



# Integrating Genome-Wide CNVs Into QTLs and High Confidence GWAScore Regions Identified Positional Candidates for Sheep Economic Traits

Jie Cheng<sup>1†</sup>, Xiukai Cao<sup>1†</sup>, Quratulain Hanif<sup>1,2,3</sup>, Li Pi<sup>4</sup>, Linyong Hu<sup>4</sup>, Yongzhen Huang<sup>1</sup>, Xianyong Lan<sup>1</sup>, Chuzhao Lei<sup>1</sup> and Hong Chen<sup>1\*</sup>

<sup>1</sup> Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China, <sup>2</sup> Computational Biology Lab, National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan, <sup>3</sup> Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan, <sup>4</sup> Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, China

## OPEN ACCESS

### Edited by:

Fabyano Fonseca Silva,  
Universidade Federal de Viçosa, Brazil

### Reviewed by:

Lin Jiang,  
Chinese Academy of Agricultural  
Sciences, China  
José Lindenberg Sarmento,  
Federal University of Piauí, Brazil

### \*Correspondence:

Hong Chen  
chenhong1212@263.net

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 16 January 2020

Accepted: 11 May 2020

Published: 17 June 2020

### Citation:

Cheng J, Cao X, Hanif Q, Pi L,  
Hu L, Huang Y, Lan X, Lei C and  
Chen H (2020) Integrating  
Genome-Wide CNVs Into QTLs  
and High Confidence GWAScore  
Regions Identified Positional  
Candidates for Sheep Economic  
Traits. *Front. Genet.* 11:569.  
doi: 10.3389/fgene.2020.00569

Copy number variations (CNVs) are important source of genetic variation, which can affect diverse economic traits through a variety of mechanisms. In addition, genome scan can identify many quantitative trait loci (QTLs) for the economic traits, while genome-wide association studies (GWAS) can localize genetic variants associated with the phenotypic variations. Here, we developed a method called GWAScore which collected GWAS summary data to identify potential candidates, and integrated CNVs into QTLs and high confidence GWAScore regions to detect crucial CNV markers for sheep growth traits. We got 197 candidate genes which were overlapping with the candidate CNVs. Some crucial genes (*MYLK3*, *TTC29*, *HERC6*, *ABCG2*, *RUNX1*, etc.) showed significantly elevated GWAScore peaks than other candidate genes. In this study, we developed the GWAScore method to excavate the potential value of candidate genes as markers for the sheep molecular breeding.

**Keywords:** copy number variations, QTL, GWAS, GWAScore, sheep economic traits

## INTRODUCTION

Chaka sheep also named Qinghai plateau half-fine wool sheep has been a geographical symbol of agricultural products by Ministry of Agriculture. Chaka sheep is famous for its excellent mutton quality, hence also known as the tributary sheep in the ancient times. Besides, its wool is exceptional for making woolen sweaters. Thus, Chaka sheep is reared as a multipurpose breed. Moreover,

**Abbreviations:** *ABCG2*, ATP binding cassette subfamily G member 2; aCGH, array comparative genomic hybridization; *ADGRB1*, adhesion G protein-coupled receptor B1; *AGPAT3*, 1-acylglycerol-3-phosphate O-acyltransferase 3; CNVs, copy number variations; CNVR, copy number variation region; *DLGAP1*, DLG associated protein 1; GWAS, genome-wide association studies; *HERC6*, HECT and RLD domain containing E3 ubiquitin protein ligase family member 6; Indels, insertions/deletions; *MYLK3*, myosin light chain kinase 3; *MYL*, myosin light chain; NGS, next-generation sequencing; QTL, quantitative trait loci; *RUNX1*, Runt-related transcription factor 1; SNPs, single nucleotide polymorphisms; *TTC29*, tetratricopeptide repeat domain 29.

Hu sheep is one of the major breeds for lamb skin in China, whereas, the Small Tailed Han Sheep (STHS) depict high growth speed with good stress resistance.

Extensive studies have been conducted to associate phenotypes with two forms of genetic variations including single nucleotide polymorphisms (SNPs) and insertion/deletions (Indels). In recent years, copy number variation (CNVs) has emerged as a potential source of genetic divergence, bearing significant phenotypic and economically important traits, however, CNVs have not been thoroughly characterized so far. CNV is defined as a DNA segment of ~50 bp or larger in size and present submicroscopic copy number variation in comparison with a reference genome, including deletions, insertions, duplications, and complex multi-site variants (Redon et al., 2006; Mills et al., 2011). Presently, in order to detect the CNVs in a genome, various genome-wide platforms have been developed, including SNP genotyping platforms (Di Gerlando et al., 2018), array comparative genomic hybridization (aCGH; Zhang et al., 2014), and next-generation sequencing (NGS) (Xu et al., 2017). For sheep, SNP arrays and CGH arrays are routinely used, and their performances in CNV detection have been reviewed in sheep in various studies (Fontanesi et al., 2011; Liu et al., 2013). These methods have moderated power in detecting duplication regions, which have been proven to be affected by cross-hybridization of repetitive sequence and low probe density (Bickhart et al., 2012). NGS is well accepted in the genome-wide identification of CNV.

