

# Genome-wide Analysis of CNVs in Three Populations of Tibetan Sheep using Whole-genome Resequencing

## Linyong Hu

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Qi Li

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Liangzhi Zhang

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Hongjin Liu

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Tianwei Xu

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Na Zhao

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Xueping Han

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Shixiao Xu

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Xinquan Zhao

Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

## Cunfang Zhang (✉ [zcf\\_1023@126.com](mailto:zcf_1023@126.com))

State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China

---

## Research Article

**Keywords:** Copy number variations, Tibetan sheep, Whole genome resequencing

**Posted Date:** December 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1145952/v1>



# Abstract

**Background:** Copy number variation (CNV), an important source of genomic structural variation, can disturb genetic structure, dosage, regulation and expression, and is associated with phenotypic diversity and adaptation to local environments in mammals. In the present study, 24 resequencing datasets were used to characterize CNVs in three ecotypic populations of Tibetan sheep and assess CNVs related to domestication and adaptations in the Qinghai-Tibetan Plateau.

**Results:** A total of 87,832 CNV events accounting for 0.3% of the sheep genome were detected in three Tibetan sheep populations. After merging the overlapping CNVs, 2777 CNV regions (CNVRs) were obtained, among which 1098 CNVRs were shared by the three populations. The average length of these CNVRs was more than 3 kb, and duplication events were more frequent than deletions. Functional analysis showed that the shared CNVRs were significantly enriched in 56 GO terms and 18 KEGG pathways that were mainly concerned with ABC transporters, olfactory transduction and oxygen transport. Moreover, 188 CNVRs overlapped with 97 quantitative trait loci (QTLs), such as growth and carcass QTLs, immunoglobulin QTLs, milk yield QTLs and fecal counts QTLs. *PCDH15*, *APP* and *GRID2* overlapped with body weight QTLs. Furthermore, Vst analysis at each CNVR showed that *RUNX1*, LOC101104348, LOC105604082 and *PAG11* were highly divergent between HTS and VTS, and *RUNX1* and LOC101111988 were significantly differentiated between VTS and OTS. Meaningfully, the duplication of *RUNX1* may facilitate the hypoxia adaptation of OTS and HTS in QTP, which deserves further research in detail.

**Conclusions:** In this study, we represented the genome-wide distribution characteristics of CNVs in Tibetan sheep and provided a valuable genetic variation resource, which will facilitate the elucidation of the genetic basis underlying the distinct phenotypic traits and local adaptation of Tibetan sheep.

## Background

Copy number variations (CNVs) and single nucleotide polymorphisms (SNPs), as significant genetic variations, play important roles in physiological and pathological processes. Unlike SNPs, which refer to the substitution, deletion or insertion of just a single nucleotide for another, CNVs involve fragmental variation even larger than 1000 bp in size. Therefore, it has been confirmed that CNVs have the potential to markedly affect organismal phenotype.

In the past twenty years, a large number of CNVs have been revealed and identified by using array comparative genome hybridization (aCGH) chips and high-density SNP chips in domestic animals, such as cattle, sheep, goats, horses, dogs, chickens, turkeys and ducks. Meanwhile, numerous CNV-covered genes have been shown to be associated with coat color, growth, fertility and production [1, 2], immune response [3–7], olfactory transduction [8–11], molecular function [12, 13], lipid metabolism [4, 9, 14–16] and environmental adaptability [17]. Compared with aCGH and SNP chips, the sensitivity of which is mainly limited by the density of the probes, whole genome resequencing can be used to detect new and rare CNVs. Therefore, an increasing number of CNVs have been detected by using high-throughput sequencing. In terms of sheep, in addition to whole genome detection of new CNVs, the primary focus was on uncovering the CNVs associated with economic traits. Yuan *et al.* found that 1855 CNVRs were associated with 166 quantitative trait loci (QTLs), including milk QTLs, carcass QTLs, and health-related QTLs in fine-wool sheep [18]. CNVR overlapping genes, such as *SHE*, *ORMDL1*, *TOP2B*, *PIGY*, and *BAG4*, were reported to be correlated with body size in sheep [19–23]. The CNVs of *BTG3*, *PTGS1* and *PSPH* were involved in fetal muscle development, prostaglandin (PG) synthesis, and bone color [24]. Meanwhile, the correlation between CNVR-harboring genes and growth traits or phenotypic traits was also attractive. The agouti signaling protein (*ASIP*)

gene duplication has been linked to typical white coat color [25]. Distal-less homeobox 3 (*DLX3*) CNV is related to wool curling in Tan sheep [26].

Tibetan sheep (TS) is one of the three ancient coarse-wool sheep breeds in China and is predominantly distributed in the Qinghai-Tibetan Plateau (QTP). Natural adaptation and artificial selection shaped the TS with the characteristics of resistance to cold, rough feeding, and hypoxic adaptation. As one of the two master domestic animals, TS can be divided into three ecotype populations, namely, Highland-type Tibetan sheep (HTS), Valley-type Tibetan sheep (VTS) and Oura-type Tibetan sheep (OTS). HTS is famous for its carpet wool production, with an average staple length of up to 25 cm. The proportion of dry head wool in the OTS was relatively high but had remarkable meat performance. The appearance of VTS is similar to that of HTS but with a smaller body size and shorter staple length. Meanwhile, the VTS was mainly distributed in valley regions with relatively low altitudes (~2000 m), in contrast to the HTS and OTS (both distributed above 3200 m).

Nevertheless, there have been few reports on the analysis of CNVs in TS. Therefore, in the present study, we exploited a large number of CNVs in 24 individuals from the three TS populations using whole genome resequencing and identified CNVs associated with the molecular basis of phenotype differences, domestication and adaptation to QTPs.

## Results

### The landscape of copy number variation in Tibetan sheep

The high-quality NGS data of 24 Tibetan sheep were produced using an Illumina HiSeq PE150 platform (Additional file 1: Table S1). The total number of raw reads obtained for a single sheep varied between 20,759,076,000 (a highland-type sheep) and 24,338,034,900 (a valley-type sheep), and the high-quality data reached 516.242 Gb, with an average of 21,510,086,113 bp per individual. The average depth was  $7.10 \times \pm 0.30 \times$ , and the coverage rate was  $83.72\% \pm 1.63\%$  at least  $4 \times$  (Additional file 1: Table S1).

The CNVs were detected using CNVcaller as described by Wang *et al*/with default parameters [27]. The number of duplication and deletion events was 21,830 and 7479 for HTS, 21,588 and 7079 for OTS, and 21,816 and 8040 for VTS, respectively. The average lengths were 1.96, 1.95 and 1.99 kb (Table 1, Additional file 2: Table S2). The frequency of duplication events was approximately 3 times higher in all three populations than that of deletions. Moreover, as shown in Figure 1A, more than 86.9% of CNVs were distributed within 0-2 kb, 11.3% were within 2-4 kb, and less than 2% were greater than 4 kb in size.

