



OPEN

Genome-wide evaluation of copy gain and loss variations in three Afghan sheep breeds

Mohammad Hossein Moradi¹✉, Roqiah Mahmodi¹, Amir Hossein Khaltabadi Farahani¹ & Mohammad Osman Karimi²

Copy number variation (CNV) is one of the main sources of variation between different individuals that has recently attracted much researcher interest as a major source for heritable variation in complex traits. The aim of this study was to identify CNVs in Afghan indigenous sheep consisting of three Arab, Baluchi, and Gadik breeds using genomic arrays containing 53,862 single nucleotide polymorphism (SNP) markers. Data were analyzed using the Hidden Markov Model (HMM) of PennCNV software. In this study, out of 45 sheep studied, 97.8% (44 animals) have shown CNVs. In total, 411 CNVs were observed for autosomal chromosomes and the entire sequence length of around 144 Mb was identified across the genome. The average number of CNVs per each sheep was 9.13. The identified CNVs for Arab, Baluchi, and Gadik breeds were 306, 62, and 43, respectively. After merging overlapped regions, a total of 376 copy number variation regions (CNVR) were identified, which are 286, 50, and 40 for Arab, Baluchi, and Gadik breeds, respectively. Bioinformatics analysis was performed to identify the genes and QTLs reported in these regions and the biochemical pathways involved by these genes. The results showed that many of these CNVRs overlapped with the genes or QTLs that are associated with various pathways such as immune system development, growth, reproduction, and environmental adaptations. Furthermore, to determine a genome-wide pattern of selection signatures in Afghan sheep breeds, the unbiased estimates of F_{ST} was calculated and the results indicated that 37 of the 376 CNVRs (~10%) have been also under selection signature, most of those overlapped with the genes influencing production, reproduction and immune system. Finally, the statistical methods used in this study was applied in an external dataset including 96 individuals of the Iranian sheep breed. The results indicated that 20 of the 114 CNVRs (18%) identified in Iranian sheep breed were also identified in our study, most of those overlapped with the genes influencing production, reproduction and immune system. Overall, this is the first attempts to develop the genomic map of loss and gain variation in the genome of Afghan indigenous sheep breeds, and may be important to shed some light on the genomic regions associated with some economically important traits in these breeds.

Livestock species play major socio-economic roles in the world, since they provide many goods and services to human populations. The sheep (*Ovis aries*) is one of the most important farm animal species, because of its high productive potential and its ability to adapt to harsh environments^{1,2}. The breeding of sheep in Afghanistan dates back to about 9000 years ago when these valuable livestock species were domesticated³. The grassland in Afghanistan is around 30 million hectares, accounting for about 47% of Afghanistan's land area. The Hindu Kush slopes are a natural and permanent pasture for sheep breeding in Afghanistan and the production of sheep is of great importance in the livestock sector. Before the war and the recent drought, approximately 90% of the foreign exchange income of the livestock sector was derived from sheep, and livestock industry have a contribution up to 40% in country's exports³. Despite the importance of sheep production in Afghanistan, the lack of accurate pedigree and phenotypic records, are the main problems that limit the implementation of animal breeding programs.

Recently, the genome-wide association studies have successfully uncovered single-nucleotide polymorphisms (SNPs) associated with complex traits such as diseases and quantitative phenotypes⁴. However, various studies showed that these variations account for a small proportion of heritability⁵. With the development of high throughput techniques, abundant structural variations have been found in the genome, of which the main variations are copy number variations (CNVs). The CNVs are increasingly recognized as an important genetic source

¹Department of Animal Science, Faculty of Agriculture and Natural Resources, Arak University, Arak 38156-8-8349, Iran. ²Department of Animal Science, Faculty of Agriculture and Natural Resources, Herat University, Herat, Afghanistan. ✉email: Hosenmoradi@ut.ac.ir; h-moradi@araku.ac.ir

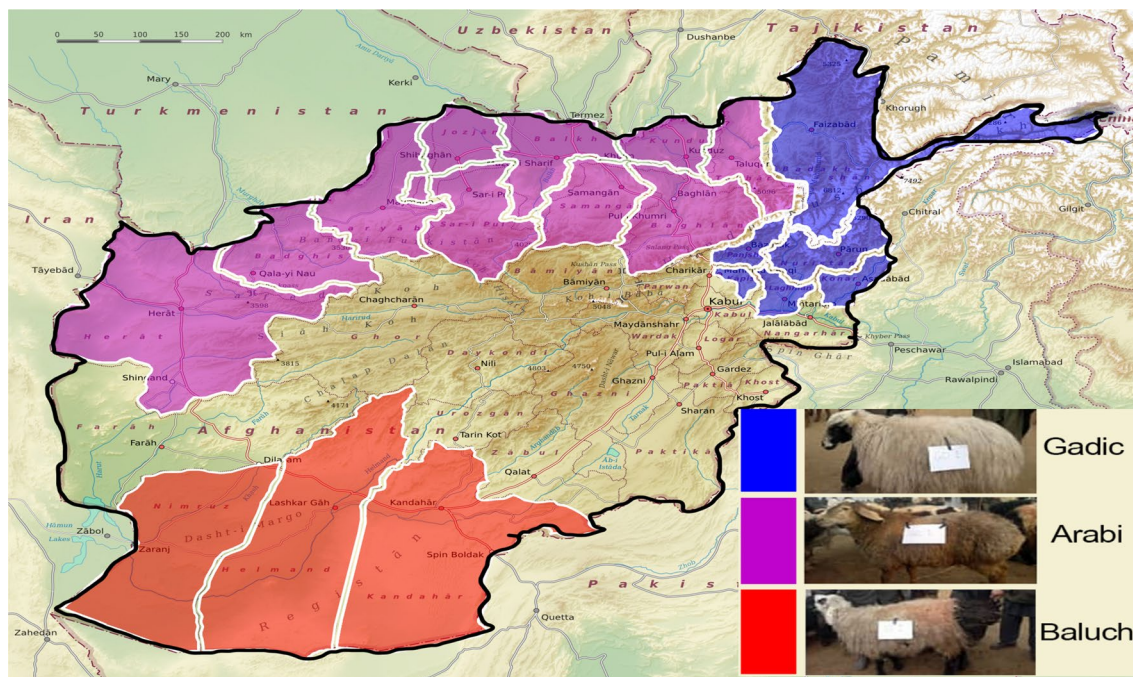


Figure 1. The studied breeds and their geographic distribution in different regions of Afghanistan (the figure has been created by using Adobe Photoshop CC 2020 v.21.0.2, <https://www.adobe.com/products/photoshop/>).

of different phenotypic traits^{6–8}. CNVs have been shown to explain around 18% of genetic variation in the gene expression, having no overlap with that explained by SNPs⁹. Initial research on CNV traces to 1936, when C. Bridges found an association between the *BAR* “gene”, duplication of a part of a chromosome, and the eye size in *Drosophila melanogaster*¹⁰. Since 2006, several studies in humans and animals have identified the association between CNV polymorphisms and different complex phenotypic traits⁷. Several studies have shown the usefulness of CNVs data for describing the genetic basis of phenotypic variations in a variety of species like cattle^{11,12}, dog¹³, chicken^{14,15}, pig^{16,17}, goat^{18,19}, and rabbit²⁰. As for sheep, Fontanesi et al.²¹ provided the first comparative map of CNVs in the sheep genome using a cross-species array comparative genome hybridization (aCGH). It has been reported that the duplication of the agouti signaling protein gene is associated with white coat color in sheep²², a gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep²² and higher levels of *KIF2A* and *PHKG2* expression in Gansu Morden sheep are associated with disease resistance²³. These studies suggest that identifying the CNVs distributed across the sheep genome can be very important for finding the genomic regions associated with some economically important traits in this species.

The aim of this study was to identify the genome-wide copy number variations (CNV) in three Afghan sheep breeds, and to study the associations between these regions of interest with different biological pathways. The results of this study, as one of the first attempts to develop the genome-wide map of CNVs in local Afghan sheep breeds, can provide a valuable information genomic regions associated with economically important traits in these breeds.

Materials and methods

Animal samples. A total of 45 individuals, sampled from three breeds including 15 samples per each breeds of Arab, Baluchi and Gadik were collected for this study from different parts of Afghanistan (Fig. 1). The breeds used in this study are the most common and main indigenous sheep breeds reared in Afghanistan that have been prioritized over the past few years by animal breeding researches and private investment sectors²⁴. Arab, Baluchi, and Gadik breeds account for around 13, 13, and 1% of total sheep reared in Afghanistan, respectively. Among these breeds, the Arabian sheep have a large body size and are the more capable for meat producing and known as a meat sheep, the Baluchi breed are well adapted to arid and desert environments and are known as dual purpose for wool and meat production and the Gadik breed as a wool sheep, is producing the best type of wool among other breeds²⁴. The color of wools for each animal was collected at the time of blood sampling. While the color of wool for all Gadik sheep breed was white, a combination of different colors was observed in other two breeds specially in Baluchi. The white coats are favored in sheep for wool production that is easily dyed.

DNA extraction, genotyping and quality control. The blood samples were collected via the jugular vein in the tubes containing EDTA (ethylene diamine tetraacetic acid) and the DNA was extracted using the GenElute Blood Genomic DNA Kit²⁴. Genotyping was carried out in all animals using Illumina Ovine SNP 50K BeadChip, which contains 53,862 SNPs spanning the ovine genome. Quality control was implemented using the PLINK 1.90 software²⁵. We removed the individuals and SNPs meeting any of the following criteria: individuals with a genotyping rate <0.90; SNPs with >0.05 missing data; and minor allele frequency (MAF) <0.02²⁶; SNPs

with a P-value of < 0.000001 for the Fisher's exact test of Hardy–Weinberg equilibrium (HWE). Finally, the SNPs without chromosomal or physical locations or located on the sex chromosome were excluded from final analysis. The markers located on the sex chromosomes were excluded from subsequent analyzes due to the different distribution of IBD (Identical by descent) pattern and the effective population size that affects identifying the CNVs²⁷. This resulted in a data set of 45 individuals genotyped for 47,519, 46,430 and 47,711 autosomal SNPs in Arab, Baluchi and Gadik breeds, respectively.

