
11. Hasin Y, Olender T, Khen M, Gonzaga-Jauregui C, Kim PM, Urban AE, et al. High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. PLoS Genet. 2008;4(11):e1000249.
12. Conrad DF, Hurles ME. The population genetics of structural variation. Nat Genet. 2007;39(7):S30–6.
13. Iskow RC, Gokcumen O, Lee C. Exploring the role of copy number variants in human adaptation. Trends Genet. 2012;28(6):245–57.
14. Axelsson E, Ratnakumar A, Arendt M-L, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. 2013;495(7441):360–4.
15. Xu L, Hou Y, Bickhart DM, Zhou Y, EHa H, Song J, et al. Population-genetic properties of differentiated copy number variations in cattle. Sci Rep. 2016; 6(1):23161.
16. Dong Y, Zhang X, Xie M, Arefnezhad B, Wang Z, Wang W, et al. Reference genome of wild goat (capra aegagrus) and sequencing of goat breeds provide insight into genic basis of goat domestication. BMC Genomics. 2015;16(1):431.
17. Guan P, Sung W-K. Structural variation detection using next-generation sequencing data: a comparative technical review. Methods. 2016;102:36–49.
18. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol. 2019;20(1):117.
19. Cai Y, Fu W, Cai D, Heller R, Zheng Z, Wen J, et al. Ancient genomes reveal the evolutionary history and origin of cashmere-producing goats in China. Mol Biol Evol. 2020;37(7):2099–109.
20. Go Y, Niimura Y. Similar numbers but different repertoires of olfactory receptor genes in humans and chimpanzees. Mol Biol Evol. 2008;25(9):1897–907.
21. Niimura Y, Nei M. Evolutionary changes of the number of olfactory receptor genes in the human and mouse lineages. Gene. 2005;346:23–8.
22. Letaief R, Rebours E, Grohs C, Meersseman C, Fritz S, Trouilh L, et al. Identification of copy number variation in French dairy and beef breeds using next-generation sequencing. Genet Sel Evol. 2017;49(1):77.
23. Niimura Y, Nei M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. Proc Natl Acad Sci. 2005;102(17):6039–44.
24. Yang L, Xu L, Zhou Y, Liu M, Wang L, Kijas JW, et al. Diversity of copy number variation in a worldwide population of sheep. Genomics. 2018; 110(3):143–8.
25. Hubbard JK, Uy JAC, Hauber ME, Hoekstra HE, Safran RJ. Vertebrate pigmentation: from underlying genes to adaptive function. Trends Genet. 2010;26(5):231–9.
26. Jiang J, Wang J, Wang H, Zhang Y, Kang H, Feng X, et al. Global copy number analyses by next generation sequencing provide insight into pig genome variation. BMC Genomics. 2014;15(1):593.
27. Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, et al. Global diversity, population stratification, and selection of human copy-number variation. Science. 2015;349(6253):aab3761.
28. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet. 2006;38(1):75–81.
29. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet. 2009;10(8):551–64.
30. Matsui A, Go Y, Niimura Y. Degeneration of olfactory receptor gene repertoires in primates: no direct link to full trichromatic vision. Mol Biol Evol. 2010;27(5):1192–200.
31. Rieder S, Taourit S, Mariat D, Langlois B, Guérin G. Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). Mamm Genome. 2001;12(6):450–5.
32. Drögemüller C, Giese A, Martins-Wess F, Wiedemann S, Andersson L, Brenig B, et al. The mutation causing the black-and-tan pigmentation phenotype of Mangalitza pigs maps to the porcine ASIP locus but does not affect its coding sequence. Mamm Genome. 2006;17(1):58–66.
33. Rochus CM, Westberg Sunesson K, Jonas E, Mikko S, Johansson AM. Mutations in ASIP and MC1R: dominant black and recessive black alleles segregate in native Swedish sheep populations. Anim Genet. 2019;50(6):712–7.
34. Norris BJ, Whan VA. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Res. 2008;18(8): 1282–93.
35. Fontanesi L, Beretti F, Riggio V, Gómez González E, Dall'Olio S, Davoli R, et al. Copy number variation and missense mutations of the Agouti

signaling protein (*ASIP*) gene in goat breeds with different coat colors. Cytogenet Genome Res. 2009;126(4):333–47.

36. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, et al. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. Genome Res. 2002;12(2):222–31.

37. Gootwine E. Placental hormones and fetal–placental development. Anim Reprod Sci. 2004;82:551–66.

38. Linzer DIH, Fisher SJ. The placenta and the prolactin family of hormones: regulation of the physiology of pregnancy. Mol Endocrinol. 1999;13(6):837–40.

39. Zhang R-Q, Wang J-J, Zhang T, Zhai H-L, Shen W. Copy-number variation in goat genome sequence: a comparative analysis of the different litter size trait groups. Gene. 2019;696:40–6.

40. Guo J, Zhong J, Li L, Zhong T, Wang L, Song T, et al. Comparative genome analyses reveal the unique genetic composition and selection signals underlying the phenotypic characteristics of three Chinese domestic goat breeds. Genet Sel Evol. 2019;51(1):70.

41. Gschwind AR, Singh A, Certa U, Reymond A, Heckel T. Diversity and regulatory impact of copy number variation in the primate *Macaca fascicularis*. BMC Genomics. 2017;18(1):144.

42. Zeder MA. Core questions in domestication research. Proc Natl Acad Sci U S A. 2015;112(11):3191–8.

43. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, et al. Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res. 2012;22(4):778–90.

44. Cozzi MC, Martinez-Ruiz CP, Roman-Ponce SI, Murillo VEV, Utrera ÁR, Montaño-Bermúdez MM, et al. Copy number variants reveal genomic diversity in a Mexican creole cattle population. Livest Sci. 2019;229:194–202.

45. Peng S-J, Cao X-K, Dong D, Liu M, Hao D, Shen X-M, et al. Integrative analysis of APOL3 gene CNV for adult cattle stature. Anim Biotechnol. 2020; 31(5):440–6.

46. Paudel Y, Madsen O, Megens H-J, Frantz LAF, Bosse M, Bastiaansen JWM, et al. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. BMC Genomics. 2013;14(1): 449.

47. Liu Y, Black J, Kisiel N, Kulesz-Martin MF. SPAF, a new AAA-protein specific to early spermatogenesis and malignant conversion. Oncogene. 2000;19(12): 1579–88.

48. Wu M, Zheng X, Wang X, Zhang G, Kuang J. 4q27 deletion and 7q36.1 microduplication in a patient with multiple malformations and hearing loss: a case report. BMC Med Genomics. 2020;13(1):31.

49. Santi CM, Martínez-López P, de la Vega-Beltrán JL, Butler A, Alisio A, Darszon A, et al. The SLO3 sperm-specific potassium channel plays a vital role in male fertility. FEBS Lett. 2010;584(5):1041–6.

50. Stafuzza NB, RMdO S, BdO F, Masuda Y, Huang Y, Gray K, et al. A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive. BMC Genomics. 2019;20(1):321.

51. Xing J, Wuren T, Simonson TS, Watkins WS, Witherspoon DJ, Wu W, et al. Genomic analysis of natural selection and phenotypic variation in high-altitude Mongolians. PLoS Genet. 2013;9(7):e1003634.

52. Lou H, Lu Y, Lu D, Fu R, Wang X, Feng Q, et al. A 3.4-kb copy-number deletion near *EPAS1* is significantly enriched in high-altitude Tibetans but absent from the Denisovan sequence. Am J Hum Genet. 2015;97(1):54–66.

53. Zhang X, Wang K, Wang L, Yang Y, Ni Z, Xie X, et al. Genome-wide patterns of copy number variation in the Chinese yak genome. BMC Genomics. 2016;17(1):379.

54. McGraw J, Waller D. Cytochrome P450 variations in different ethnic populations. Expert Opin Drug Metab Toxicol. 2012;8(3):371–82.