A total of 2777 CNV regions (CNVRs) were obtained by merging overlapping CNVs in the three populations, and the numbers of duplication types in HTS, OTS and VTS were 1575, 1457 and 1484, respectively. Meanwhile, the numbers of deletion types were 521, 548 and 519. The duplication to deletion ratio of CNVRs is consistent with that of the CNVs. The average size of these CNVRs in the three populations was more than 3 kb, accounting for 0.3% of the sheep genome (Oar\_v4.0, Table 1). The size distribution of all the CNVRs showed an L-shaped pattern, with 58.2% of the CNVRs located within 1-2 kb, 28.7% within 2-4 kb, 6.5% within 6-8 kb, and others greater than 8 kb in size (Figure 1B). Moreover, as shown by the Venn diagram, 1002 CNVRs were shared by all three populations: approximately 203 CNVRs were distributed in HTS, 201 CNVRs were distributed in VTS, and 183 CNVRs were distributed in OTS (Figure 2, Additional file 3: Table S3). Furthermore, the distribution of these regions on the *Ovis aries* chromosome was analyzed by the karyoploteR package of Bioconductor. The results indicated that these CNVRs were nonuniformly distributed across each chromosome, and the number of CNVRs had a significant

positive linear relationship with the corresponding chromosome size ( $R^2 = 0.66$ , Figure 3). More CNVRs were distributed on OARX (1180), OAR1 (162), OAR 3 and OAR 10 (138), and the fewest CNVRs were distributed on OAR 24 (11), as shown in Additional file 4: Figure S1 and Additional file 5: Table S4. Functional classification showed that 1839 CNVRs were located in intergenic regions, 812 were contained within introns, and the other 85 were located in exonic regions.

## Functional annotation of CNVRs

Functional enrichment analysis was performed for all 2777 detected CNVRs. Among the CNVRs shared by the three populations of Tibetan sheep, 56 GO terms ( $p < 0.01$ ) were enriched, including 23 biological processes, 8 cellular components and 25 molecular functions (Additional file 6: Table S5). These GO terms involved ion transport (GO:0006811, GO:0006812), sensory perception system (GO:0007600, GO:0050906, and GO:0050907), gas transport (GO:0015669), and pigmentation (GO:0032400, GO:0051875). KEGG pathway analysis showed that the shared CNVR-harboring genes were enriched in 18 pathways ( $p < 0.05$ , Additional file 7: Table S6), including disease defense (ko04612, ko05332, ko05330, ko05320 and ko04672), nutrition metabolism (ko02010, ko00020, and ko00910), hematopoietic cell lineage (ko04640), and ABC transporters (ko02010), among others. In particular, the CNVR-harboring genes specifically distributed in HTS mainly participated in pigmentation (GO:0043473, GO:0033059, GO0048753, GO:0051875, and GO:0032400) processes.

## QTLs overlapping with identified CNVRs

To further reveal the CNVRs associated with sheep economic traits and confirm their hereditary effects, the detected CNVRs were compared with QTLs in the sheep QTL database. We found 188 CNVRs overlapping with 97 quantitative trait loci (QTLs), including milk production and quality (61 CNVRs), fecal egg counts (56 CNVRs), tail fat deposition (21 CNVRs), immunoglobulin A level (20 CNVRs), bone development (14 CNVRs), growth and carcass traits (36 CNVRs) (Additional file 8: Table S7). Some CNVR-overlapping genes related to slaughter performance were uncovered, such as *PCDH15* (protocadherin related 15) gene located in body weight (slaughter) QTL (95871), carcass bone percentage QTL (95870), hot carcass weight QTL (95872) and muscle weight in carcass QTL (95853); *APP* (Amyloid precursor protein) gene located in muscle weight in carcass QTL (95851) and total muscle area QTL (95852). Meanwhile, the *ASIP*, LOC101111988 and LOC105606907 genes located in the tail fat deposition QTL (127012) were also detected.

## CNVRs showing highly divergent between populations

The  $V_{st}$  statistic was used to analyze the population differentiation of CNVRs among HTS, VTS and OTS. This method is similar to  $F_{ST}$  in estimating the population-specific selective pressure at the gene level but uses the protein-coding genes annotated by CNVR. The average values of  $V_{st}$  across all of the detected responding CNVRs were 0.1061 for 'HTS vs OTS', 0.099 for 'HTS vs VTS', and 0.096 for 'VTS vs OTS' (Additional file 9: Table S8). As shown in Figure 4 and Table S8, the divergent CNVRs were distributed unevenly on the chromosome. Eighteen CNVR overlapping genes, including RUNX family transcription factor 1 (*RUNX1*), glutamate receptor 4 (*GRIA4*), LOC101104348, LOC105604082, pregnancy-associated glycoproteins 11 (*PAG11*) and LOC106990378, were the top 1% of genes that showed significant divergence between HTS and VTS. There were 15 overlapping genes, including *RAPGEF1*, *MED27*, LOC105615522, *RUNX1*, LOC101113153, LOC105607734 and LOC101111988 (located upstream of *ASIP*), among others, which were the top 1% genes highly differentiated between the VTS and OTS (Table 2). Among them, only *RUNX1*, LOC101104348, LOC105604082, *PAG11* and LOC101111988 were located in exonic or intronic regions. Unfortunately, few significantly divergent genes located in exonic or intronic regions were detected between HTS and OTS. Notably, *RUNX1* was detected in both the 'HTS vs VTS' pairs and 'VTS vs OTS' pairs. It is

worth noting that *ASIP*, playing a vital role in coat color, also showed significant variation between 'VTS vs OTS' pairs and 'HTS vs OTS' pairs.

## CNV Validation by qPCR

qPCR was used to verify the accuracy of CNVRs screened by resequencing. Therefore, 11 CNVRs were selected according to their VST value ( $P < 0.01$ ) (Additional file 11: Table S9). The 11 selected CNVRs included five duplication types of CNVR-harboring genes *LOC101111526*, *PLCB1*, *GUCY1A2*, *GRIA4*, and *TMEM144* and six deletion types of *LOC105602432*, *RUNX1*, *PAG11*, *LOC101113153*, and *ASIP*. The qPCR results suggested that the copy number variants of these target genes were all consistent with the results based on resequencing (Additional file 10: Figure S2).

## Discussion

To date, there are many reports regarding CNVs of sheep using aCGH, SNP array and genome resequencing. Due to the difference in CNV detection platforms, the quantity of detected CNVs varies widely in different studies. In terms of the CNVs captured by aCGH and SNP chips, the number usually ranges from dozens to hundreds, and the average length is approximately 100-300 kb [28–32]. Using whole genome resequencing, Cheng *et al* [33] reported that 4301 CNVRs were identified in three Chinese breeds. A total of 7228 CNVRs were obtained, including 861 duplications, 6345 deletions, and 22 both events in Chinese indigenous fine-wool sheep [18]. Similarly, there were differences in the amounts of CNVR, and the proportion accounted for chromosomes. In the present study, we obtained 2777 CNVRs, including 1965 duplications and 812 deletions. These CNVRs accounted for ~0.3% of the sheep reference genome. This coverage ratio is comparable with previous reports [29, 33] but lower than the 10% reported by Salehian-Dehkordi *et al*. [34]. There is no rule about which (duplication or deletion) happened more frequently in previous studies. In addition to the detection platform, the reason for these variations may be the differences in calling algorithms and standards. Meanwhile, the breed and sample size used in the study were other considerable factors [35].

The GO enrichment analysis showed that the detected CNVRs harboring genes shared by the three populations were significantly enriched in ion transport, sensory perception system, gas transport and pigmentation. Among them, sensory perception was also significantly enriched in yaks [36, 37] and other sheep breeds [18, 29]. The three TS populations live in alpine grasslands where the weather conditions are much harsher. As yak, TS faced the same living conditions as the lack of herbage in the cold season. The well-developed sensory perception system was important to improve the yak and sheep's ability to acquire food and avoid noxious weeds. The gas transport term was also enriched in TS, and efficient oxygen exchange may facilitate hypoxia adaptation. In the KEGG pathway analysis, it is noteworthy that the CNVR-harboring genes shared by the three populations were significantly enriched in nutrition metabolism, ABC transporters, disease defense and hematopoietic cell lineage. The enrichment of nutrient metabolism terms, including nitrogen metabolism, citrate cycle, and bile secretion, was important for the digestion of nutrients in TS, especially in the cold season. Similar results showing the enrichment of CNVR-harboring genes in ABC transporters were also reported in humans [38, 39], cattle [10, 40, 41] and goats [42]. In mammals, ABC transporters can carry a broad array of endogenous metabolites, such as amino acids, peptides, and sugars, across lipid membranes, which facilitate the absorption and utilization of these nutrients. Hypoxia can promote hematopoiesis. Here, CNVR-harboring genes were enriched in the hematopoietic cell lineage in Tibetan sheep, which may make it suitable for their adaptation to high altitudes with low oxygen. Overall, the number of enriched CNVR overlapping genes associated with forage consumption and oxygen transport may be helpful for their adaptation to the local environment.