Identification of CNV and CNVR. CNVs were inferred within each individual using PennCNV v.1.0.5 software²⁸. A Hidden Markov Model (HMM) algorithm as employed in PennCNV incorporates multiple parameters, such as signal intensity ratios (log R Ratio: LRR) and allelic frequencies (B allele frequency: BAF) values of each marker in each individual, and the population frequency of B allele (PFB) of SNPs. The software extracts the signal intensity of each allele from the array genotyping data and then combines allelic frequency information and single nucleotide polymorphism distance to identify CNVs. For this purpose, the values of X and Y, which is the signal intensity of alleles A and B were used. R is calculated as the total signal intensity of two alleles, or $R = X + Y$. As a normalized measure of total signal intensity, the LRR value for each SNP is then calculated as $LRR = \log_2(R_{\text{observed}}/R_{\text{Expected}})$, which R_{observed} is computed from linear interpolation of canonical genotype clusters²⁹ obtained from a set of reference samples. The θ value measures the relative allelic intensity ratio of two alleles and is calculated from Eq. (1) which ranges from 0 to 1.

$$\theta = \arctan\left(\frac{Y}{X}\right) / \left(\frac{\pi}{2}\right) \quad (1)$$

The B Allele Frequency (BAF) refers to a normalized measure of relative signal intensity ratio of the B and A alleles²⁸:

$$BAF = \begin{cases} 0, & \text{if } \theta < \theta_{AA} \\ 0.5(\theta - \theta_{AA}) / (\theta_{AB} - \theta_{AA}), & \text{if } \theta_{AA} \leq \theta \leq \theta_{AB} \\ 0.5 + 0.5(\theta - \theta_{AB}) / (\theta_{BB} - \theta_{AB}), & \text{if } \theta_{AB} \leq \theta \leq \theta_{BB} \\ 1, & \text{if } \theta \geq \theta_{BB} \end{cases} \quad (2)$$

in which θ_{AA} , θ_{AB} and θ_{BB} are the θ values for three canonical genotype clusters generated from external reference samples (such as HapMap samples). We retrieved signal intensity ratios (LRR) and allelic frequencies (BAF) for each SNP using the Illumina Genome Studio v1.0 software. To increase the confidence of the detected CNVs, quality control was performed by employing standard exclusions of the LRR (standard deviation of LRR) < 0.3 , a BAF drift < 0.01 and a waviness factor < 0.05 (<http://www.openbioinformatics.org>).

CNV regions (CNVR) were determined by aggregating overlapped CNVs, identified in different animals, as reported previously³⁰. For construction of the CNVR map, we classified the status of these CNVR into three categories, 'Loss' (CNVR containing deletion), 'Gain' (CNVR containing duplication) and 'Both' (CNVR containing both deletion and duplication). Chromosomal distribution of CNVRs in Afghan sheep breeds has been developed in R v4.0.2. The VENN diagram web tool was also used to create a Venn diagram showing the overlap between CNVRs identified in different breeds [<https://bioinformatics.psb.ugent.be/webtools/Venn/> (accessed on July 3, 2021)].

Finally, to evaluate the CNVRs identified in the present study in more detail, we compared them to the CNVRs identified in an Iranian sheep breed. In particular, 96 animals belong to Iranian Baluchi sheep breed were used. This breed is rearing in eastern parts of Iran, close to the borders of Afghanistan, and genotyped using the Illumina Ovine SNP50k Beadchip that fully described previously by Gholizadeh et al.³¹. The same statistical methods used in this study was applied to detect CNV and CNVRs in Iranian Baluchi sheep.

Genome-wide detection of selection signatures. To determine a genome-wide pattern of selection signatures in Afghan sheep breeds, the unbiased estimates of F_{ST} (Theta) was calculated as described by Weir and Cockerham³². Three comparisons consisting Arabi vs Baluchi, Arabi vs Gadik and finally Baluchi vs Gadik was performed. The values can theoretically range from 0 (showing no differentiation) to 1 (indicating complete differentiation, i.e. populations are fixed for different alleles). For each set of 5 adjacent SNPs, the average of F_{ST} values was calculated and termed windowed F_{ST} . This is an approximate method of looking for regions where selection is apparent over multiple markers, rather than one-off high values. A window of 5 markers was chosen as it appeared to provide the better signal compare to other arbitrary window sizes²⁶. The windowed F_{ST} values were then plotted against genome location. All scripts were written and performed in R v 4.0.2 (<https://cloud.r-project.org/>).

Gene and QTL content and Cytoscape analysis. Genes harbored in the previously identified CNVRs were obtained from the Ensembl Genes 64 Database using BioMart software (<https://www.ensembl.org/biomart/>) based on the *Ovis aries* gene sequence assembly (Oar_v3.1). To functionally annotate these CNVRs, the gene ontology analysis (GO)²⁹ was conducted using the Database for Annotation, Visualization, and Integrated Discovery Cytoscape functional annotation tool (<http://www.cytoscape.org/>). Also, to better understand the functions of the genes within the detected CNVRs, the *O. aries* Ensembl gene IDs were converted to human ortholog Ensembl gene IDs using BioMart because annotation of the sheep genome was limited³³. In addition, to study whether these CNVRs were associated with economically important traits, these CNVRs were compared

Descriptive statistics	Arab	Baluchi	Gadik	All animals
Number of CNVs	306	62	43	411
Total length of CNV (kb)	128,897	9864	5872	144,635
The proportion of CNVs length divided by whole-genome length per each animal ¹	3.44	0.26	0.17	1.31
The average length of CNVs (kb)	421	159	137	352
The median length of CNVs (kb)	301	103	102	214
Number (percentage) of CNVs range 20–100 kb	51 (17%)	28 (45%)	21 (49%)	100 (24%)
Number (percentage) of CNVs range 100–500 kb	173 (56%)	32 (52%)	22 (51%)	227 (55%)
Number (percentage) CNVs with the length of > 500 kb	82 (27%)	2 (3%)	0	84 (21%)

Table 1. Summary of CNVs identified in Afghan sheep breeds. ¹The length of the sheep genome is considered to be 2500,000 kb (<https://www.ensemble.org>).

with QTLs in the sheep QTL database (<https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>, sheep QTL database, accessed on 10 May 2022).

Approval for animal experiments. The experimental protocols were approved by the Biomedical Ethics Committee of Arak University. All methods were carried out in accordance with relevant guidelines and regulations. The authors also complied with the ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines.

Results and discussion

Genome-wide detection of CNVs. The large-scale CNV explorations was performed and the results identified 411 CNVs across 45 individuals from three breeds, where the total length was 144 Mb and the average length for each individual was 9.13 Mb (Table 1). Among them 306, 62 and 43 CNVs was observed in Arabian, Baluchi and Gadik sheep breeds, respectively. The length of CNVs compared to whole-genome length, per each animal was 1.13%. The average length of the CNVs was 352 kb, and the median length was 214 kb. We found that approximately 24% of the CNVs range from 20 to 100 kb, 55% of CNVs range from 100 to 500 kb in size, and 21% being larger fragment CNVs (> 500 kb) (Table 1).

In this study, CNVs were observed in all autosomal chromosomes, that was in agreement with the results obtained in previous studies in Chinese sheep³⁴, native Korean Hanwoo cattle³⁵, Chicken¹⁴ and human³⁶. The average of CNVs per each animal in the present study was 9.13. A variety of CNVs number per each individual have been reported, including less than 1²³ to 16.9²¹ in sheep, 17.9 in goat¹⁸, 9.6 in chicken¹⁴, 3.1 for horse³⁷, and 19.16 in Iranian native cattle³⁸. The number and the length of CNVs are influenced by a variety of factors including the sample size, the statistical method, species, and markers density^{39,40}. For instance, while 45 samples from three Afghan sheep breeds (15 samples per each breed) have been used in the present study, Fontanesi et al.²¹ studied 11 samples (1 up to 2 samples per each breeds) to develop the first comparative map of CNVs in Italian dairy or dual-purpose sheep breeds using a cross-species array comparative genome hybridization (aCGH). Ma et al.²³ were identified the CNVs in the genomes of eight sheep breeds (20 samples per each breed) using the sheep SNP50k BeadChip genotyping array, where, a total of 173 CNVs were discovered by the average and median sizes of 123 kb and 100 kb, respectively. The number of CNVs obtained in Afghan sheep breeds were in the range of previously reported studies by Fontanesi et al.²¹ and Ma et al.²³ and the differences may be caused due to different breeds of animal and the number of samples.

Genome-wide identification of autosomal CNVRs. CNVRs were determined by merging the overlapping CNVs identified in all samples, as reported previously^{6,30}. A total of 376 autosomal CNVRs (125 losses, 242 gains, and 7 both event) were identified. The average and median sizes of CNVRs were 361 kb and 224 kb, respectively. Approximately 25% of the CNVRs ranged from 20 to 100 kb, 53% ranged from 100 to 500 kb and 22% of the CNVRs were larger than 500 kb (Table 2). Detailed information on CNVRs in Afghan sheep breeds is presented in Table 2.

The map of the CNVRs distribution on autosomal chromosomes was created in total samples of Afghan sheep and has been shown in Fig. 2.