Quantitative trait loci (QTLs) is genetic basis of economic trait, such as muscle weight in carcass, body weight, milk protein percentage. However, such traits are mostly based on multiple genomic variants. Since the genes are pleiotrophic, multi-traits GWAScore can improve the confidence of crucial gene variants (Flint et al., 2014). Candidate gene method is one of the traditional methods to identify the important gene. It is easy to use, and the effect of candidate gene on livestock phenotypes can be inferred based on the known-function genes in the model animal (Cheng et al., 2019). However, for genes with unknown functionality, this method could not possess reliable results. At present, 3D genomics is a powerful method to identify crucial CNV impacting phenotypes (Lupiañez et al., 2015; Will et al., 2017; Zhang et al., 2018). However, the costs of NGS and the downstream analysis are significantly high. Our method identifies CNVs using resequencing data which then integrates with the existing genome-wide association studies (GWAS) and QTL data. QTL are regions often associated with one or more important traits conferring phenotypes. GWAS can identify significant genetic variants associated with one or more interest phenotypes. GWAScore could quantitate the probability distribution of key variants in genome regions.

Single nucleotide polymorphisms have been extensively used for GWAS data, but integrating GWAS data based on CNVs for sheep has been rarely reported. Here, we used NGS to detect the CNVs in sheep. As a result, we have developed a new method called GWAScore to look for important regions across the genome by summarizing the GWAS data, integrating QTLs and high confidence GWAScore regions with CNVs to

identify positional candidates for sheep breeding to accelerate genetic improvement.

## MATERIALS AND METHODS

### Ethics Statement

The protocols were approved by the Faculty of Animal Policy and Welfare Committee of Northwest A&F University (FAPWC-NWAFU, protocol number, NWAFAC1008).

### Sample Collection and Sequencing

Chaka, Hu, and STHS are three distinct breeds from China, which were collected from Qinghai, Jiangsu, and Shandong provinces, respectively (detail in **Supplementary Table S1**). DNA were extracted from whole blood samples of Chaka sheep ( $n = 10$ ) as previously described (Pan et al., 2018). The libraries were subjected to NGS with the HiSeq 4000 platform with paired-end mode. Trimmomatic (Leading:20 Trailing:20 Slidingwindow:3-15 Avgqual:20 Minlen:35 Tophred:33) and FASTQC software were employed to trim the sequences to assess the quality of the raw sequence data. Sequence reads were mapped, sorted and deduped to the reference sheep genome<sup>1</sup> by BWA-MEM (0.7.13-r1126) and Picard v2.18.2<sup>2</sup>. The command of BWA-MEM “mem -t 12 -M -R” was used for each individual read. The PICARD SortSam and MarkDuplicates were also used to sort and dedupe. The average sequencing depth was  $7.68\times$  for the 10 Chaka sheep. Raw sequencing data of Hu and STHS have been submitted to the NCBI Sequence Read Archive with accession number SRP066883 (Pan et al., 2018). Sequencing data of Ten Chaka sheep have been submitted to NCBI (project number PRJNA594811).

### Whole-Genome Copy Number Variation Identification

To assess the copy number variations of sheep including Chaka ( $n = 10$ ), Hu ( $n = 10$ ) and Small Tailed Han ( $n = 9$ ) Sheep, we first genotyped the 29 individuals that represent three important Chinese sheep breeds. The CNVs were identified using CNVcaller methods utilizing NGS data (Wang et al., 2017). The CNVs were calculated in 1000 bp sliding window with 500 bp steps along the autosomes. Next, the absolute number of copies in each window were calculated and reads with high similarity ( $\geq 97\%$ ) were merged to determine the copy number variation regions. Primary CNVR were further processed into merged CNVR for genotyping of the data set. We combined the identified CNVs among all individuals.

### Integrating CNV with GWAS and QTL Data

To detect significant loci with phenotypic effects, we integrated GWAS and QTL aggregate with CNV. GWAS and QTL data

<sup>1</sup>[https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000298735.2/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2/), Oar\_v4.0

<sup>2</sup><http://broadinstitute.github.io/picard/>

about growth traits, were retrieved from Animal QTLdb<sup>3</sup>. All data were mapped to sheep genome Oar\_v4.0 and visualized using CIRCOS.