Notably, the olfactory transduction pathway was significantly enriched specifically in OTSs. Enrichment of the olfactory transduction pathway has been reported in cattle [11, 16, 43, 44], yaks [45], sheep [18, 28, 46] and goats [42]. It has been revealed to influence food consumption [47] and as a factor to assess feed efficiency and performance in crossbred beef cattle [48] and residual feed intake in pigs [49]. Meanwhile, more CNVR harboring genes were enriched in amino acid and VFA metabolism pathways in the OTS. OTS has a much better meat performance than HTS and VTS. More CNVR harboring genes enriched in the olfactory transduction pathway, together with protein and energy metabolism pathways, may explain the better production performance in OTS.

QTLs, which contain genetic variants affecting the economic traits of domestic animals, can be used to select candidate CNVR-harboring genes that affect phenotypes in sheep. After integrating CNVs into QTLs, we found 188 CNVRs overlapping with 97 sheep QTL regions in this study. Many CNVs harboring genes, such as *PCDH15*, *APP* and *GRID2*, are located in growth and carcass QTL regions. The *PCDH15* gene was identified to be associated with the concentration of the neurotransmitter glutamate (Glu) in cattle [50]. *APP* is one of the well-known pathogeneses in Alzheimer's disease. Zheng *et al.* indicated that homozygous *APP* deficiency leads to a 15–20% reduction in body weight [51]. An *et al.* reported that adipocyte-specific and mitochondria-targeted *APP*-overexpressing mice display increased body mass [52]. *GRID2* was also identified as being associated with body weight in rats [53].

Selective sweeping can reveal putative regions that undergo environmental and artificial selection during local adaptation and domestication. To screen the critical CNVR significantly divergent between different populations, the pairwise *Vst* value was estimated [18, 50, 54]. Here, the five CNVR harboring genes *RUNX1*, *LOC101104348*, *LOC105604082*, *PAG11* and *LOC101111988* showed significant pairwise differentiation among HTS, OTS and VTS. It is well known that *RUNX1* plays a crucial role in hematopoiesis, leukemogenesis and neural development. Lin *et al.* reported that GWAS hit SNPs associated with colostrum albumin concentration were enriched in *RUNX1* in Chinese Holsteins, and these mutations might initiate the hyperactivation of inflammatory and innate immunity [55]. The GWAS hit SNP within *RUNX1* is associated with the mean corpuscular volume level in pigs [56]. Moreover, HIF-1 $\alpha$  facilitates *RUNX1* transcriptional activity under hypoxic conditions and triggers hematopoietic stem cell differentiation, which ultimately improves oxygen transport to peripheral tissues [57]. Chaka sheep were a cultivated breed mainly with Tibetan sheep as basic ewes and adapted to a low oxygen environment for a long time. Hence, we speculated that CNVR-harboring *RUNX1* underwent strict selection pressure in Tibetan and Chaka sheep, possibly helpful for their adaptation to a low oxygen environment at high altitudes. *PAG11* is a pregnancy-associated glycoprotein that is expressed in the trophoblast of the ruminant placenta and influences embryo growth and survival [58]. HTS should face a shortage of forage and cold environments in their late pregnancy stage, in contrast to VTS. We thought the divergence of CNVR-harboring *PAG11* between HTS and VTS may be related to their different habitat conditions, especially in the pregnancy stage. Notably, the *ASIP* gene, duplication of CNVR-harboring *LOC101111988*, located upstream of *ASIP*, was divergent between VTS vs OTS (*Vst* = 0.489) and HTS vs OTS (*Vst* = 0.284). The *ASIP* gene has been widely studied in mammals for its effect on animal coat color. Individuals with normal or duplication alleles of the *ASIP* gene are generally white or gray coat color, but individuals with normal or single deletion alleles in *ASIP* almost entirely have solid-black coat color in sheep [59–61]. In our study, the duplication of the *LOC101111988* gene in VTS and HTS might account for their mainly white coat color. Meanwhile, the deletion in *LOC101111988* CNV might be the basis for the primary brown covering color, especially in the neck in OTS.

## Conclusions

In this study, the whole genomic characteristics were described in three ecotypic populations of Tibetan sheep. A total of 87,802 CNV events and 2777 CNVRs covering 0.3% of the sheep genome were captured. A few CNVR-harbored genes, such as *PCDH15*, *APP*, *RUNX1*, *PAG11*, and *ASIP*, were uncovered and may be associated with body weight, environmental adaptation, fertility, and coat color. Above all, our results provide a valuable genome-wide variation resource in Tibetan sheep for the elucidation of the genetic basis underlying the distinct phenotypic traits and local adaptation of Tibetan sheep.

## Materials And Methods

### Animals and resequencing

Twenty-four blood samples of Tibetan sheep were collected from their core distribution region in this study, including 8 highland-type sheep (HTS, Qilian County), 8 valley-type sheep (VTS, Huangyuan County) and 8 outer-type sheep (OTS, Henan County). The samples were stored in EDTA antifreezing tubes at -20°C. Genomic DNA was extracted with a Qiagen DNA purification kit (Qiagen, Dusseldorf, DE), and then the DNA integrity and concentration were checked with 1% agarose gel electrophoresis and Qubit<sup>®</sup> 3.0 Fluorometer (Invitrogen, USA), respectively. A sequencing library was built with 0.2 µg genomic DNA from each sample using the NEB Next<sup>®</sup> Ultra<sup>™</sup> DNA Library Prep Kit (NEB, USA) following the manufacturer's recommendations. Briefly, genomic DNA samples were fragmented by sonication to a size of 350 bp, and then DNA fragments were endpolished, A-tailed, and ligated with the full-length adapter, followed by further PCR amplification. After PCR products were purified by the AMPure XP system (Beckman Coulter, Beverly, USA), DNA concentration was measured by a Qubit<sup>®</sup> Fluorometer (Invitrogen, USA). The DNA libraries were sequenced using an Illumina HiSeq X-Ten platform (Illumina, USA), and 150 bp paired-end reads were generated and stored in FASTQ format. Paired reads with more than 10% unidentified nucleotides in either read, with low-quality bases (Phred quality value <5) over 50%, and with more than 10 bp aligned to the adapter were removed by using Fastp (0.19.7) to obtain clean data.