Keeping in mind that the detection rate of CNVRs is affected by many factors⁴¹, the results of this study were compared with previous studies in sheep. Between 111⁴² and 3488⁴³ CNVRs have been detected in previous studies, with CNVR lengths of 10.56²¹ up to 120.53²³ Mbs. The CNVRs identified in these studies were different to some extent, which may be due to the differences in sheep breeds, CNV detection platforms and CNV calling algorithms used. Similar findings were also observed in CNVR studies conducted in different mammals including pig⁴⁴ dairy and beef cattle⁴⁵ and goat breeds⁴⁶. The use of different CNV calling algorithms has a substantial effect on the results of CNVR studies. Currently, three algorithms for CNV detection based on SNP arrays have been developed, which are available in different programs, including PennCNV, cnvPartition and QuantiSNP²³. Compared to other algorithms, PennCNV provides a standard proven method that is more reliable for assessing the number of copies when using Illumina arrays because it incorporates the allelic intensity ratio at each SNP marker and the total signal intensity, the allele frequency of SNPs, the distance between neighboring SNPs, and

Descriptive statistics	Arab	Baluchi	Gadik	All
Number of CNVRs in deletion state	50	44	31	125
Number of CNVRs in duplication state	233	3	8	242
Number of CNVRs in deletion/duplication state	3	3	1	7
Total length of CNVR (kb)	122,811	7667	5275	135,756
Average of CNVRs length (kb)	429	153	132	361
Median of CNVRs length (kb)	306	100	100	224
Number (percentage) of CNVRs range 20–100 kb	46 (16%)	27 (54%)	20 (50%)	93 (25%)
Number (percentage) of CNVRs range 100–500 kb	160 (56%)	21 (42%)	20 (50%)	201 (53%)
Number (percentage) CNVRs with the length of > 500 kb	80 (28%)	2 (4%)	0	82 (22%)

Table 2. Summary of CNVRs identified in Afghan sheep breeds.

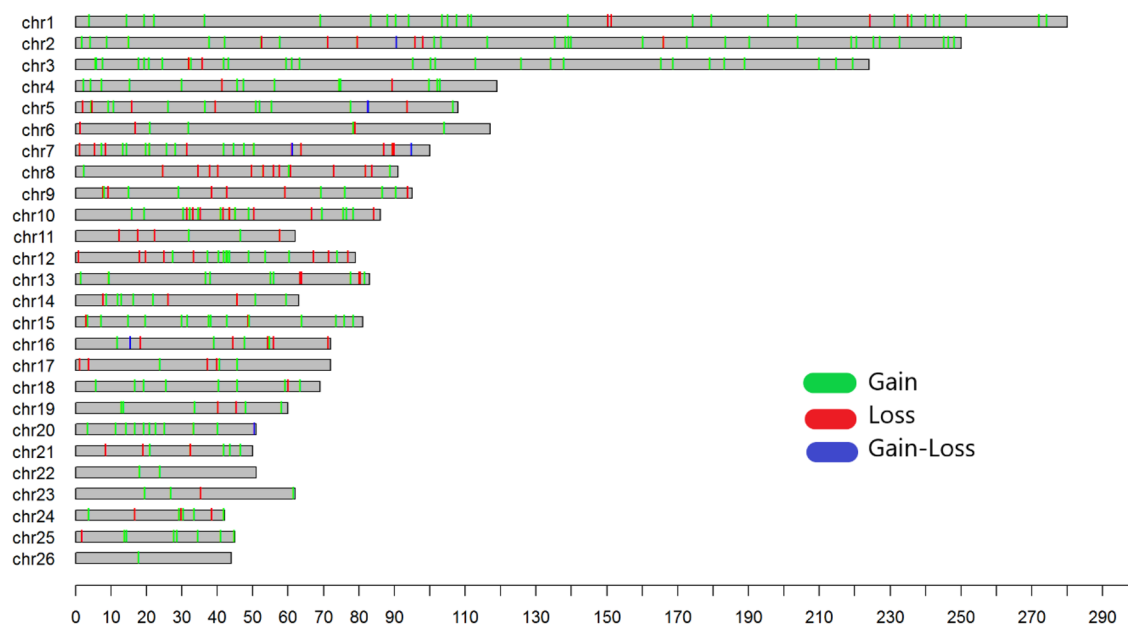


Figure 2. Chromosomal distribution of copy number variant regions (CNVRs) in Afghan sheep breeds, developed using R v4.0.2 (<https://cloud.r-project.org/>): in this figure the chromosome length (Mb) is displayed on the X-axis and CNVR variations with deletion, duplication, and deletion/duplication state showed on different chromosomes on Y-axis.

the GC content to overcome biases⁴⁷. Also, the software packages currently commonly used for CNV detection include PennCNV, CNVcaller, and CNVnator, each of them have their own advantages and disadvantages. PennCNV software has been extensively applied to Illumina chip data⁴⁸ while, CNVcaller and CNVnator software are using read depth methods to detect CNVs in resequencing data^{33,49}.

The results indicate that the number of CNVRs in the chromosomes varies from 1 to 39 (Additional file Table S1), with the top three highest rate of CNVRs (more than 30) located on chromosomes 2, 1, and 3, which contained 39, 34, and 31 CNVRs respectively. Chromosome 26 had the smallest number of CNVRs. This observation was in agreement with Yuan et al.⁵⁰ where they reported that the chromosome length had a significant positive linear relationship with the number of CNVRs ($R^2 = 0.87$) with the maximum found in chromosome 1 and the minimum found on 26. Conflicting results have also reported by Jenkins et al.⁴³ in Texel, Coopworth, Perendale, Romania, and Merino sheep and Zhu et al.⁴⁸, in the study of CNVR-containing genomic regions in Tibetan sheep, stated that chromosome 3 had the highest number of CNVRs. These results may be influenced by the number of samples and arrays used. For example, while 50k chips were used in the present study, Zhu et al.⁴⁸ used the chips with a 600k marker density. However, the medium density of Illumina OvineSNP50 BeadChip array used in our study, provides a standard proven method to detect CNVs and has been widely used in CNV detection in different sheep breeds^{41,51–53}. It was also observed in the present study that CNV duplication state is more than the deletion state, which may be because the deletions are under stronger purifying selection (which removes deleterious variants from the population) than duplications, therefore deletions should on average, be both less frequent and smaller than duplications³⁰.

A comparison of the CNVRs among breeds is shown in the Venn diagram (Fig. 3). The highest number of shared CNVRs was observed between Arabi and Gadik ($n = 8$), and the lowest number, was shared between

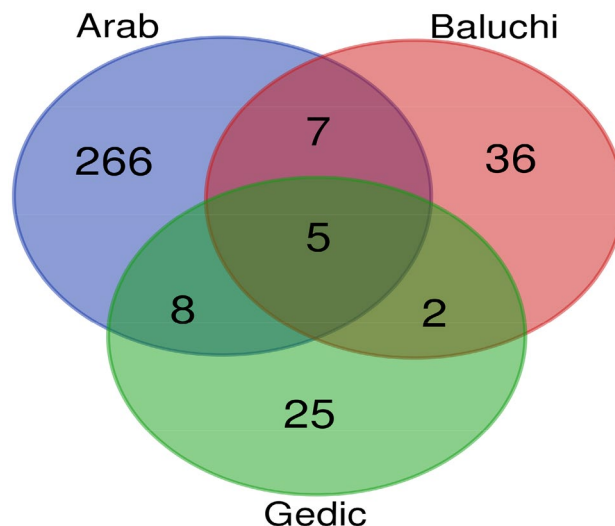


Figure 3. Venn diagram representing common and unique CNVRs found among the three Afghan sheep breeds. This figure was created using VENN diagram web tool (<https://bioinformatics.psb.ugent.be/webtools/Venn/>).

Baluchi and Gadic ($n=2$). Long-term adaptation to different environmental conditions or different selection schemes increases the variation in CNVRs among breeds³⁰. Similar environmental conditions, gene flow, and shared ancestral components may have led to an increase in shared CNVRs among breeds⁴¹. The more CNVRs identified in Arab sheep and the highest shared CNVRs observed between Arab and Gadic are in agreement with the distribution of these breeds in Afghanistan. Arab sheep breed is rearing in a wider geographic ranges and therefore, is adapted to different climatic conditions. Also, Arab and Gadic sheep breeds are rearing close to each other and gene flow may have occurred during last generations, leading to an increase in shared CNVRs among breeds (Fig. 1).

Gene content and functional annotation. The BioMart system in ENSEMBL was used to retrieve the gene content in the 376 CNVRs (125 losses, 242 gains, and 7 both event). Our results revealed that some CNVRs did not overlap with any annotated gene that was in agreement with the results reported in previous studies on cattle⁵⁴ and sheep⁵². This may be due to the genome assembly version, especially in non-human species⁵⁴, and also several studies have highlighted the genome-wide distribution of CNVs in regions covering non-coding sequences, thus affecting the regulation of distant target genes⁵⁵. However, many genes identified in these regions had economically important biological functions including *IL37* genes, affecting immune response, *DCT* gene with an important role in melanin synthesis and color expression in skin and feathers, *CACNA1A* gene influencing contraction process muscles and the *TCF12* gene affecting skeletal muscle, thyroid, B and T cells (GeneCards). The *IL37* gene is located on chromosome 3, which is involved in pro-inflammatory and anti-inflammatory reactions that it is one of the candidate genes influencing the immune response and resistance to mastitis in cattle⁵⁶. The *DCT* gene is located on chromosome 10, which is involved in the production of the pigment Eumelanin and Feomelanin. In mammals, skin and hair color are depending on the relative ratio of Eumelanin and Feomelanin⁵². The *CACNA1A* gene on chromosome 5 involve in muscle contraction, hormone or neurotransmitter secretion, and gene expression (GeneCards). Zhu et al.⁵⁷ have shown that this gene is associated with fat deposition in fat tailed sheep breeds. The *TCF12* gene which is situated on chromosome 7, is expressed in many tissues, including skeletal muscle, thyroid, B, and T cells (GeneCards). Dunner et al.⁵⁸ reported that this gene affects muscle fatty acid production in *Bos Taurus* cattle.