Array density, phenotypic measurement errors, and statistical methods lead to poor reproducibility of GWAS (Barendse, 2011; Schaid et al., 2018; Wijmenga and Zhernakova, 2018). In our study, GWAScore was developed to study genomic regions which impact the phenotypic effects using GWAS data. The method excels these factors by concisely summarizing GWAS SNPs of different studies and QTLs. Here, GWAScore was calculated with 100 kb window (Cao et al., 2019) as

$$\text{GWAScore}_{(x,x+0.1)} = \frac{-n_{\text{gwas}(x,x+0.1)} \times \sum_{i=1}^{n_{\text{snp}(x,x+0.1)}} \log P}{0.1 \times n_{\text{snp}(x,x+0.1)}}$$

$n_{\text{gwas}(x,x+0.1)}$  is the number of GWAS in the genomic region ( $x, x + 0.1$ ),  $n_{\text{snp}(x,x+0.1)}$  is the number of SNPs in genomic region ( $x, x + 0.1$ ),  $P$  is the  $P$  value of significant SNPs. Whereas, ( $x, x + 0.1$ ) means the positions of each windows, and the size of window was 0.1 Mb. GWAScore derived from the significance SNP loci was used to quantify the probability of key mutations in genome (GWAScore > 0).

## Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Pathway Analysis

It was in QTLs and high confidence GWAScore regions that identified genome-wide CNVs and then annotated positional candidate gene for sheep growth traits. Entering that, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were completed ( $P < 0.01$ ) by metascap online software<sup>4</sup> (Zhou et al., 2019).

<sup>3</sup><http://www.animalgenome.org/>

<sup>4</sup><http://metascap.org/gp/index.html#/main/step1>

## RESULTS

### The Analysis of CNVs

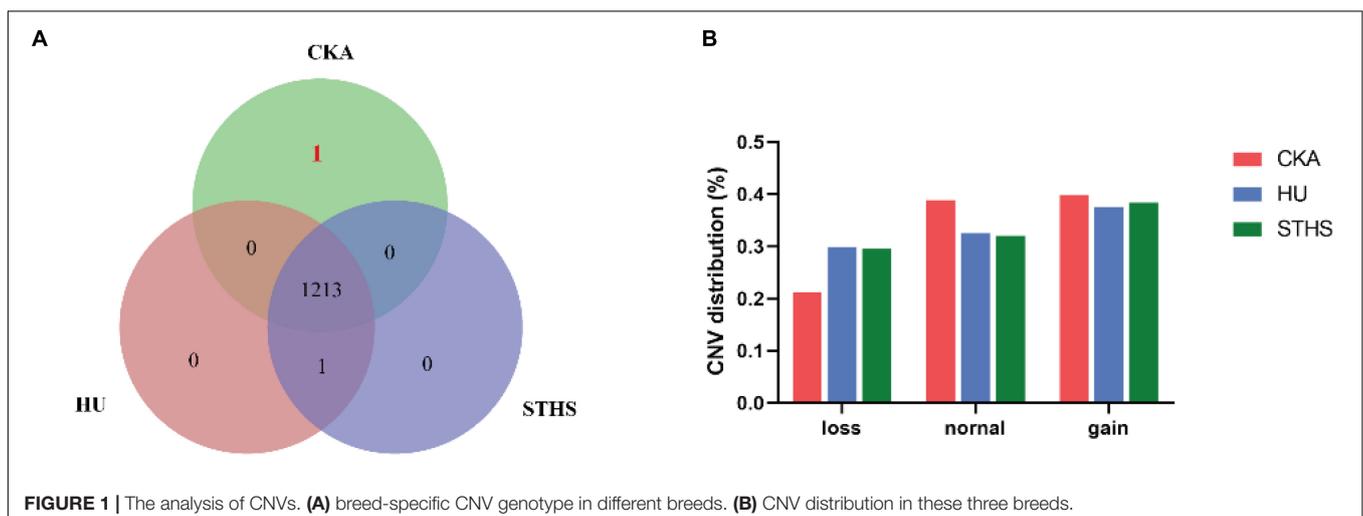
Copy number variations (CNVs) are an important source of genetic variations. We obtained 4301 CNV regions in total by using the whole genome sequencing data from the three breeds. The average length of the CNVs was about 4 kb. Among them, we identified 19%, 18% and 56% CNVs located in exonic, intronic and intergenic regions, respectively (detail in **Supplementary Table S2**). Furthermore, 4301 CNVs were commonly shared by all the breeds in the current study.