### CNV and CNVR detection and annotation

The following steps were required before CNV detection: (i) the clean reads of each sample were aligned against the reference genome of *Ovis aries* (Oar\_v4.0, <http://www.ncbi.nlm.nih.gov/genome/?term=sheep>) using BWA (Burrows–Wheeler Aligner) [62] and (ii) alignment files were converted to BAM files using SAMtools software [63]. (iii) SAMtools was also used to remove potential PCR duplications. If multiple read pairs have identical external coordinates, only the pair with the highest mapping quality is retained. CNVs were detected using CNVcaller with default parameters (-w: 800 bp, -l: 0.2, -u: 0.7, -g: 0.5) [64]. The individual candidate CNV windows are defined using 2 criteria: (i) The normalized read depth (RD) must be significantly higher or lower than the normalized mean RD (deletions < 1 - 2 \* STDEV; duplications > 1 + 2 \* STDEV). (ii) Considering that the normalized RD of heterozygous deletions and duplications should be approximately 0.5 and 1.5, respectively, an empirical standard for the normalized RD (deletions < 0.65; duplications > 1.35) must also be achieved. Low-frequency windows with low sequencing coverage were removed, and windows with an allele frequency ≥0.05 or at least 3 homozygous duplicated/deleted individuals were selected for further validation. If Pearson's correlation index was significant at the  $P = 0.05$  level by Student's  $t$  test, these 2 adjacent non-overlapping windows were merged into 1 call. To avoid putative positive CNVRs, the population-level candidate CNVRs for the three populations and TS as a whole were analyzed. CNV region definition, the distance between the 2 initial calls is less than 20% of their combined length,

and the Pearson's correlation index of the 2 CNVRs is significant at the  $P = 0.01$  level. Functional annotation of CNVs was completed by ANNOVAR[65] and classified as intronic, exonic or intergenic.

## Functional enrichment analysis of CNVR-harboring genes

CNVR-harboring genes were retrieved from the NCBI database by BioMart software (<http://www.biomart.org/>), and completely or partially ( $\geq 50\%$ ) overlapping genes were all reserved for later analysis. GO and KEGG functional enrichment analyses were performed according to Huang *et al.* [66]. The P value was calculated and subjected to FDR correction. The merged CNVRs were compared with QTLs in the animal QTL database (<https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>), to further assess the CNVRs that were correlated with economic traits in three ecotypes of Tibetan sheep.

## Sweep selective analysis of the CNVR

Three ecotypes of Tibetan sheep have formed their own special characters under positive selection throughout long-term domestication and adaptation in the plateau. Therefore, the  $V_{st}$  was calculated similar to population differentiation index  $F_{ST}$  using the equation:  $V_{st} = (V_{total} - [V_{pop1} \times N_{pop1} + V_{pop2} \times N_{pop2}] / N_{total}) / V_{total}$ , where  $V_{total}$  is the total variance,  $V_{pop}$  is the CN variance for each respective population,  $N_{pop}$  is the sample size for each respective population, and  $N_{total}$  is the total sample size. Subsequently, the CNVRs with the top five  $V_{st}$  values were selected as candidate CNVRs, and functional enrichment analysis of these regions was performed.

## qPCR and high depth resequencing validation

Quantitative real-time PCR was employed to validate the accuracy of CNVs detected in our study. Ten CNVs were selected that were identified by CNVcaller. The primers were designed using Primer 5 software based on NCBI reference sequences, and the DGAT1 gene was chosen as a reference gene (all primers are shown in Additional file 11: Table S9). qPCR was performed using the TB Green PCR reagent kit (Takara Bio). Three replicates per sample and blank controls were required in the PCR. The results were analyzed using the  $2^{-\Delta\Delta Ct}$  method, and the copy number of the targeted genes was calculated [38-40].

## Declarations

### *Ethics Statement*

All experimental protocols and procedures were carried out in accordance with institutional guidelines and were approved by the Animal Care and Ethics Committee of Northwest Institute of Plateau Biology, Chinese Academy of Sciences (NWIPB20160302). We obtained the permission from the owner of the sheep to take samples.

### *Consent for publication*

Not applicable.

## ***Availability of data and materials***

The raw sequencing data of 24 sheep for the identification of CNVs in this study have been deposited in the Genome Sequence Archive in BIG Data Center (<https://ngdc.cnca.ac.cn/>), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under the accession number CRA005573.

## ***Competing interests***

The authors declare that they have no competing interests.

## ***Funding***

This study was supported in part by The Second Tibetan Plateau Scientific Expedition and Research (STEP) program (Grant No. 2019QZKK0302), Key R & D and Transformation Project of Qinghai Province (2019-NK-173), and Basic Applied Study of Qinghai Province (2014-ZJ-710).

## ***Authors' contributions***

LH, SX, XZ and CZ initiated the study and designed the project. LH, HL, TX, NZ and XH collected the samples. LH, QL, and CZ performed the bioinformatics and experimental analysis. LH, QL and CZ summarized the results. LH and CZ wrote the manuscript. All authors read and approved the final manuscript.

## ***Acknowledgements***

Not applicable.

## ***Authors details***

<sup>1</sup> Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810008, China

<sup>2</sup> State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China

## **References**

1. Zhou Y, Utsunomiya YT, Xu L, Hay EHA, Bickhart DM, Alexandre PA, Rosen BD, Schroeder SG, Carneiro R, de Rezende Neves HH *et al*: **Genome-wide CNV analysis reveals variants associated with growth traits in *Bos indicus***. *Bmc Genomics*2016, **17**.
2. de Lemos MVA, Peripolli E, Berton MP, Braga Feitosa FL, Olivieri BF, Stafuzza NB, Tonussi RL, Kluska S, Justino Chiaia HL, Mueller *et al*: **Association study between copy number variation and beef fatty acid profile of Nellore cattle**. *Journal of Applied Genetics*2018, **59**(2):203-223.
3. Zhang X, Wang K, Wang L, Yang Y, Ni Z, Xie X, Shao X, Han J, Wan D, Qiu Q: **Genome-wide patterns of copy number variation in the Chinese yak genome**. *BMC Genomics*2016, **17**:379.

4. Xu Y, Jiang Y, Shi T, Cai H, Lan X, Zhao X, Plath M, Chen H: **Whole-genome sequencing reveals mutational landscape underlying phenotypic differences between two widespread Chinese cattle breeds.** *PLoS One*2017, **12**(8):e0183921.
5. Sasaki S, Watanabe T, Nishimura S, Sugimoto Y: **Genome-wide identification of copy number variation using high-density single-nucleotide polymorphism array in Japanese Black cattle.** *BMC Genet*2016, **17**:26.
6. Chen L, Chamberlain AJ, Reich CM, Daetwyler HD, Hayes BJ: **Detection and validation of structural variations in bovine whole-genome sequence data.** *Genet Sel Evol*2017, **49**(1):13.
7. Chen J, Lee MO, Womack JE: **Genetic variation and gene conversions within the bovine NK-lysin gene family.** *Anim Genet*2017, **48**(2):225-227.
8. Zhou Y, Utsunomiya YT, Xu L, Hay el HA, Bickhart DM, Sonstegard TS, Van Tassell CP, Garcia JF, Liu GE: **Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations.** *Sci Rep*2016, **6**:29219.
9. de Lemos MVA, Peripolli E, Berton MP, Feitosa FLB, Olivieri BF, Stafuzza NB, Tonussi RL, Kluska S, Chiaia HLJ, Mueller *et al.*: **Association study between copy number variation and beef fatty acid profile of Nellore cattle.** *J Appl Genet*2018, **59**(2):203-223.
10. Upadhyay M, da Silva VH, Megens HJ, Visker M, Ajmone-Marsan P, Balteanu VA, Dunner S, Garcia JF, Ginja C, Kantanen *et al.*: **Distribution and Functionality of Copy Number Variation across European Cattle Populations.** *Front Genet*2017, **8**:108.
11. Mielczarek M, Fraszczak M, Nicolazzi E, Williams JL, Szyda J: **Landscape of copy number variations in *Bos taurus*: individual - and inter-breed variability.** *Bmc Genomics*2018, **19**(1):410.
12. Pierce MD, Dzama K, Muchadeyi FC: **Genetic Diversity of Seven Cattle Breeds Inferred Using Copy Number Variations.** *Front Genet*2018, **9**:163.
13. Letaief R, Rebours E, Grohs C, Meersseman C, Fritz S, Trouilh L, Esquerre D, Barbieri J, Klopp C, Philippe *et al.*: **Identification of copy number variation in French dairy and beef breeds using next-generation sequencing.** *Genet Sel Evol*2017, **49**(1):77.
14. Gao Y, Jiang J, Yang S, Hou Y, Liu GE, Zhang S, Zhang Q, Sun D: **CNV discovery for milk composition traits in dairy cattle using whole genome resequencing.** *Bmc Genomics*2017, **18**.
15. Goyache F, Fernandez I, Tapsoba ASR, Traore A, Menendez-Arias NA, Alvarez I: **Functional characterization of Copy Number Variations regions in Djallonke sheep.** *Journal of Animal Breeding and Genetics*2021, **138**(5):600-612.
16. Gao Y, Jiang J, Yang S, Hou Y, Liu GE, Zhang S, Zhang Q, Sun D: **CNV discovery for milk composition traits in dairy cattle using whole genome resequencing.** *BMC Genomics*2017, **18**(1):265.
17. Wang MD, Dzama K, Rees DJG, Muchadeyi FC: **Tropically adapted cattle of Africa: perspectives on potential role of copy number variations.** *Anim Genet*2016, **47**(2):154-164.
18. Yuan C, Lu Z, Guo T, Yue Y, Wang X, Wang T, Zhang Y, Hou F, Niu C, Sun *X et al.*: **A global analysis of CNVs in Chinese indigenous fine-wool sheep populations using whole-genome resequencing.** *BMC Genomics*2021, **22**(1):78.
19. Jiang R, Cheng J, Cao X-K, Ma Y-L, Chaogetu B, Huang Y-Z, Lan X-Y, Lei C-Z, Hu L-Y, Chen H: **Copy Number Variation of the SHE Gene in Sheep and Its Association with Economic Traits.** *Animals*2019, **9**(8).
20. Wang X, Cao X, Wen Y, Ma Y, Elnour IE, Huang Y, Lan X, Chaogetu B, Hu L, Chen H: **Associations of ORMDL1 gene copy number variations with growth traits in four Chinese sheep breeds.** *Arch Anim Breed*2019, **62**(2):571-