55. Guengerich FP. Cytochrome P450s and other enzymes in drug metabolism and toxicity. AAPS J. 2006;8(1):E101–11.

56. Martis S, Mei H, Vijzelaar R, Edelmann L, Desnick RJ, Scott SA. Multi-ethnic cytochrome-P450 copy number profiling: novel pharmacogenetic alleles and mechanism of copy number variation formation. Pharmacogenomics J. 2013;13(6):558–66.

57. Fukami T, Nakajima M, Yamanaka H, Fukushima Y, McLeod HL, Yokoi T. A novel duplication type of CYP2A6 gene in African-American population. Drug Metab Dispos. 2007;35(4):515–20.

58. Ramamoorthy A, Skaar TC. Gene copy number variations: it is important to determine which allele is affected. Pharmacogenomics. 2011;12(3):299–301.

59. Ramamoorthy A, Flockhart DA, Hosono N, Kubo M, Nakamura Y, Skaar TC. Differential quantification of *CYP2D6* gene copy number by four different

60. quantitative real-time PCR assays. Pharmacogenet Genomics. 2010;20(7): 451–4.

60. Jarvis JP, Peter AP, Shaman JA. Consequences of *CYP2D6* copy-number variation for pharmacogenomics in psychiatry. Front Psychiatry. 2019;10:432.

61. Hendrickson SL. A genome wide study of genetic adaptation to high altitude in feral Andean horses of the páramo. BMC Evol Biol. 2013;13:273.

62. Yang W, Qi Y, Bi K, Fu J. Toward understanding the genetic basis of adaptation to high-elevation life in poikilothermic species: a comparative transcriptomic analysis of two ranid frogs, *Rana chensinensis* and *R kukunoris*. BMC Genomics. 2012;13:588.

63. Hou C-L, Meng F-H, Wang W, Wang S-Y, Xing Y-P, Cao J-W, et al. Genome-wide analysis of copy number variations in Chinese sheep using array comparative genomic hybridization. Small Ruminant Res. 2015;128:19–26.

64. Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, et al. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. Sci Rep. 2016;6:38932.

65. Li X, Su R, Wan W, Zhang W, Jiang H, Qiao X, et al. Identification of selection signals by large-scale whole-genome resequencing of cashmere goats. Sci Rep. 2017;7(1):15142.

66. Drögemüller C, Rüfenacht S, Wichert B, Leeb T. Mutations within the *FGF5* gene are associated with hair length in cats. Anim Genet. 2007;38(3):218–21.

67. Wang X, Yu H, Lei A, Zhou J, Zeng W, Zhu H, et al. Generation of gene-modified goats targeting *MSTN* and *FGF5* via zygote injection of CRISPR/ Cas9 system. Sci Rep. 2015;5:13878.

68. Wang X, Cai B, Zhou J, Zhu H, Niu Y, Ma B, et al. Disruption of *FGF5* in cashmere goats using CRISPR/Cas9 results in more secondary hair follicles and longer fibers. PLoS One. 2016;11(10):e0164640.

69. Li W-R, Liu C-X, Zhang X-M, Chen L, Peng X-R, He S-G, et al. CRISPR/Cas9-mediated loss of FGF5 function increases wool staple length in sheep. FEBS J. 2017;284(17):2764–73.

70. Li Y, Song S, Liu X, Zhang Y, Wang D, He X, et al. Deletion of an enhancer in FGF5 is associated with ectopic expression in goat hair follicles and the cashmere growth phenotype. bioRxiv. 2020; 2020.08.24 264754.

71. Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics. 2009;25(14):1754–60.

72. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.

73. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333–9.

74. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009;25(21):2865–71.

75. Wang X, Zheng Z, Cai Y, Chen T, Li C, Fu W, et al. CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. GigaScience. 2017;6(12):1–12.

76. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

77. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444–54.

78. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

79. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2012;14(2):178–92.

80. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.

81. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.

82. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.

83. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38(6):1358–70.

84. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. G:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47(W1):W191–8.

Publisher's Note