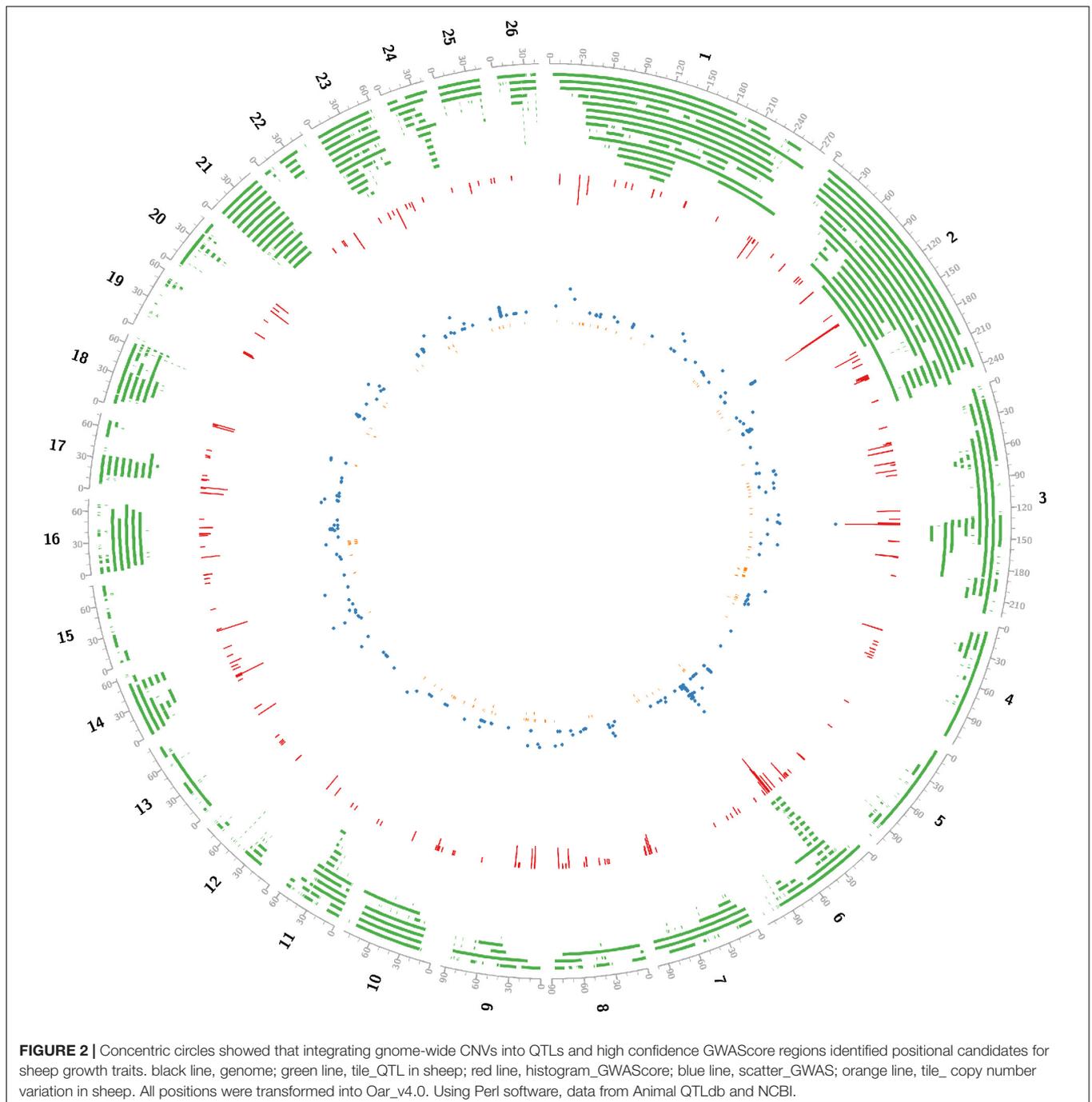
CNV caller was used to genotype CNVs into “dd, Ad, AA, AB, BB, BC, and M.” Moreover, “dd and Ad” were grouped into “Loss genotype,” “AA” into “Normal genotype” and “AB, BB, BC, and M” into “Gain genotype.” If all individual of one breed have the same CNV genotypes which didn’t appear in other two breeds, we will take it as a breed-specific CNV. Then, we found the number of breed-specific CNV genotype was very low and only one specific CNV genotype for Chaka sheep. The specific and shared CNV genotypes are represented in a venn diagram for all the three breeds (**Figure 1A**). The only one breed-specific CNV in Chaka sheep was found at chr11: 27042001-27044500. CNVs are distributed into loss, normal and gain mutations among the three breeds and are presented in **Figure 1B**.

### Integrative Analysis Annotates Candidate Genes of Sheep Economic Traits

A total of 4301 CNV regions, 729 carcass QTLs (detail in **Supplementary Table S3**) and 440 GWAS markers were collected. GWAScore was designed to use independent GWAS to calculate the phenotypic effects of a given genomic region. The GWAScore map of 190 regions ranging from 1.30 to 16.55, covering 19.0 Mb of the genome was presented (**Supplementary Table S4**).

Critical CNV markers were revealed by integrative analysis of 729 carcass QTL and 440 GWAS markers as shown in





**TABLE 1 |** Integrative analysis annotated candidate genes (top five) of sheep traits.

Chromosome	Start	End	GWAScore	The number of QTL	Overlap genes	Location
chr14	14618001	14622000	7.815613972	4	<i>MYLK3</i>	intronic
chr17	11330501	11334500	6.915066425	1	<i>TTC29</i>	intronic
chr6	36222001	36224500	6.882224908	17	<i>HERC6</i>	exonic
chr6	36353001	36356000	6.882224908	17	<i>ABCG2</i>	UTR5
chr1	264845001	264848500	6.649751982	2	<i>RUNX1</i>	intronic

**Figure 2**, in which we identified 197 coding candidate genes located in common regions shared by GWAScore regions, QTL and CNVs regions. Some crucial candidate genes were identified with higher GWAScore such as myosin light chain kinase 3 (*MYLK3*), tetratricopeptide repeat domain 29 (*TTC29*), HECT and RLD domain containing E3 ubiquitin protein ligase family member 6 (*HERC6*), ATP binding cassette subfamily G member 2 (*ABCG2*) and Runt-related transcription factor 01 (*RUNX1*), which may be associated with important economic traits. Among the 197 candidate genes, we identified 23 CNVs located in the exon of gene (*HERC6*, etc.) and five CNVs located in the upstream or 5'UTR of gene (*ABCG2*, etc.) (**Table 1**, detail in **Supplementary Table S5**).