To determine the underlying biological functions of the genes overlapped with CNVRs, the Gene Ontology (GO) analysis was used and the genes were evaluated for their affects on biological process (BP), cellular component (CC), molecular function (MF), and immune system. Our results showed that the most of the genes were associated with sensory perception and immune system activity, interleukins, and growth performance (Table 3). The same pathways have previously been reported for the development of skeletal muscle and immune system in Altai and Tibetan sheep⁴⁸ and in antigenic processes and immune responses by Fontanesi et al.²¹ and Hou et al.³⁴. Liu et al.⁶ reported the genes that contributed in the adaptation to environmental conditions, which is consistent with the results of the present study.

A circular diagram of the biological pathways involved by the genes reported in CNVR-containing regions of all Afghan sheep breeds has been shown in Fig. 4. The results showed that 11.73% of the genes affected the positive regulation of cell development. This pathway refers to any process that increases the speed, frequency, or rate of development of a cell from its formation into a mature structure (QuickGO). The ratios of 8.02, 8.02, and 6.79%, were also influencing the pathways of intracellular organelle part, actin cytoskeleton organization, and binding membrane of organelles, respectively (Fig. 4). The identification of GOTerms in this study and their comparisons with other studies showed a number of similar pathways. Most of these similar pathways include

GOID	GO term	P-value
GO:0007600	Sensory perception	1.9E-6
GO:0050907	Detection of chemical stimulus involved in sensory perception	9.5E-5
GO:0050906	Detection of stimulus involved in sensory perception	1.4E-11
GO:0007606	Sensory perception of chemical stimulus	1.8E-11
GO:0007608	Sensory perception of smell	1.4E-13
GO:0048588	Developmental cell growth	9.8E-3
GO:0005149	Interleukin-1 receptor binding	4.5E-5

Table 3. Summary of the results from the ontology of genes identified in CNV regions in all samples.

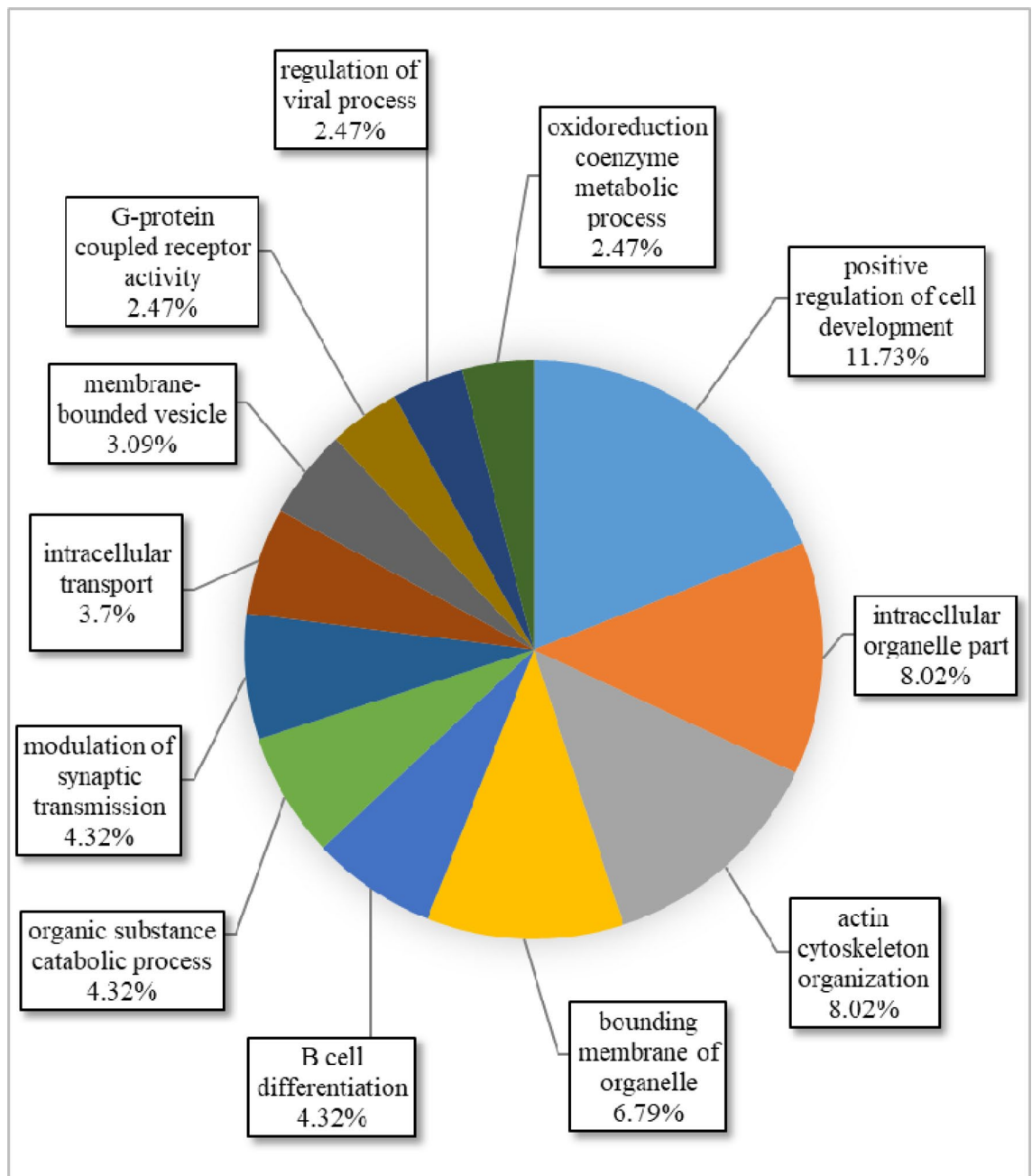


Figure 4. Circular diagram of the biological pathways involved by the genes reported in CNV-containing regions of all Afghan sheep breeds produced by Cytoscape functional annotation tool (<http://www.cytoscape.org/>): in this figure, the name of each process and the degree of the gene participation have been reported. The pathways with a participation rate of less than 2% are not shown.

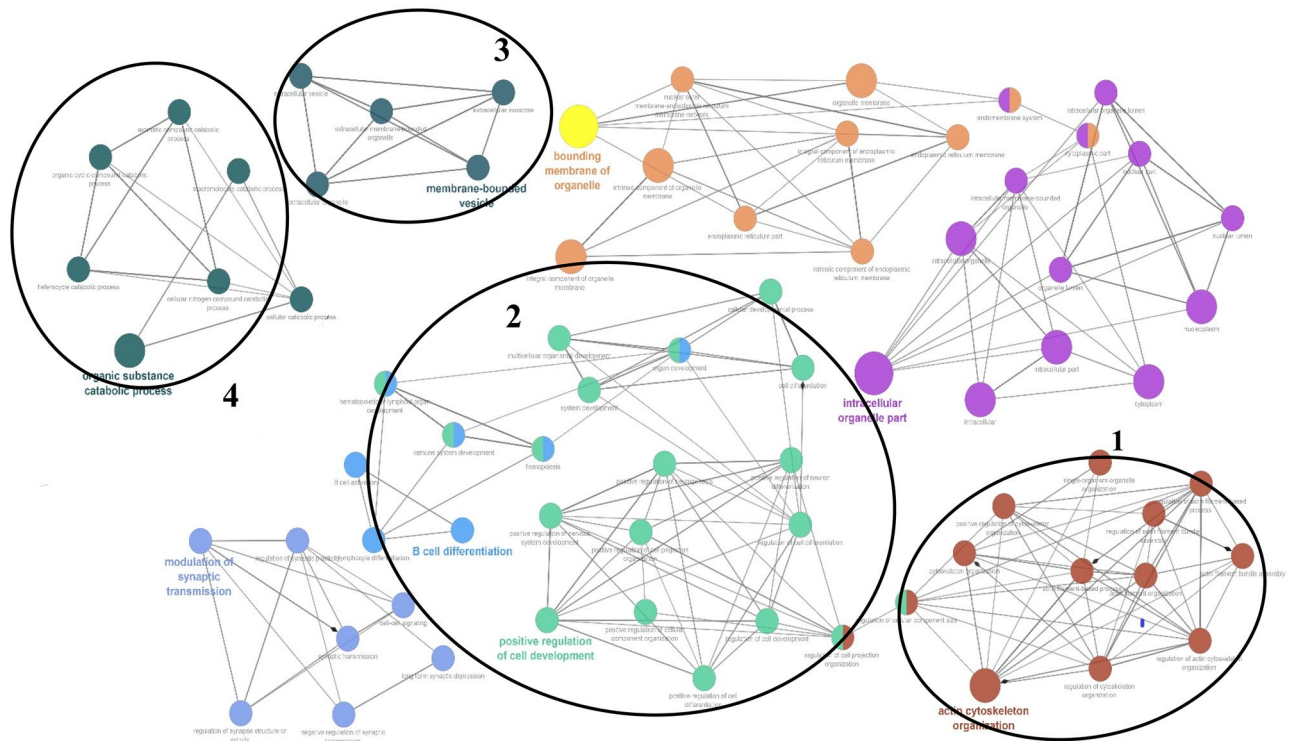


Figure 5. Gene-network interactions between all the genes reported in the regions containing CNVRs in Afghan sheep breeds, produced by Cytoscape functional annotation tool (<http://www.cytoscape.org/>). Circles of different colors represent (biochemical pathways) and the lines between these circles indicate the interactions between these pathways with each other. Zone 1 include the genes that affecting the skeletal muscle, zone 2 influence the birth weight, fertility, embryonic development and skeletal morphology, zone 3 is containing the genes that effect on skeletal development, cartilage structure, and joint function and zone 4 plays a role in fetal development and development.

the G-Protein coupled receptor signaling pathway (GO:0007186)^{6,34}, Sensory perception (GO:0007600)⁶, neurological system process (GO:0050877)³⁴, Cation binding (GO:0043169)⁷, metal ion binding (GO:0046872)⁷ and enzyme binding (GO:0019899)²³. The study of these GO Terms in other species also showed some reported pathways that are consistent with the present study, for instance, G-Protein coupled receptor signaling pathway (GO:0007186) in cattle⁵⁹.