21. Toremurat Z, Ibrahim EE, Huang Y-Z, Lan X, Pi L, Chaogetu B, Hu L, Chen H: **Copy number variations of TOP2B gene are associated with growth traits in Chinese sheep breeds.** *Animal Biotechnology*2020.
22. Feng Z, Li X, Cheng J, Jiang R, Huang R, Wang D, Huang Y, Pi L, Hu L, Chen H: **Copy Number Variation of the PIGY Gene in Sheep and Its Association Analysis with Growth Traits.** *Animals*2020, **10**(4).
23. Yang Z, Cao X, Ma Y, Cheng J, Song C, Jiang R, Wang X, Huang Y, Buren C, Lan X *et al.*: **Novel copy number variation of the BAG4 gene is associated with growth traits in three Chinese sheep populations.** *Animal Biotechnology*2021, **32**(4):461-469.
24. Yang L, Xu L, Zhou Y, Liu M, Wang L, Kijas JW, Zhang H, Li L, Liu GE: **Diversity of copy number variation in a worldwide population of sheep.** *Genomics*2018, **110**(3):143-148.
25. Bickhart DM, Liu GE: **The challenges and importance of structural variation detection in livestock.** *Front Genet*2014, **5**.
26. Ma Q, Liu X, Pan J, Ma L, Ma Y, He X, Zhao Q, Pu Y, Li Y, Jiang L: **Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600 K SNP array.** *Sci Rep*2017, **7**(1):912.
27. Wang X, Zheng Z, Cai Y, Chen T, Li C, Fu W, Jiang Y: **CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations.** *Gigascience*2017, **6**(12):1-12.
28. Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F *et al.*: **Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array.** *BMC Genomics*2013, **14**:229.
29. Ma Y, Zhang Q, Lu Z, Zhao X, Zhang Y: **Analysis of copy number variations by SNP50 BeadChip array in Chinese sheep.** *Genomics*2015, **106**(5):295-300.
30. Zhu C, Fan H, Yuan Z, Hu S, Ma X, Xuan J, Wang H, Zhang L, Wei C, Zhang Q *et al.*: **Genome-wide detection of CNVs in Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP arrays.** *Sci Rep*2016, **6**:27822.
31. Wu, Kai-Feng, Zhou, Huan-Min, Zhang, Dong, Hou, Cheng-Lin, Liu, Chun-Xia: **Genome-wide analysis of copy number variations in Chinese sheep using array comparative genomic hybridization.** *Small Ruminant Research*2015.
32. Jenkins GM, Goddard, M. E., Black, M. A., Brauning, R., Auvray, B., Dodds, K. G., Kijas, J. W., Cockett, N., & McEwan, J. C.: **Copy number variants in the sheep genome detected using multiple approaches.** *BMC Genomics*2016, **17**:441.
33. Cheng J, Cao X, Hanif Q, Pi L, Hu L, Huang Y, Lan X, Lei C, Chen H: **Integrating Genome-Wide CNVs Into QTLs and High Confidence GWAScore Regions Identified Positional Candidates for Sheep Economic Traits.** *Front Genet*2020, **11**.
34. Salehian-Dehkordi H, Xu Y-X, Xu S-S, Li X, Luo L-Y, Liu Y-J, Wang D-F, Cao Y-H, Shen M, Gao L *et al.*: **Genome-Wide Detection of Copy Number Variations and Their Association With Distinct Phenotypes in the World's Sheep.** *Front Genet*2021, **12**.
35. Yuan C, Lu Z, Guo T, Yue Y, Wang X, Wang T, Zhang Y, Hou F, Niu C, Sun X *et al.*: **A global analysis of CNVs in Chinese indigenous fine-wool sheep populations using whole-genome resequencing.** *Bmc Genomics*2021, **22**(1).
36. Qiu QZ, Guojie; Ma, Tao; Qian, Wubin; Wang, Junyi; Ye, Zhiqiang; Cao, Changchang; Hu, Qianjun; Kim, Jaebum; Larkin, Denis M.; Auvil, Loretta; Capitanu, Boris; Ma, Jian; Lewin, Harris A.; Qian, Xiaojun; Lang, Yongshan; Zhou, Ran; Wang, Lizhong: **The yak genome and adaptation to life at high altitude.** *Nature Genetics*2012, **44**(8):946.