## GO and KEGG Analysis

We analyzed 190 regions with GWAScores, and identified important CNVs to annotate 197 common genes. We dissected these 197 genes by GO and the KEGG Pathway Analysis (**Figure 3**). The genes enriched axonogenesis (GO:0007409), factors involved in megakaryocyte development and platelet production (R-HSA-983231), negative regulation of CD4-positive, alpha-beta T cell differentiation (GO:0043371), and muscle cell migration (GO:0014812).

## DISCUSSION

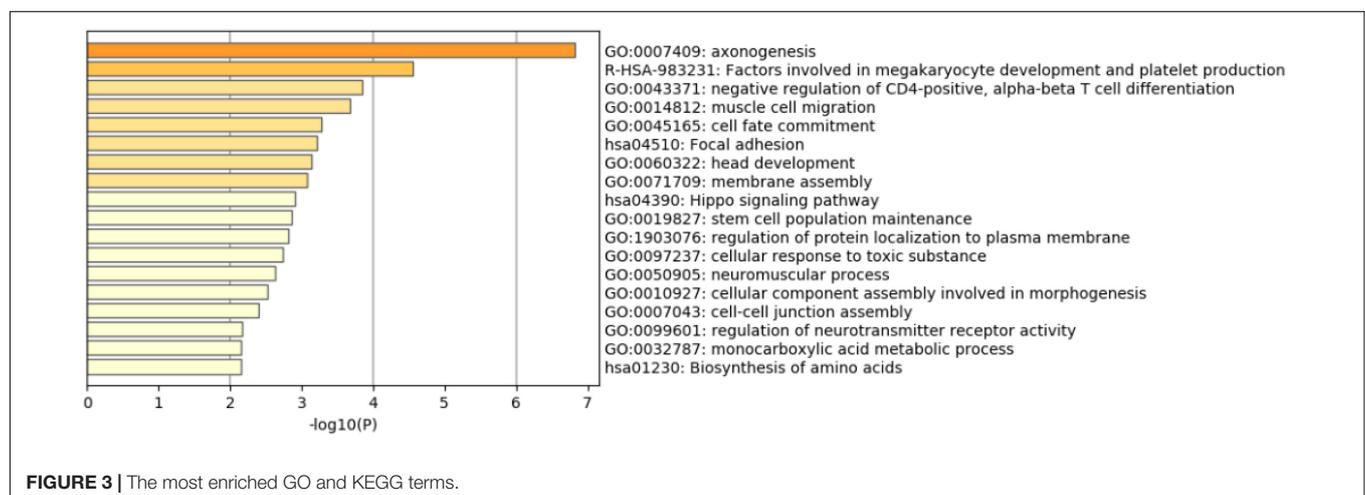
Researchers have extensively utilized SNPs as genetic markers for GWAS, but the identification of CNVs has gathered the highlight for its associated with complex traits. Genome-wide association studies based on CNV markers have been reported in livestock, such as milk production traits, milk somatic cell score in Holsteins (Xu et al., 2014; Ben Sassi et al., 2016; Duran Aguilar et al., 2017), feed conversion ratio in beef cattle (de Almeida Santana et al., 2016), growth traits in *Bos indicus* (Zhou et al., 2016), quantitative traits in Brown Swiss cattle (Prinsen et al., 2017), and umbilical hernia in swine (Long et al., 2016). Although the well-powered GWAS provided by the integrated array of CNVs may be a perfect solution in

livestock breeding, it didn't integrate all the GWAS data from different papers and besides it is still not financially viable to serve the majority of the scientific community under a sea of samples (McCarroll and Altshuler, 2007). Integrating GWAS data based on CNVs for sheep are rarely reported. Cao et al. (2018) have integrated cattle genome CNV with various meta-QTL and identified CNV of *GBP4* as a positional candidate for cattle stature. But GWAS based QTL were omitted in their study which may provide additional information for phenotypic CNV identification (Cao et al., 2018).

Candidate gene method is one of the traditional methods to identify the important gene. The effect of candidate gene on livestock phenotypes can be inferred based on the known-function genes in the model animal (Cheng et al., 2019). For unknown-function genes, we can't use candidate gene method. In this study, we used a novel method, developed by our group called GWAScore which could quantify the probability to be a QTL of a given region by GWAS data to reveal novel candidates. This method has helped to detect key genes associated with muscle development from transcriptome (Cao et al., 2019). QTLs, regions in the genome, contain genetic variants affecting the quantitative traits of interest. Genome-wide association studies is identifying genetic variants associated with one or more interest phenotypes (Georges et al., 2019). In our study, by using public data of GWAS from Animal QTLdb<sup>3</sup>, we integrated whole-genome CNVs into QTLs and high confidence GWAScore regions for whole-genome analysis.