The gene networks involved by the genes identified in the present study are shown in Fig. 5. Four gene networks related to different traits were created and several co-associated genes in these networks are involved in biological processes previously found to be associated with production and reproduction performance traits. Network 1 in stance, included different genes, such as *ZNF* and *CAPZB* that were previously reported to be associated with muscle structure and meat production⁶⁰. Genes in cluster 2 primarily included *HINT2*, *TNFRSF19*, *CCIN*, *FNDC3A*, *SYCP2*, *TUBGCP3*, and *RECK*, which were related with birth weight, body size, fertility, embryonic development, as well as skeletal morphology^{61,62}. Cluster 3 included some genes such as *ACAN* and cluster 4 the genes like *NID2* which were previously reported to be associated with skeletal development and cartilage structure³⁶ and embryonic development⁶³, respectively.

As the current annotation of the sheep genome is not as comprehensive as *H. sapiens*²⁶, the regions of interest in *O. aries* were compared to the corresponding areas in human (Fig. 6). The results of gene ontology showed that 8.85% of the genes that overlapped with CNV regions affected the actin filament bundle assembly pathway. This pathway consists a set of actin filament bands that are aligned in an axis with identical or opposite poles (approximately 30–60 nm apart) (QuickGO). Also, the ratios of 8.46, 7.31, and 5.38% were associated with cell projection, negative regulation of dendritic cell antigen presentation, and supramolecular fiber organization, respectively (Fig. 6). Given that actin and myosin are one of the most important components of muscle development, it seems that the development of the skeletal and muscle systems has been one of the most important targets during the evolution of the Afghan sheep. These pathways had a P-value of $3.4E-7$, $3.2E-5$ and $1.2E-11$ in *O. aries* and $2.6E-3$, $4.5E-5$ and $6.6E-9$ in *H. sapiens*, respectively.

Also, the gene networks developed by the genes identified in orthologous regions of *H. sapiens* were also evaluated and the results revealed that some networks are associated with muscle skeletal, embryonic, reproductive stages, birth weight, melanocyte, pigmentation, enteric nerves, skeletal development, immune response, antigens and antibodies (Fig. 7). Network 1 in stance, included different genes, such as *ROCK-1*, *ROCK-2* that were previously reported to be associated with muscle structure system in sheep⁶⁴ and network 2 included *PTGDR*, *HINT2*, *SPAG8*, *CCIN*, *TNFRSF19* and *ACAN* genes which were related with sex determination⁶⁵, fertility, fetal development, birth and carcass weights⁶⁶ and skeletal development and cartilage structure³⁶. Cluster 3 included

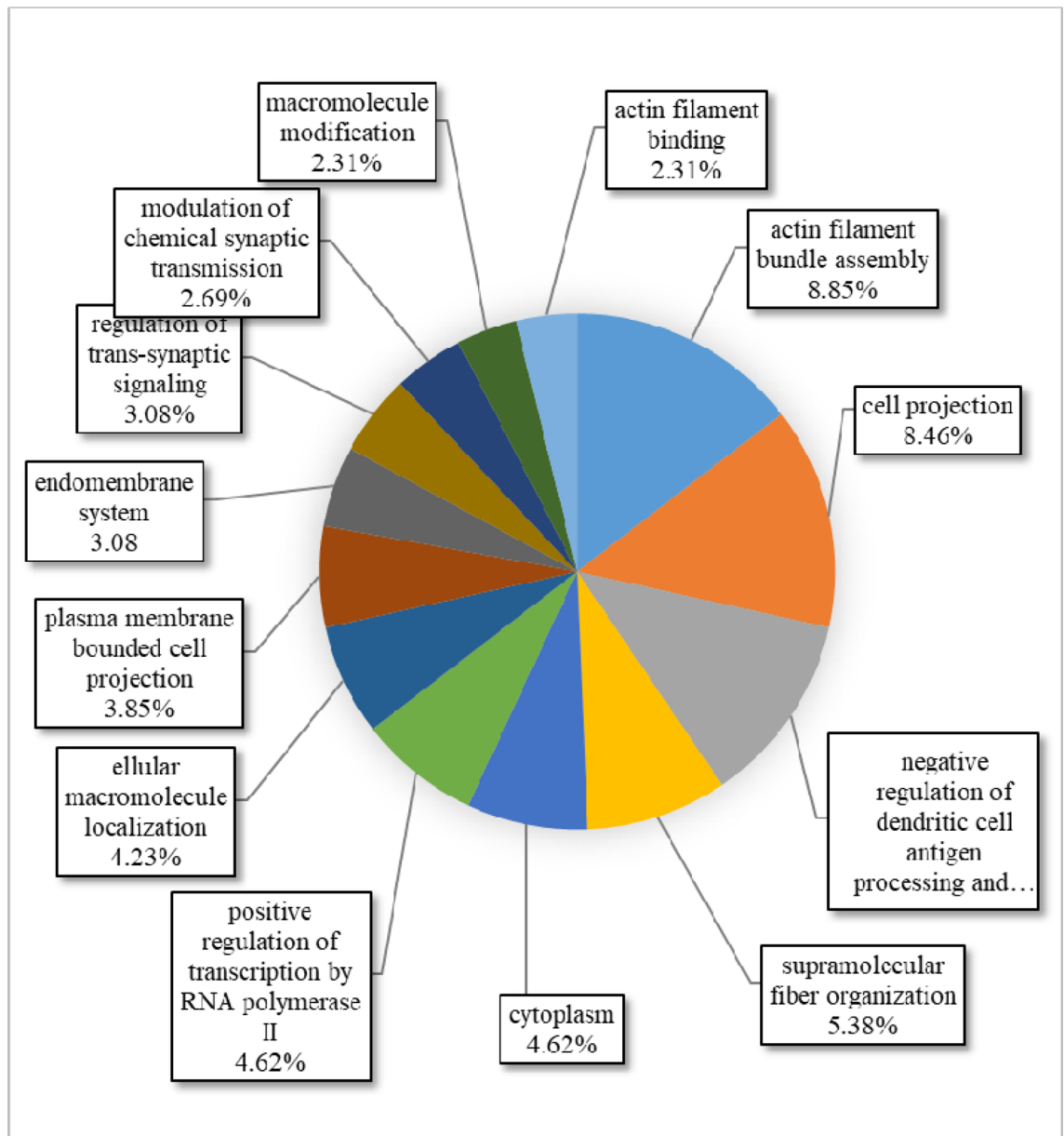


Figure 6. Circular diagram of the biological pathways involved by the genes reported in orthologous regions containing CNVs in *H. sapiens*, created by Cytoscape functional annotation tool (<http://www.cytoscape.org/>): In this figure the name of each process and its percentage has been showed. Pathways with a contribution of less than 2% were removed from the graph.

some genes such as *SYCP2*, *NPR2*, *EDN3* influencing birth weight, skeletal morphology and body size⁶⁷, and finally network 4 including the genes like *PRSS55*, *CRISP2* that are relate with immune system, antigens and antibodies⁶⁸.

Due to the high conservation of the genes between humans and sheep⁶⁹, genes that are known to be related to complex human traits may also be important for related traits in sheep. However, further research will be needed to formally test the functional relevance of these genes.

QTLs overlapping with identified CNVRs. To further reveal the CNVRs associated with sheep economic traits, the detected CNVRs were compared with quantitative trait loci (QTLs) in the sheep QTL database. We found 121 CNVRs overlapping with 171 QTLs, including milk production and quality, growth and carcass traits, reproduction performance, fleece yield, fecal egg counts, tail fat deposition, horn type and health-related traits (Additional file Table S1).

Many studies have shown that CNVRs contain QTLs associated with important economic traits in animals^{50,70}. Therefore, the CNVRs detected in this study were compared with the QTLs reported in the sheep QTL database. The QTL categories found in this study were basically identical to those found in previously reported studies^{50,52,71}. For example, the health-related QTLs found included fecal egg count QTLs, and worm count QTLs

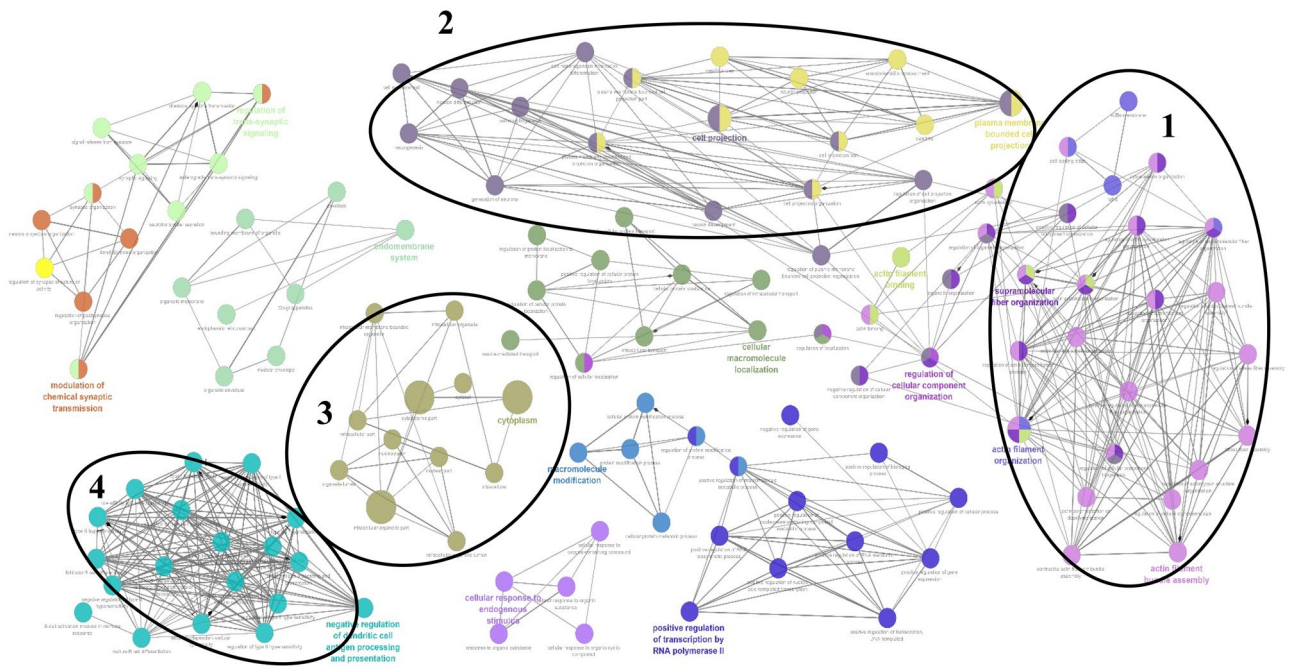


Figure 7. Gene-network interactions between all the genes reported in the orthologous areas of *H. sapiens*, for the genomic regions containing CNVRs in Afghan sheep breeds [the figure developed by Cytoscape functional annotation tool (<http://www.cytoscape.org/>)]. Circles of different colors represent biochemical paths and the lines between these circles indicate the interactions between these paths with each other, whose intensity is a measure of the assurance of a functional partnership. Region 1 contains the genes affecting the muscle skeletal region, region 2, the sex determination of males and females, birth weight, carcass weight, fetal development, skeletal development and cartilage structure region 3 is associated with birth weight and skeletal morphology and body size, and finally, region 4, affect the immune response, antigens, and antibodies.