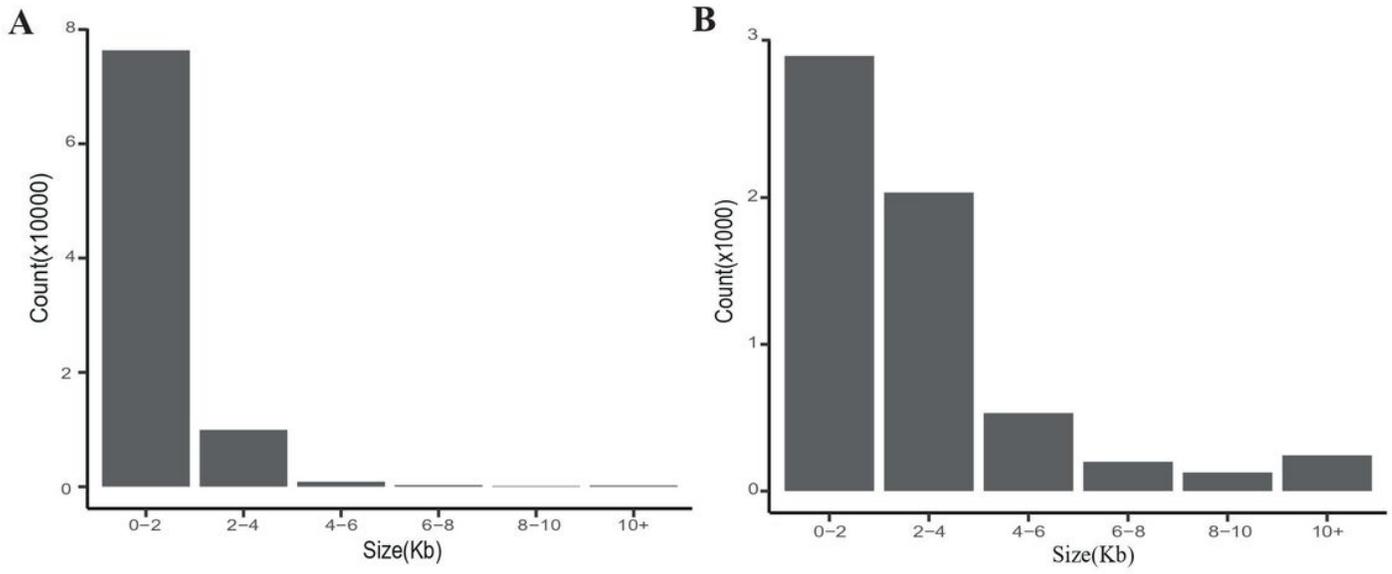
37. Hui W, Chai Z, Dan H, Ji Q, Zhong J: **A global analysis of CNVs in diverse yak populations using whole-genome resequencing.** *BMC Genomics*2019, **20**(1).
38. Sun Y, Shi N, Lu H, Zhang J, Ma Y, Qiao Y, Mao Y, Jia K, Han L, Liu F *et al*: **ABCC4 copy number variation is associated with susceptibility to esophageal squamous cell carcinoma.** *Carcinogenesis*2014, **35**(9):1941-1950.
39. Dron JS, Wang J, Berberich AJ, Iacocca MA, Cao H, Yang P, Knoll J, Tremblay K, Brisson D, Netzer C *et al*: **Large-scale deletions of the ABCA1 gene in patients with hypoalphalipoproteinemia.** *Journal of Lipid Research*2018, **59**(8):1529-1535.
40. Liu GE, Brown T, Hebert DA, Cardone MF, Hou Y, Choudhary RK, Shaffer J, Amazu C, Connor EE, Ventura M *et al*: **Initial analysis of copy number variations in cattle selected for resistance or susceptibility to intestinal nematodes.** *Mamm Genome*2011, **22**(1-2):111-121.
41. Lee KT, Chung WH, Lee SY, Choi JW, Kim J, Lim D, Lee S, Jang GW, Kim B, Choy YH *et al*: **Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity.** *BMC Genomics*2013, **14**:519.
42. Guan D, Martínez A, Castelló A, Landi V, Amills M: **A genome-wide analysis of copy number variation in Murciano-Granadina goats.** *Genetics Selection Evolution*2020, **52**(1).
43. **Identification of copy number variations in Qinchuan cattle using BovineHD Genotyping Beadchip array.** *Molecular Genetics & Genomics*2015, **290**(1):319-327.
44. Lemos MD, Peripolli E, Berton MP, Feitosa F, Olivieri BF, Stafuzza NB, Tonussi RL, Kluska S, Chiaia H, Mueller L: **Association study between copy number variation and beef fatty acid profile of Nelore cattle.** *Journal of Applied Genetics*2018.
45. Guang-Xin E, Yang BG, Zhu YB, Duang XH, An TW: **Genome-wide selective sweep analysis of the high-altitude adaptability of yaks by using the copy number variant.** *3 Biotech*2020, **10**(6).
46. Jiasen, Liu, Li, Zhang, Lingyang, Xu, Hangxing, Ren, Jian, Lu: **Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array.** *BMC Genomics*2013, **14**(1):1-11.
47. Ma, Minghong: **Encoding olfactory signals via multiple chemosensory systems.** *Critical Reviews in Biochemistry & Molecular Biology*2008, **42**(6):463-480.
48. Abo-Ismael MK, Voort GV, Squires... J: **Single nucleotide polymorphisms for feed efficiency and performance in crossbred beef cattle.** *Bmc Genetics*2014, **15**(1).
49. **Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake.** *Frontiers in Genetics*2014, **5**(5):307.
50. Chen Q, Qu K, Ma Z, Zhan J, Lei C: **Genome-Wide Association Study Identifies Genomic Loci Associated With Neurotransmitter Concentration in Cattle.** *Frontiers in Genetics*2020, **11**.
51. Hui Z, Jiang M, Trumbauer ME, Sirinathsinghji D, Ploeg L: **beta-Amyloid precursor protein-deficient mice show reactive gliosis and decreased locomotor activity.** *Cell*1995, **81**(4):525-531.
52. An YA, Crewe C, Asterholm IW, Sun K, Chen S, Zhang F, Shao M, Funcke JB, Zhang Z, Straub L: **Dysregulation of amyloid precursor protein impairs adipose tissue mitochondrial function and promotes obesity.** *Nature Metabolism*.
53. Keele GR, Prokop JW, He H, Holl K, Littrell J, Deal A, Francic S, Cui L, Gatti DM, Broman K *et al*: **Genetic Fine-Mapping and Identification of Candidate Genes and Variants for Adiposity Traits in Outbred Rats.** *Obesity*2018, **26**(1):213-222.

54. Salehian-Dehkordi H XY, Xu SS, Li X, Luo LY, Liu YJ, Wang DF, Cao YH, Shen M, Gao L, Chen ZH, Glessner JT, Lenstra JA, Esmailizadeh A, Li MH, Lv FH: **Genome-Wide Detection of Copy Number Variations and Their Association With Distinct Phenotypes in the World's Sheep.** *Front Genet*2021, **12**:670582.
55. Lin S, Wan Z, Zhang J, Xu L, Sun D: **Genome-Wide Association Studies for the Concentration of Albumin in Colostrum and Serum in Chinese Holstein.** *Animals*2020, **10**(12):2211.
56. Lee J-B, Yoo C-K, Park H-B, Cho I-C, Lim H-T: **Association of the Single Nucleotide Polymorphisms in RUNX1, DYRK1A, and KCNJ15 with Blood Related Traits in Pigs.** *Asian-Australasian Journal of Animal Sciences*2016, **29**(12):1675-1681.
57. Lee SH, Manandhar S, Lee YM: **Roles of RUNX in Hypoxia-Induced Responses and Angiogenesis.** *Oxygen Transport to Tissue XXXIII*2017, **962**:449-469.
58. Shorten PR, Ledgard AM, Donnison M, Pfeffer PL, McDonald RM, Berg DK: **A mathematical model of the interaction between bovine blastocyst developmental stage and progesterone-stimulated uterine factors on differential embryonic development observed on Day 15 of gestation.** *Journal of Dairy Science*2018, **101**(1):736-751.
59. Han JL, Yang M, Yue YJ, Guo TT, Liu JB, Niu CE, Yang BH: **Analysis of agouti signaling protein (ASIP) gene polymorphisms and association with coat color in Tibetan sheep (*Ovis aries*).** *Genetics and Molecular Research*2015, **14**(1):1200-1209.
60. Fontanesi L, Dall'Olio S, Beretti F, Portolano B, Russo V: **Coat colours in the Massese sheep breed are associated with mutations in the agouti signalling protein (ASIP) and melanocortin 1 receptor (MC1R) genes.** *Animal*2011, **5**(1):8-17.
61. 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):1282-1293.
62. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics*2009, **25**(14):1754-1760.
63. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics*2009, **25**(16):2078-2079.
64. Abyzov A, Urban AE, Snyder M, Gerstein M: **CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing.** *Genome Res*2011, **21**(6):974-984.
65. Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.** *Nucleic Acids Res*2010, **38**(16):e164.
66. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nature Protocols*2009, **4**(1):44-57.