These 197 candidate genes were analyzed by the GO Analysis ( $P < 0.01$ ) and the KEGG Pathway. These genes were enriched in axonogenesis (GO:0007409), factors involved in megakaryocyte development and platelet production (R-HSA-983231), negative regulation of CD4-positive, alpha-beta T cell differentiation (GO:0043371), and muscle cell migration (GO:0014812), which showed that the function of these gene were associated with neural development, immunity regulation (Cloutier et al., 2018) and muscle migration, imposing strong signals toward the economic traits.

We extracted crucial genes with highest GWAScore and overlapped CNVs in these regions (**Table 1**, detail in



Supplementary Table S5), including the well-known gene *MYLK3*, *TTC29*, *HERC6*, *ABCG2*, *RUNX1*, etc. *MYLK3* is the member of the myosin light chain (MYL) family. Importantly, myosin is the main component of the myofibrillary filament, which plays a crucial role in the growth and contraction of muscle (Zhang et al., 2009). *TTC29* is associated with muscle mass in informative regions of the QTL (Kärst et al., 2011), and *TTC29* has also been recognized as candidate genes for residual feed intake in pigs (Do et al., 2014). *HERC6* has significant effects on fat and protein yield (Cohen-Zinder et al., 2005). There is significant association between *ABCG2* polymorphisms and milk production trait in Holsteins, Chinese Holsteins, Brown Swiss cows, and Sheep (Cohen-Zinder et al., 2005; Yue et al., 2011; Árnýasi et al., 2013; Cecchinato et al., 2014). It was reported that *RUNX1* plays a vital role in neural development, hematopoiesis, and leukemogenesis (Speck and Gilliland, 2002; Yoshikawa et al., 2007). The GWAS hit SNP within *RUNX1* is associated with the mean corpuscular volume level in pigs. We identified 23 CNVs located in the exon of genes, such as *DLGAP1*, *ADGRB1*, and *AGPAT3*. DLG associated protein 1 (DLGAP1) which may regulate DLG1, is associated with control of infection by ovine lentivirus (White et al., 2012). Adhesion G protein-coupled receptor B1 (*ADGRB1*) gene has antagonistic or sensitive effects on *Escherichia coli* F17 resistance to diarrhea (Jin et al., 2018). 1-Acylglycerol-3-phosphate O-acyltransferase 3 (*AGPAT3*) gene may make triacylglycerol and phospholipid synthesis in milk biosynthesis (Bionaz and Loor, 2008).

## CONCLUSION

We used a new method called GWAScore and integrated whole-genome CNVs into QTLs and high confidence GWAScore regions, which identified potential candidate for the first time in sheep including the well-known gene *MYLK3*, *TTC29*, *HERC6*, *ABCG2*, *RUNX1*, etc.

## DATA AVAILABILITY STATEMENT

Raw sequencing data of Hu and Small Tailed Han Sheep have been submitted to the NCBI Sequence Read

Archive under accession number SRP066883. Sequencing data of Ten Chaka sheep have been submitted to NCBI under submission number SUB6671624 (project number PRJNA594811).

## ETHICS STATEMENT

The animal study was reviewed and approved by the Faculty of Animal Policy and Welfare Committee of Northwest A&F University (FAPWC-NWAFU). Written informed consent was obtained from the owners for the participation of their animals in this study.

## AUTHOR CONTRIBUTIONS

JC and XC designed the project. JC wrote the manuscript. QH and XC revised the manuscript. LP and LH provided the sample of Chaka sheep. YH, XL, CL, QH, and HC provided the correction of the manuscript.

## FUNDING

This work was supported by the Qinghai Province major science and technology projects (2017-NK-153).

## ACKNOWLEDGMENTS

Thanks for Laboratory for Agricultural Genomic Big Data in College of Animal Science and Technology, Northwest A&F University.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00569/full#supplementary-material>