that previous studies have reported that worm disease infection rates in sheep can exceed 70% in many countries, causing huge losses to the livestock industry⁷². Relative to barn-fed livestock, grazing livestock are more likely to be infected with worms. These results indicate that CNVs, which are a critical type of genetic variation, may have an important effect on sheep health.

Comparison of identified CNVRs with the regions that have been under selection signatures. To identify loci that have been targets of selection across different Afghan sheep breeds, the windowed F_{ST} was plotted against genomic location (Fig. 8). Variants with unusually large F_{ST} values are typically interpreted as being the targets of local selective pressures due to the hitch-hiking effect²⁶.

Comparison of identified CNVRs with the regions that have been under selection signatures indicated that, 37 of the 376 CNVRs (~10%) were also identified to be under selection pressure and confirmed by selective sweep signatures (Fig. 8 and Additional file Table S1). Almost all of these regions overlapped with the genes or QTLs that had previously been identified to be associated with production, reproduction and environmental adaptations traits in sheep (Additional file Table S1, Sheet 2). Zhu et al.⁵⁷ identified the CNVs and selection signatures on the X chromosome in Chinese indigenous sheep breeds with different tail types. Their results revealed 20 common genes that were shared by CNVRs and selection signatures, and were associated with fat, energy metabolism, and immune system.

Comparison of identified CNVRs with Iranian sheep breed. Finally, to characterize the CNVRs identified in the present study in more detail, we compared them to the CNVRs identified in an Iranian sheep breed namely Baluchi. This breed is fat-tailed and dual-purpose breed (wooly and meaty), which is well adapted to the dry and hot climate conditions of eastern Iran. Iranian Baluchi sheep has the largest population than other breeds of sheep in Iran (about 30% sheep population) and have selected for this comparison since, it is mostly rearing in the regions of Iran that are close to the boarder of Afghanistan³¹. After quality control, 45,673 autosomal SNPs from 87 animals were used for CNVRs detection. The results indicated that 20 of the 114 CNVRs (18%) identified in Iranian sheep breed were overlapped or located close to the CNVRs identified in our study (Table 4). Study of the genes reported in these regions of interest revealed that most of those overlapped with the genes influencing production, reproduction and immune system (Table 4).

The CNVRs identified in different breeds have previously compared in a variety of species including sheep²³, cattle⁹⁹ and chicken¹⁰⁰ to evaluate the reliability of the detected regions consisting CNVs. The high proportion of overlapped CNVRs observed in the current study between Iranian and Afghanistan sheep breeds may probably

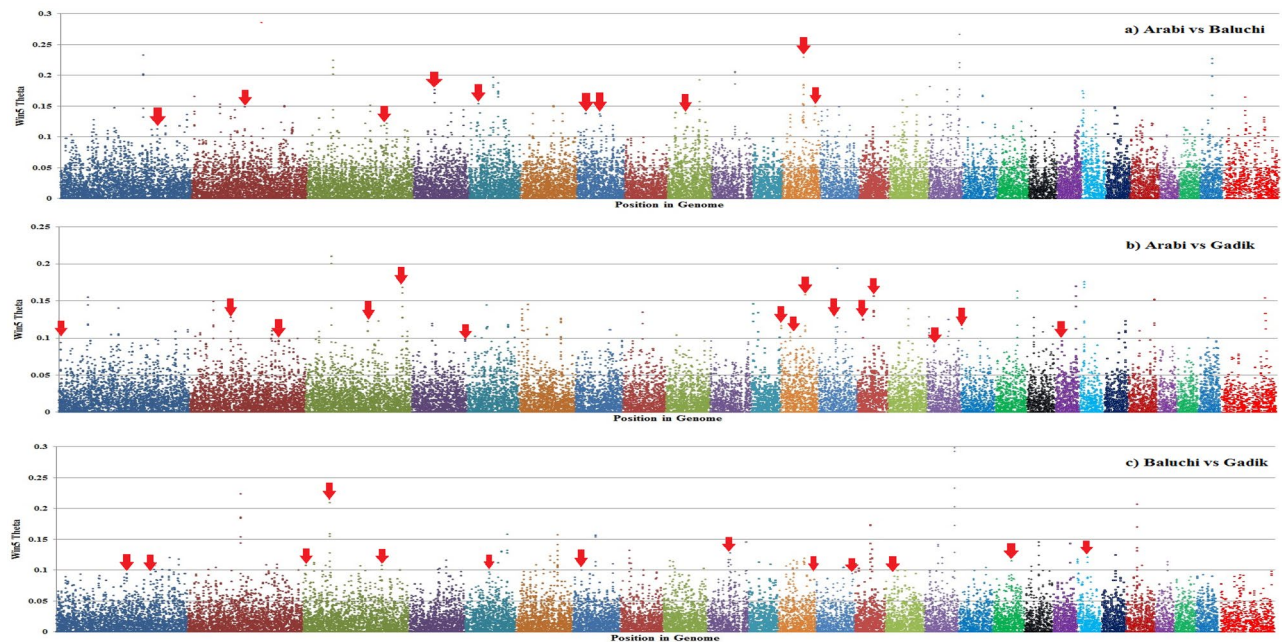


Figure 8. Distribution of selective sweep signatures for different comparisons in Afghan sheep breeds by chromosome. SNP position in the genome is shown on the X-axis and each chromosome was highlighted in different colors, and windowed F_{ST} (Theta) is plotted on the Y-axis. The red arrows are highlighting the peaks that overlapped by CNV regions identified in our study. This figure was created using R v 4.0.2 (<https://cloud.r-project.org/>).

reflect this fact that these breeds have been influenced by similar evolutionary events, mostly due to their close geographic distribution and may some gene flow occurred between these breeds in the past.

In general, this study is an important step towards the generation of a CNV map and provides an important resource for studies of ovine genomic variation in indigenous sheep of Afghanistan. However, the number of animals used in the current study may be one of the limitations affecting the number of CNVs identified in this study. Although, it should be considered that this is the only genomic information available from Afghan local sheep breeds and most of the similar published reports have been performed with almost the same or lower sample size in different indigenous sheep breeds^{7,21,42,50,57,94,101}, however, there were validation or phenotypes in most of the previous studies. For example, quantitative PCR (qPCR) has been used to confirm subsets of CNVs identified in different studies^{42,101}. Given that no validation of our findings has been performed in the current study, the results presented in this study is purely exploratory and external validation will be necessary to confirm the findings. It seems by increasing the number of genotyped samples in the future researches, we would be able to shed some new lights on the domestication event and the history of Afghan sheep. Also, further research on phenotypic associations would help understanding the role of CNVs in the formation of the local sheep breeds. Our results offer meaningful genomic insights that will help to guide future research and to provide a preliminary basis for the future exploration of the relationship between CNVs and economically important phenotypes of sheep.

Conclusion

Overall, the first map of CNVs on the genome of Afghan sheep breeds was developed in this study using genome-wide information. A total number of 411 CNVs and 376 CNVRs were identified that cover 144 Mb of the genome. Since each CNV variant contains more length of the DNA, comparing to single nucleotide polymorphism (SNP) markers, and a variety of previously published studies suggest that these regions have been associated with economically important traits, it seems CNVs can provide better information for realizing the missing inheritance related to complex traits in Afghan sheep breeds. The results of this study showed that the genes reported in the genomic regions containing CNVRs in this study are associated with various metabolic pathways including reproductive traits, carcass characteristics, and body weight, immune and musculoskeletal system. The results of the present study provide valuable information for analyzing the association of CNVs with economically important phenotypes in Afghan sheep.

Data availability

All CNVRs identified in Afghan sheep breeds have been deposited in figshare repository (<https://www.figshare.com>) with accession no. <https://doi.org/10.6084/m9.figshare.16577954.v1>. The genomic information used in this study belongs to private sheep breeding programs, so we do not have authorization to share it and are available from the corresponding authors upon reasonable request.