## Tables

Due to technical limitations, tables are only available as a download in the Supplemental Files section.

## Figures



**Figure 1**

Size distribution of CNVs (A) and CNVRs (B) in Tibetan sheep

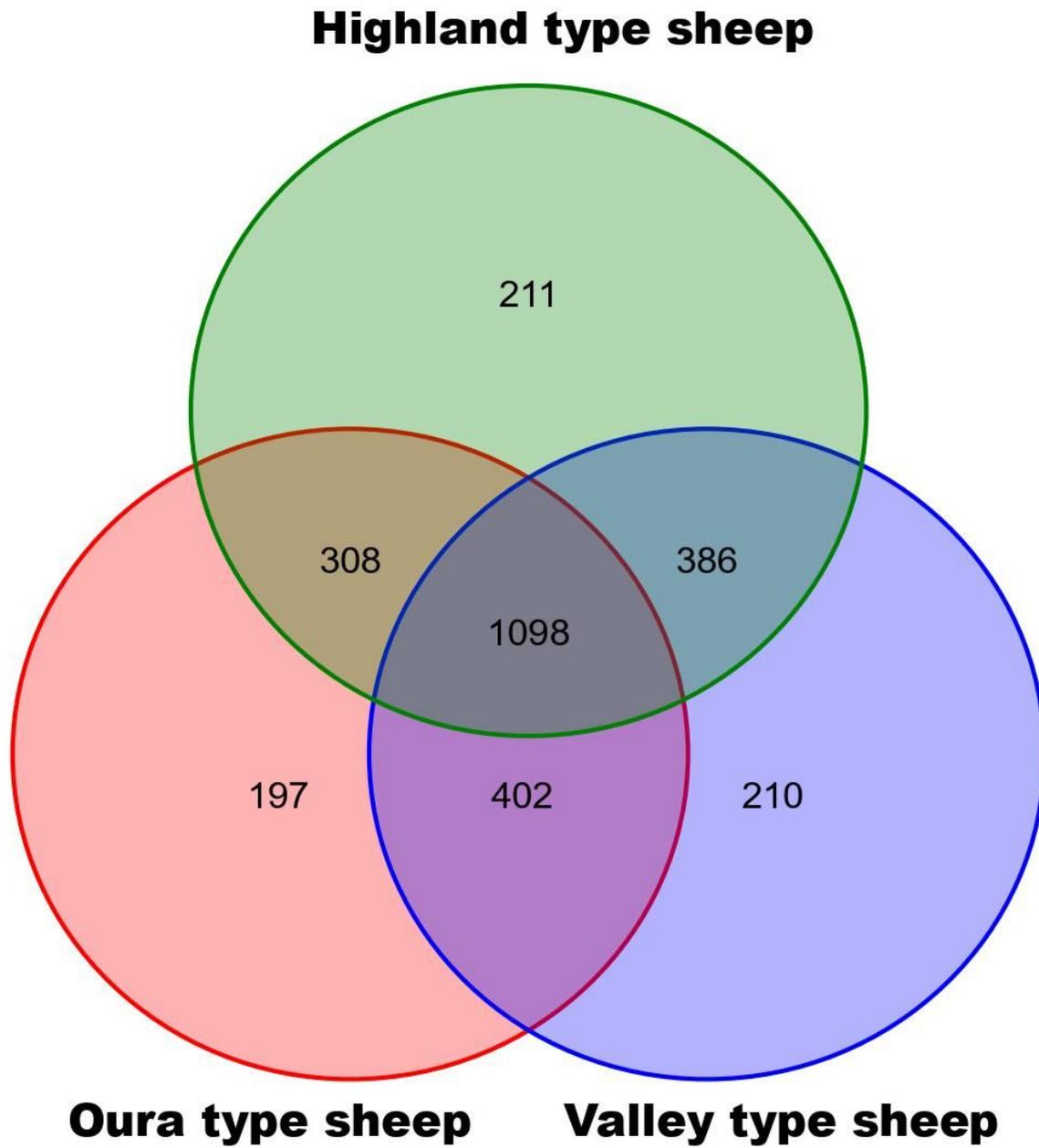
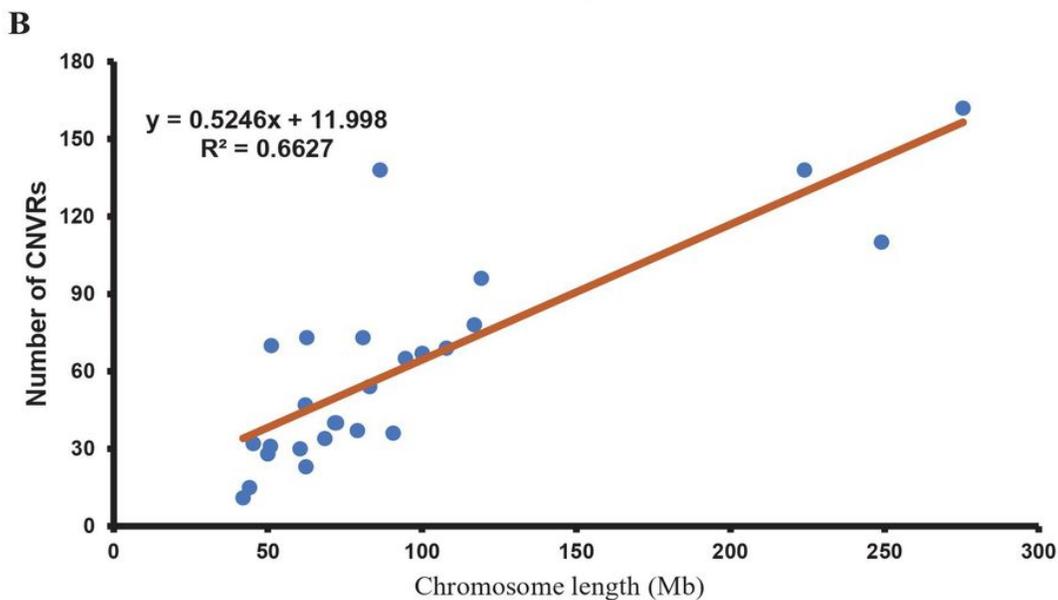
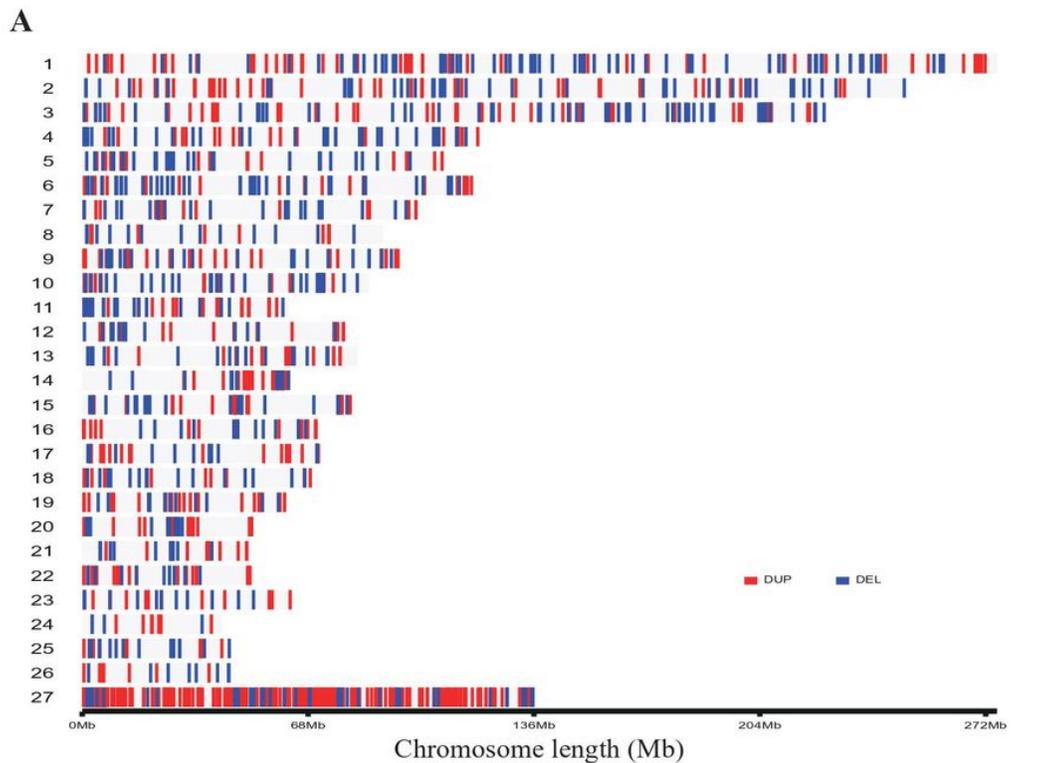