## REFERENCES

- Árnýasi, M., Komlósi, I., Kent, M., Czeglédi, L., Gulyás, G., and Jávora, A. (2013). Investigation of polymorphisms and association of the *ABCG2* gene with milk production traits in sheep. *Livestock Sci.* 154, 64–68. doi: 10.1016/j.livsci.2013.03.010
- Barendse, W. (2011). The effect of measurement error of phenotypes on genome wide association studies. *BMC Genomics* 12:232. doi: 10.1186/1471-2164-12-232
- Ben Sassi, N., Gonzalez-Recio, O., De Paz-Del Rio, R., Rodriguez-Ramilo, S. T., and Fernandez, A. I. (2016). Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. *J. Dairy Sci.* 99, 6371–6380. doi: 10.3168/jds.2015-10487
- Bickhart, D. M., Hou, Y., Schroeder, S. G., Alkan, C., Cardone, M. F., Matukumalli, L. K., et al. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* 22, 778–790. doi: 10.1101/gr.133967.111
- Bionaz, M., and Loor, J. J. (2008). ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. *J. Nutr.* 138, 1019–1024. doi: 10.1093/jn/138.6.1019
- Cao, X.-K., Cheng, J., Huang, Y.-Z., Wang, X.-G., Ma, Y.-L., Peng, S.-J., et al. (2019). Growth performance and meat quality evaluations in three-way cross cattle developed for the tibetan plateau and their molecular understanding by integrative omics analysis. *J. Agric. Food Chem.* 67, 541–550. doi: 10.1021/acs.jafc.8b05477
- Cao, X.-K., Huang, Y.-Z., Ma, Y.-L., Cheng, J., Qu, Z.-X., Ma, Y., et al. (2018). Integrating CNVs into meta-QTL identified GBP4 as positional candidate for adult cattle stature. *Funct. Integrat. Genomics* 18, 559–567. doi: 10.1007/s10142-018-0613-0
- Cecchinato, A., Ribeca, C., Chessa, S., Cipolat-Gotet, C., Maretto, F., Casellas, J., et al. (2014). Candidate gene association analysis for milk yield, composition, urea nitrogen and somatic cell scores in Brown Swiss cows. *Animal* 8, 1062–1070. doi: 10.1017/S1751731114001098