CNVs positions in Afghan sheep	CNVs positions in Iranian sheep	Distance (kb)	The genes reported in this location	Function of the gene	References
Chr1:242324429–242574074	Chr1:241583858–241653162	671	<i>SLITRK3</i>	Milk production traits	73
Chr2:172582236–172616581	Chr2:172272878–172424726	158	<i>MBD5</i> and <i>EPC2</i>	Growth traits	74
Chr5:39479796–39659262	Chr5:39400849–39585301	0	<i>COL23A1</i>	Fertility	75
Chr6:78371034–79041954	Chr6:78371034–78600405	0	<i>EXOC1L</i>	Reproduction	76
			<i>CEP135</i>	Milk production	77
			<i>AASDH</i>	Metabolism	78
Chr6:78796013–79165093	Chr6:79007752–79203136	0	<i>HOPX</i>	Sperm traits	79
Chr8:52949829–52997226	Chr8:52676280–52949829	0	<i>BACH2</i>	Immune system	80
Chr9:7615407–8155540	Chr9:7838010–7948027	0			
Chr10:19347073–19637375	Chr10:18607317–18670320	677	<i>RUBCNL</i>	Vertebrate autophagy	81
			<i>LRCH1</i>	Immune system	82
			<i>HTR2A</i>	Adaptive traits	83
Chr10:66683508–66747944	Chr10:65063529–65212062	977	<i>SLITRK6</i>	Milk production traits	73
Chr10:83028244–85695176	Chr10:85120520–85216381	0	<i>HS6ST3</i>	Growth traits	84
			<i>STK24</i>	Immune response	85
			<i>SLC15A1</i>	Development of growth	75
			<i>GPR18</i>	Immune system	86
			<i>TM9SF2</i>	Fertility	87
			<i>DOCK9</i>	Disease resistance	88
Chr12:27323634–27702125	Chr12:27961834–28032557	260	<i>DUSP10</i>	Immunity inflammation	89
Chr14:12743159–13111196	Chr14:13401157–13465959	290	<i>FOXC2</i>	Immune system	90
Chr15:75842476–75936079	Chr15:76144260–76212569	208			
Chr16:47568518–47871748	Chr16:46887875–47002471	566			
Chr17:23675899–23828049	Chr17:23538968–23738050	0			
Chr17:23675899–23828049	Chr17:23988299–24240466	0			
Chr21:18899984–19383548	Chr21:18655157–18820162	80	<i>LRFN4</i>	Ovarian cancer	91
Chr21:46501772–46700862	Chr21:46501772–46628225	0	<i>TENM4</i>	Nervous system development	92
			<i>NARS2</i>	Energy deficiency disorders	93
Chr25:14234573:14644765	Chr25:14561562–14644765	0	<i>PHYHIPL</i>	Growth and wool traits	94
Chr25:34269680:34689602	Chr25:34845271–34945170	156	<i>KCNMA1</i>	Growth traits	95
			<i>DLG5</i>	Muscle growth	96
			<i>POLR3A</i>	Immune response	97
			<i>RPS24</i>	Heat stress	98

Table 4. The list of CNVs overlapped between Afghan and Iranian sheep breeds and the genes and their functions reported in these locations.

Received: 9 February 2022; Accepted: 16 August 2022

Published online: 22 August 2022

References

- Kijas, J. W. *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* **10**, e1001258 (2012).
- Moradi, M. H., Khaltabadi-Farahani, A. H., Khodaei-Motlagh, M., Kazemi-Bonchenari, M. & McEwan, J. Genome-wide selection of discriminant SNP markers for breed assignment in indigenous sheep breeds. *Ann. Anim. Sci.* **21**, 807–831 (2021).
- Rashiq, M. H. Building carbohydrates from a livestock nutrition perspective. *Kabul Univ. J.* **1**, 15–24 (1995).
- Khanzadeh, H., Ghavi Hossein-Zadeh, N. & Ghovvati, S. Genome wide association studies, next generation sequencing and their application in animal breeding and genetics: A review. *Iran. J. Appl. Anim. Sci.* **10**, 395–404 (2020).

5. Zhu, H. & Zhou, X. Statistical methods for SNP heritability estimation and partition: A review. *Comput. Struct. Biotechnol. J.* **18**, 1557–1568 (2020).
6. Liu, J. *et al.* Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* **14**, 229 (2013).
7. Yan, J. *et al.* Genome-wide detection of autosomal copy number variants in several sheep breeds using Illumina OvineSNP50 BeadChips. *Small Rumin. Res.* **155**, 24–32 (2017).
8. Huang, Y. *et al.* An atlas of CNV maps in cattle, goat and sheep. *Sci. China Life Sci.* **64**, 1747–1764 (2021).
9. Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
10. Bridges, C. B. The bar “gene” a duplication. *Science* **83**, 210–211 (1936).
11. Bae, J. S. *et al.* Identification of copy number variations and common deletion polymorphisms in cattle. *BMC Genomics* **11**, 1–10 (2010).
12. Butty, A. M. *et al.* Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle. *J. Dairy Sci.* **104**, 8050–8061 (2021).
13. Alvarez, C. E. & Akey, J. M. Copy number variation in the domestic dog. *Mamm. Genome* **23**, 144–163 (2012).
14. Wang, X., Nahashon, S., Feaster, T. K., Bohannon-Stewart, A. & Adefope, N. An initial map of chromosomal segmental copy number variations in the chicken. *BMC Genomics* **11**, 1–10 (2010).
15. Fernandes, A. C. *et al.* Genome-wide detection of CNVs and their association with performance traits in broilers. *BMC Genomics* **22**, 1–18 (2021).
16. Fadista, J., Nygaard, M., Holm, L.-E., Thomsen, B. & Bendixen, C. A snapshot of CNVs in the pig genome. *PLoS One* **3**, e3916 (2008).
17. Ramayo-Caldas, Y. *et al.* Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics* **11**, 1–10 (2010).
18. Fontanesi, L. *et al.* An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics* **11**, 1–15 (2010).
19. Liu, M. *et al.* Genome-wide CNV analysis revealed variants associated with growth traits in African indigenous goats. *Genomics* **112**, 1477–1480 (2020).
20. Fontanesi, L. *et al.* Exploring copy number variation in the rabbit (*Oryctolagus cuniculus*) genome by array comparative genome hybridization. *Genomics* **100**, 245–251 (2012).
21. Fontanesi, L. *et al.* A first comparative map of copy number variations in the sheep genome. *Genomics* **97**, 158–165 (2011).
22. Norris, B. J. & Whan, V. A. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res.* **18**, 1282–1293 (2008).
23. Ma, Q. *et al.* Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600 K SNP array. *Sci. Rep.* **7**, 912 (2017).
24. Karimi, M. O. Investigation of genetic diversity and selection markers of genome level in some Afghani sheep breeds. PhD thesis, Ferdowsi University of Mashhad (2016).
25. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
26. Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G. & McEwan, J. C. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet.* **13**, 10 (2012).
27. Curik, I., Ferenčaković, M., Karapandza, N., Cubric-Curik, V. & Sölkner, J. Estimation of inbreeding and effective population size in Istrian cattle using molecular information. *Acta Agraria Kaposváriensis* **18**, 30–34 (2014).
28. Wang, K. *et al.* PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
29. Peiffer, D. A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* **16**, 1136–1148 (2006).
30. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
31. Gholizadeh, M., Rahimi-Mianji, G., Nejati-Javaremi, A., De Koning, D. J. & Jonas, E. Genomewide association study to detect QTL for twinning rate in Baluchi sheep. *J. Genet.* **93**, 489–493 (2014).
32. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
33. Wang, X. *et al.* CNVcaller: Highly efficient and widely applicable software for detecting copy number variations in large populations. *GigaScience* **6**, gix115 (2017).
34. Hou, C.-L. *et al.* Genome-wide analysis of copy number variations in Chinese sheep using array comparative genomic hybridization. *Small Rumin. Res.* **128**, 19–26 (2015).
35. Mohanty, T. R. *et al.* Molecular variation in pigmentation genes contributing to coat colour in native Korean Hanwoo cattle. *Anim. Genet.* **39**, 550–553 (2008).
36. Weedon, M. N. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nat. Genet.* **40**, 575–583 (2008).
37. Metzger, J. *et al.* Analysis of copy number variants by three detection algorithms and their association with body size in horses. *BMC Genomics* **14**, 1–15 (2013).
38. Karimi, K., Esmailzadeh, A., Wu, D. D. & Gondro, C. Mapping of genome-wide copy number variations in the Iranian indigenous cattle using a dense SNP data set. *Anim. Prod. Sci.* **58**, 1192–1200 (2017).
39. Nosrati, M. & Tahmoursipour, M. Identification of variation in genome copy number of 15 Italian sheep breeds using a 50 kb sheep chip. *Iran. J. Anim. Sci. Res.* **8**, 489–501 (2016).
40. Pielberg, G. R. *et al.* A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nat. Genet.* **40**, 1004–1009 (2008).
41. Di Gerlando, R. *et al.* Identification of copy number variations and genetic diversity in Italian insular sheep breeds. *Animals* **12**, 217 (2022).
42. Ma, Y., Zhang, Q., Lu, Z., Zhao, X. & Zhang, Y. Analysis of copy number variations by SNP50 BeadChip array in Chinese sheep. *Genomics* **106**, 295–300 (2015).
43. Jenkins, G. M. *et al.* Copy number variants in the sheep genome detected using multiple approaches. *BMC Genomics* **17**, 441 (2016).
44. Wang, J. *et al.* Enhancing genome-wide copy number variation identification by high density array CGH using diverse resources of pig breeds. *PLoS One* **9**, e87571 (2014).
45. Letaief, R. *et al.* Identification of copy number variation in French dairy and beef breeds using next-generation sequencing. *Genet. Sel. Evol.* **49**, 1–15 (2017).
46. Liu, M. *et al.* Diversity of copy number variation in the worldwide goat population. *Heredity* **122**, 636–646 (2019).
47. Marenne, G. *et al.* Assessment of copy number variation using the Illumina Infinium 1M SNP-array: A comparison of methodological approaches in the Spanish Bladder Cancer/EPICURO study. *Hum. Mutat.* **32**, 240–248 (2011).
48. Zhu, C. *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.* **6**, 1–9 (2016).

49. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).
50. Yuan, C. *et al.* A global analysis of CNVs in Chinese indigenous fine-wool sheep populations using whole-genome resequencing. *BMC Genomics* **22**, 78 (2021).
51. Goyache, F. *et al.* Functional characterization of copy number variations regions in Djallonké sheep. *J. Anim. Breed. Genet.* **138**, 600–612 (2021).
52. Di Gerlando, R. *et al.* Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PLoS One* **14**, e0215204 (2019).
53. Yang, L. *et al.* Diversity of copy number variation in a worldwide population of sheep. *Genomics* **110**, 143–148 (2018).
54. Prinsen, R. *et al.* A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. *Livest. Sci.* **202**, 7–12 (2017).
55. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015).
56. Usman, T. *et al.* Novel SNPs in IL-17F and IL-17A genes associated with somatic cell count in Chinese Holstein and Inner-Mongolia Sanhe cattle. *J. Anim. Sci. Biotechnol.* **8**, 5 (2017).
57. Zhu, C., Li, M., Qin, S., Zhao, F. & Fang, S. Detection of copy number variation and selection signatures on the X chromosome in Chinese indigenous sheep with different types of tail. *Asian Australas. J. Anim. Sci.* **33**, 1378 (2020).
58. Dunner, S. *et al.* Genes involved in muscle lipid composition in 15 European *Bos taurus* breeds. *Anim. Genet.* **44**, 493–501 (2013).
59. da Silva, V. H. *et al.* Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. *PLoS One* **11**, e0157711 (2016).
60. Taye, M. *et al.* Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. *BMC Genet.* **18**, 1–14 (2017).
61. Wei, C. *et al.* Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* **16**, 1–12 (2015).
62. Meuwissen, R. L. *et al.* A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes. *EMBO J.* **11**, 5091–5100 (1992).
63. Forde, N. *et al.* Effects of low progesterone on the endometrial transcriptome in cattle. *Biol. Reprod.* **87**, 121–124 (2012).
64. Levent, A. & Büyükaşar, K. Expression of Rho-kinase (ROCK-1 and ROCK-2) and its substantial role in the contractile activity of the sheep ureter. *Br. J. Pharmacol.* **143**, 431–437 (2004).
65. Pispá, J., Mikkola, M. L., Mustonen, T. & Thesleff, I. Ectodysplasin, Edar and TNFRSF19 are expressed in complementary and overlapping patterns during mouse embryogenesis. *Gene Expr. Patterns* **3**, 675–679 (2003).
66. Li, R., Tang, X., Miao, S., Zong, S. & Wang, L. Regulation of the G2/M phase of the cell cycle by sperm associated antigen 8 (SPAG8) protein. *Cell Biochem. Funct. Cell. Biochem. Modul. Act. Agents Dis.* **27**, 264–268 (2009).
67. Zheng, Y. H. *et al.* Expression pattern of meiosis associated SYCP family members during germline development in chickens. *Reproduction* **138**, 483 (2009).
68. Arangasamy, A., Kasimanickam, V. R., DeJarnette, J. M. & Kasimanickam, R. K. Association of CRISP2, CCT8, PEBP1 mRNA abundance in sperm and sire conception rate in Holstein bulls. *Theriogenology* **76**, 570–577 (2011).
69. Wang, Z. *et al.* Genome-wide detection of CNVs and association with body weight in sheep based on 600K SNP arrays. *Front. Genet.* **11**, 558 (2020).
70. Cao, X.-K. *et al.* Integrating CNVs into meta-QTL identified GBP4 as positional candidate for adult cattle stature. *Funct. Integr. Genomics* **18**, 559–567 (2018).
71. Hu, L. *et al.* Genome-wide analysis of CNVs in three populations of Tibetan sheep using whole-genome resequencing. *Mapp. Intim.* <https://doi.org/10.21203/rs.3.rs-1145952/v1> (2021).
72. Raza, M. A., Younas, M. & Schlecht, E. Prevalence of gastrointestinal helminths in pastoral sheep and goat flocks in the Cholistan Desert of Pakistan. *J. Anim. Plant Sci.* **24**, 127–134 (2014).
73. Li, C. *et al.* RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. *Sci. Rep.* **6**, 1–11 (2016).
74. Zhang, L. *et al.* Genome-wide association studies for growth and meat production traits in sheep. *PLoS One* **8**, e66569 (2013).
75. Liu, L., Liu, H., Ning, L. & Li, F. Rabbit SLC15A1, SLC7A1 and SLC1A1 genes are affected by site of digestion, stage of development and dietary protein content. *Animal* **13**, 326–332 (2019).
76. Osawa, Y. *et al.* EXOC1 plays an integral role in spermatogonia pseudopod elongation and spermatocyte stable syncytium formation in mice. *Elife* **10**, e59759 (2021).
77. Rui, L. I. U. *et al.* Fine mapping QTLs affecting milk production traits on BTA6 in Chinese Holstein with SNP markers. *J. Integr. Agric.* **12**, 110–117 (2013).
78. Ramayo-Caldas, Y. *et al.* From SNP co-association to RNA co-expression: Novel insights into gene networks for intramuscular fatty acid composition in porcine. *BMC Genomics* **15**, 1–15 (2014).
79. Serrano, M. *et al.* Genome-wide association studies for sperm traits in Assaf sheep breed. *Animal* **15**, 100065 (2021).
80. Marroquí, L. *et al.* BACH2, a candidate risk gene for type 1 diabetes, regulates apoptosis in pancreatic β -cells via JNK1 modulation and crosstalk with the candidate gene PTPN2. *Diabetes* **63**, 2516–2527 (2014).
81. Li, Y., Liu, R., Wu, J. & Li, X. Self-eating: Friend or foe? The emerging role of autophagy in fibrotic diseases. *Theranostics* **10**, 7993 (2020).
82. Cavani, L. *et al.* Genomic study of *Babesia bovis* infection level and its association with tick count in Hereford and Braford cattle. *Front. Immunol.* **11**, 1905 (2020).
83. Garza-Brenner, E. *et al.* Influence of genetic markers on the feeding behavior of yearling bulls. *Revista Colombiana de Ciencias Pecuarias* **32**, 14–20 (2019).
84. Huang, C. *et al.* Copy number variation of the HPGDS gene in the Ashidan yak and its associations with growth traits. *Gene* **772**, 145382 (2021).
85. Hsu, H.-P., Wang, C.-Y., Hsieh, P.-Y., Fang, J.-H. & Chen, Y.-L. Knockdown of serine/threonine-protein kinase 24 promotes tumorigenesis and myeloid-derived suppressor cell expansion in an orthotopic immunocompetent gastric cancer animal model. *J. Cancer* **11**, 213 (2020).
86. Neumann, A. *et al.* Computational investigations on the binding mode of ligands for the cannabinoid-activated G protein-coupled receptor GPR18. *Biomolecules* **10**, 686 (2020).
87. Gross, N., Peñagaricano, F. & Khatib, H. Integration of whole-genome DNA methylation data with RNA sequencing data to identify markers for bull fertility. *Anim. Genet.* **51**, 502–510 (2020).
88. Abied, A. *et al.* Genomic diversity, population structure, and signature of selection in five Chinese native sheep breeds adapted to extreme environments. *Genes* **11**, 494 (2020).
89. Jiménez-Martínez, M., Stamatakis, K. & Fresno, M. The dual-specificity phosphatase 10 (DUSP10): Its role in cancer, inflammation, and immunity. *Int. J. Mol. Sci.* **20**, 1626 (2019).
90. Liebl, J. *et al.* Cdk5 controls lymphatic vessel development and function by phosphorylation of Foxc2. *Nat. Commun.* **6**, 1–13 (2015).
91. Li, H., Li, M., Tang, C. & Xu, L. Screening and prognostic value of potential biomarkers for ovarian cancer. *Ann. Transl. Med.* **9**, 1007 (2021).

92. Karimi, K., Farid, A. H., Myles, S. & Miar, Y. Detection of selection signatures for response to Aleutian mink disease virus infection in American mink. *Sci. Rep.* **11**, 1–13 (2021).
93. Napierala, J. S. *et al.* Comprehensive analysis of gene expression patterns in Friedreich's ataxia fibroblasts by RNA sequencing reveals altered levels of protein synthesis factors and solute carriers. *Dis. Model. Mech.* **10**, 1353–1369 (2017).
94. Mohammadi, H., Rafat, S. A., Moradi Shahrabak, H., Shodja, J. & Moradi, M. H. Genome-wide association study and gene ontology for growth and wool characteristics in Zandi sheep. *J. Livest. Sci. Technol.* **8**, 45–55 (2020).
95. Amaral, A. J., Pavão, A. L. & Gama, L. T. Genomic tools for the conservation and genetic improvement of a highly fragmented breed—The Ramo Grande cattle from the Azores. *Animals* **10**, 1089 (2020).
96. Bloch, M. *et al.* KCNMA1 gene amplification promotes tumor cell proliferation in human prostate cancer. *Oncogene* **26**, 2525–2534 (2007).
97. Aliloo, H., Mrode, R., Okeyo, A. M. & Gibson, J. P. Ancestral haplotype mapping for GWAS and detection of signatures of selection in admixed dairy cattle of Kenya. *Front. Genet.* **11**, 544 (2020).
98. Mehla, K. *et al.* Genome-wide analysis of the heat stress response in Zebu (Sahiwal) cattle. *Gene* **533**, 500–507 (2014).
99. Upadhyay, M. *et al.* Distribution and functionality of copy number variation across European cattle populations. *Front. Genet.* **8**, 108 (2017).
100. Gorla, E. *et al.* Genomic variability in Mexican chicken population using copy number variants. *BMC Genet.* **18**, 1–11 (2017).
101. Igoshin, A. V. *et al.* Copy number variants in genomes of local sheep breeds from Russia. *Anim. Genet.* **53**, 119–132 (2021).

Acknowledgements

The authors gratefully acknowledge Dr. Nicolo Macciotta and Dr. Giustino Gaspa from Sassari University, Italy for providing the facilities to conduct the experimental work and genotyping the animal samples.

Author contributions

The data were collected by M.H.M. and M.O.K., and analyzed by M.H.M., R.M., and A.H.F. The paper was drafted by M.H.M. and R.M. All authors have contributed to the editing of the article, and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-18571-4>.

Correspondence and requests for materials should be addressed to M.H.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022