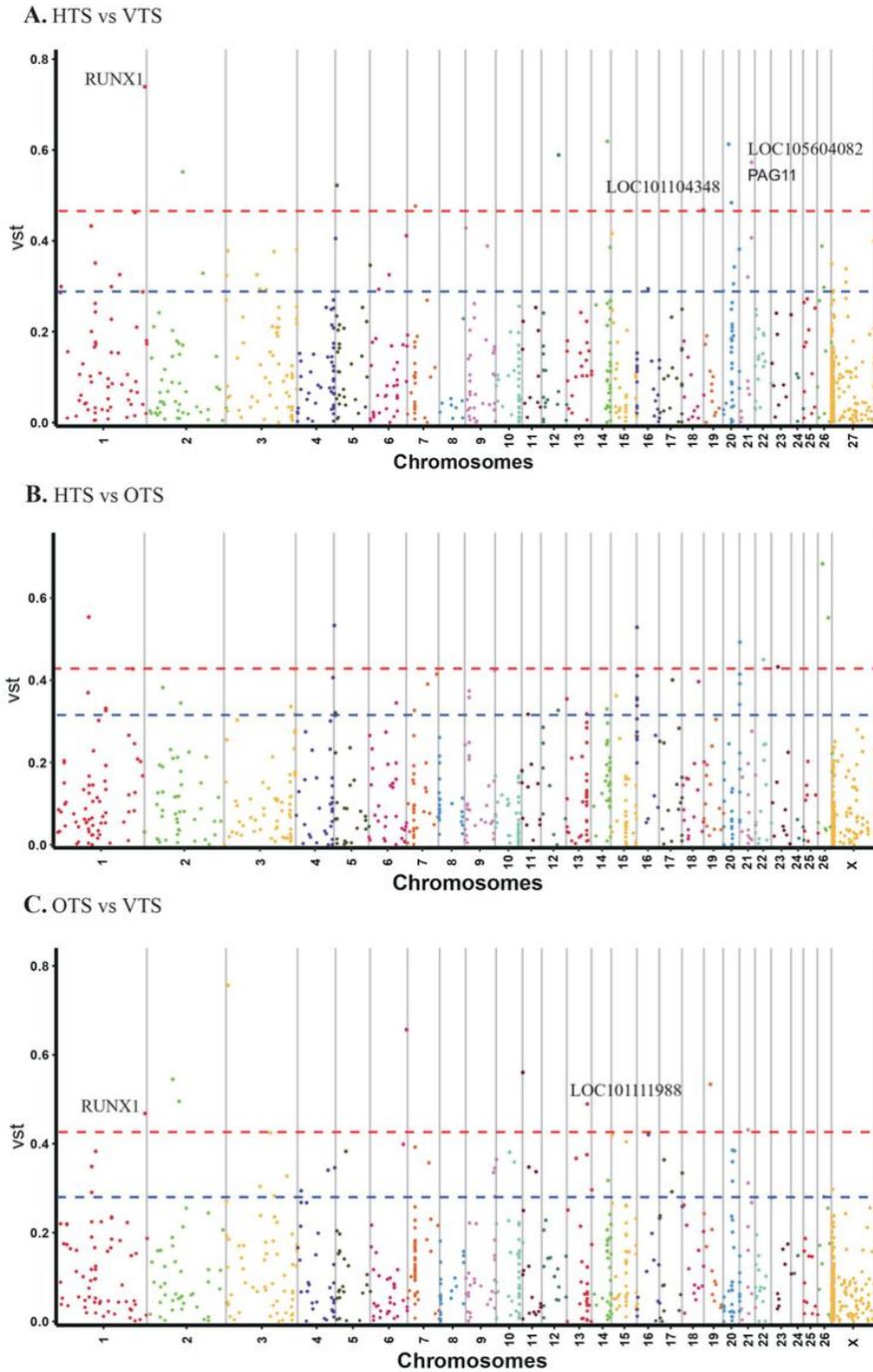
Figure 2

Venn diagram of CNVR numbers identified in three Tibetan sheep populations



**Figure 3**

Genomic landscape of CNVRs in Tibetan sheep. A. CNVRs distributed in each chromosome, and 27 represents the X chromosome. B. The number of CNVRs had a positive linear relationship with the corresponding chromosome size



**Figure 4**

Genome-wide Vst value plots for CNVRs in three population pairs. A, B, and C show the divergent CNVRs between population pairs of HTS and VTS, HTS and OTS, OTS and VTS. The plots distributed above the dotted line with red and blue colors represent the top 1% and 5% divergent CNVRs, respectively

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1TableS1.Basicinformationrelatingtothesheepandtheresequencingdatausedinthepresentstudy.xlsx](#)

- [Additionalfile2TableS2.TheidentifiedCNVsinthreeTSpopulations.xlsx](#)
- [Additionalfile3TableS3.DetailedinformationofthedetectedCNVRs.xlsx](#)
- [Additionalfile4FigureS1.ThelengthofCNVRsdistributedoneachchromosomeinthreeTSpopulations.pdf](#)
- [Additionalfile5TableS4.ThenumberandlengthofCNVRsdistributedonchromosomesinthreesheep populations.xlsx](#)
- [Additionalfile6TableS5.FunctionalenrichmentofGOtermsforCNVRharboredgenes.xlsx](#)
- [Additionalfile7TableS6.FunctionalenrichmentofKEGGpathwaysforCNVRharboredgenes.xlsx](#)
- [Additionalfile8TableS7.InformationaboutQTLsoverlappingwithidentifiedCNVRs.xlsx](#)
- [Additionalfile9TableS8.ListofVstvaluesandannotatedgenes.xlsx](#)
- [Additionalfile10FigureS2.TheqPCRvalidationresultsfor10selectedCNVloci.pdf](#)
- [Additionalfile11TableS9.Informationofthe10selectedCNVsandprimerpairsforqPCRvalidation.xlsx](#)
- [Table1SummaryofCNVsandCNVRsidentifiedinthreeTibetansheep populations.xlsx](#)
- [Table2CNVRharboringgenesshowinghighdivergenceintwopopulationpairs.xlsx](#)