- Cheng, J., Jiang, R., Cao, X. K., Liu, M., Huang, Y. Z., Lan, X. Y., et al. (2019). Association analysis of SSTR2 copy number variation with cattle stature and its expression analysis in Chinese beef cattle. *J. Agric. Sci.* 157, 365–374. doi: 10.1017/s0021859619000674
- Cloutier, N., Allaey, I., Marcoux, G., Machlus, K. R., Mailhot, B., Zufferey, A., et al. (2018). Platelets release pathogenic serotonin and return to circulation after immune complex-mediated sequestration. *Proc. Natl. Acad. Sci. U.S.A.* 115, E1550–E1559. doi: 10.1073/pnas.1720553115
- Cohen-Zinder, M., Seroussi, E., Larkin, D. M., Loo, J. J., Everts-Van Der Wind, A., Lee, J.-H., et al. (2005). Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 15, 936–944. doi: 10.1101/gr.3806705
- de Almeida Santana, M. H., Oliveira Junior, G. A., Mello Cesar, A. S., Freua, M. C., Gomes, R. D. C., Da Luz, E., et al. (2016). Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. *J. Appl. Genet.* 57, 495–504. doi: 10.1007/s13353-016-0344-7
- Di Gerlando, R., Sardina, M. T., Tolone, M., Suter, A. M., Mastrangelo, S., and Portolano, B. (2018). Genome-wide detection of copy-number variations in local cattle breeds. *Anim. Prod. Sci.* 59, 815–822.
- Do, D. N., Strathe, A. B., Ostensen, T., Pant, S. D., and Kadarmideen, H. N. (2014). Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. *Front. Genet.* 5:307. doi: 10.3389/fgene.2014.00307
- Duran Aguilar, M., Roman Ponce, S. I., Ruiz Lopez, F. J., Gonzalez Padilla, E., Vasquez Pelaez, C. G., Bagnato, A., et al. (2017). Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *J. Anim. Breed. Genet.* 134, 49–59. doi: 10.1111/jbg.12238
- Flint, J., Bolormaa, S., Pryce, J. E., Reverter, A., Zhang, Y., Barendse, W., et al. (2014). A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet.* 10:e1004198. doi: 10.1371/journal.pgen.1004198
- Fontanesi, L., Beretti, F., Martelli, P. L., Colombo, M., Dall'olio, S., Occidente, M., et al. (2011). A first comparative map of copy number variations in the sheep genome. *Genomics* 97, 158–165. doi: 10.1016/j.ygeno.2010.11.005
- Georges, M., Charlier, C., and Hayes, B. (2019). Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* 20, 135–156. doi: 10.1038/s41576-018-0082-2
- Jin, C., Bao, J., Wang, Y., Chen, W., Wu, T., Wang, L., et al. (2018). Changes in long non-coding RNA expression profiles related to the antagonistic effects of *Escherichia coli* F17 on lamb spleens. *Sci. Rep.* 8:16514. doi: 10.1038/s41598-018-34291-0
- Kärst, S., Cheng, R., Schmitt, A. O., Yang, H., De Villena, F. P. M., Palmer, A. A., et al. (2011). Genetic determinants for intramuscular fat content and water-holding capacity in mice selected for high muscle mass. *Mammalian Genome* 22, 530–543. doi: 10.1007/s00335-011-9342-6
- Liu, J., Zhang, L., Xu, L., Ren, H., Lu, J., Zhang, X., et al. (2013). Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* 14:299. doi: 10.1186/1471-2164-14-229
- Long, Y., Su, Y., Ai, H., Zhang, Z., Yang, B., Ruan, G., et al. (2016). A genome-wide association study of copy number variations with umbilical hernia in swine. *Anim. Genet.* 47, 298–305. doi: 10.1111/age.12402
- Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025. doi: 10.1016/j.cell.2015.04.004
- McCarroll, S. A., and Altshuler, D. M. (2007). Copy-number variation and association studies of human disease. *Nat. Genet.* 39, S37–S42.
- Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., et al. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature* 470, 59–65. doi: 10.1038/nature09708
- Pan, Z., Li, S., Liu, Q., Wang, Z., Zhou, Z., Di, R., et al. (2018). Whole-genome sequences of 89 Chinese sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-feralization. *GigaScience* 7:giy019. doi: 10.1093/gigascience/giy019
- Prinsen, R. T. M. M., Rossoni, A., Gredler, B., Bieber, A., Bagnato, A., and Strillacci, M. G. (2017). A genome wide association study between CNVs and quantitative traits in brown swiss cattle. *Livestock Sci.* 202, 7–12. doi: 10.1016/j.livsci.2017.05.011
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.
- Schaid, D. J., Chen, W., and Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* 19, 491–504. doi: 10.1038/s41576-018-0016-z
- Speck, N. A., and Gilliland, D. G. (2002). Core-binding factors in haematopoiesis and leukaemia. *Nat. Rev. Cancer* 2, 502–513. doi: 10.1038/nrc840
- Wang, X., Zheng, Z., Cai, Y., Chen, T., Li, C., Fu, W., et al. (2017). CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. *GigaScience* 6:gix115. doi: 10.1093/gigascience/gix115
- White, S. N., Mousel, M. R., Herrmann-Hoesing, L. M., Reynolds, J. O., Leymaster, K. A., Neiberger, H. L., et al. (2012). Genome-wide association identifies multiple genomic regions associated with susceptibility to and control of ovine lentivirus. *PLoS One* 7:e47829. doi: 10.1371/journal.pone.0047829
- Wijmenga, C., and Zernakova, A. (2018). The importance of cohort studies in the post-GWAS era. *Nat. Genet.* 50, 322–328. doi: 10.1038/s41588-018-0066-3
- Will, A. J., Cova, G., Osterwalder, M., Chan, W.-L., Wittler, L., Brieske, N., et al. (2017). Composition and dosage of a multipartite enhancer cluster control developmental expression of *Ihh* (Indian hedgehog). *Nat. Genet.* 49, 1539–1545. doi: 10.1038/ng.3939
- Xu, L., Cole, J. B., Bickhart, D. M., Hou, Y., Song, J., Vanraden, P. M., et al. (2014). Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. *BMC Genomics* 15:683. doi: 10.1186/1471-2164-15-683
- Xu, Y., Jiang, Y., Shi, T., Cai, H., Lan, X., Zhao, X., et al. (2017). Whole-genome sequencing reveals mutational landscape underlying phenotypic differences between two widespread Chinese cattle breeds. *PLoS One* 12:e0183921. doi: 10.1371/journal.pone.0183921
- Yoshikawa, M., Senzaki, K., Yokomizo, T., Takahashi, S., Ozaki, S., and Shiga, T. (2007). Runx1 selectively regulates cell fate specification and axonal projections of dorsal root ganglion neurons. *Dev. Biol.* 303, 663–674. doi: 10.1016/j.ydbio.2006.12.007
- Yue, W., Fang, X., Zhang, C., Pang, Y., Xu, H., Gu, C., et al. (2011). Two novel SNPs of the ABCG2 gene and its associations with milk traits in Chinese Holsteins. *Mol. Biol. Rep.* 38, 2927–2932. doi: 10.1007/s11033-010-9955-y
- Zhang, L., Jia, S., Yang, M., Xu, Y., Li, C., Sun, J., et al. (2014). Detection of copy number variations and their effects in Chinese bulls. *BMC Genomics* 15:480. doi: 10.1186/1471-2164-15-480
- Zhang, S.-Z., Xu, Y., Xie, H.-Q., Li, X.-Q., Wei, Y.-Q., and Yang, Z.-M. (2009). The possible role of myosin light chain in myoblast proliferation. *Biol. Res.* 42, 121–132.
- Zhang, X., Choi, P. S., Francis, J. M., Gao, G. F., Campbell, J. D., Ramachandran, A., et al. (2018). Somatic superenhancer duplications and hotspot mutations lead to oncogenic activation of the KLF5 transcription factor. *Cancer Discov.* 8, 108–125. doi: 10.1158/2159-8290.CD-17-0532
- Zhou, Y., Utsunomiya, Y. T., Xu, L., Hay, E. H. A., Bickhart, D. M., Alexandre, P. A., et al. (2016). Genome-wide CNV analysis reveals variants associated with growth traits in *Bos indicus*. *BMC Genomics* 17:419. doi: 10.1186/s12864-016-2461-4
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523. doi: 10.1038/s41467-019-09234-6

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cheng, Cao, Hanif, Pi, Hu, Huang, Lan, Lei and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.