Journal of Animal Science, 2022, **101**, 1–9
https://doi.org/10.1093/jas/skac383
Advance access publication 19 November 2022
Animal Genetics and Genomics

Design and characterization of a high-resolution multiple-SNP capture array by target sequencing for sheep

Yingwei Guo,[†] Fengting Bai,[†] Jintao Wang,[†] Shaoyin Fu,[‡] Yu Zhang,[†] Xiaoyi Liu,[†] Zhuangbiao Zhang,[†] Junjie Shao,[†] Ran Li,[†] Fei Wang,[†] Lei Zhang,[†] Huiling Zheng,[†] Xihong Wang,[†] Yongbin Liu,[‖] and Yu Jiang[†,1]

[†]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China
[‡]Institute of Animal Science, Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot 010031, China
[‖]School of Life Science, Inner Mongolia University, Hohhot 010070, China

[1]Corresponding author: yu.jiang@nwafu.edu.cn

Abstract

The efficiency of molecular breeding largely depends on inexpensive genotyping arrays. In this study, we aimed to develop an ovine high-resolution multiple-single-nucleotide polymorphism (**SNP**) capture array, based on genotyping by target sequencing (**GBTS**) system with capture-in-solution (liquid chip) technology. All the markers were from 40K captured regions, including genes located within selective sweep regions, breed-specific regions, quantitative trait loci (**QTL**), and the potential functional SNPs on the sheep genome. The results showed that a total of 210K high-quality SNPs were identified in the 40K regions, indicating a high average capture ratio (99.7%) for the target genomic regions. Using genotyped data (n = 317) from liquid chip technology, we further performed genome-wide association studies (**GWAS**) to detect the genetic loci affecting sheep hair types and teat number. A single significant association signal for hair types was identified on 6.7-7.1 Mb of chromosome 25. The *IRF2BP2* gene (chr25: 7,067,974-7,071,785), which is located within this genomic region, has been previously known to be involved in hair/wool traits in sheep. The results further showed a new candidate region around 26.4 Mb of chromosome 13, between the *ARHGAP21* and *KIAA1217* genes, that was significantly related to teat number in sheep. The haplotype patterns of this region also showed differences in animals with 2, 3, or 4 teats. Advances in using the high-accuracy and low-cost liquid chip are expected to accelerate sheep genomic and breeding studies in the coming years.

Lay Summary

Large-scale genotyping platforms are valuable tools for animal selection and breeding programs. The bead chip has been widely used in both research and commercial applications for a long time. A highly efficient and economical genotyping platform has been developed recently. In the present study, by combining the advantages of resequencing and bead chips, we developed a high-resolution capture array based on target sequencing with capture-in-solution technology (liquid chip), including updated functional probes according to the latest research. We further evaluated this approach by using 317 individuals and found that 210K single-nucleotide polymorphisms can be accurately genotyped, confirming the ratio of the captured regions compared with the designed rations is around 99.7%. Genome-wide association studies conducted using this chip suggested *IRF2BP2* gene may be involved in hair types and ARHGAP21-KIAA1217 locus may be related to teats number. The liquid chip with high accuracy and low cost can be widely used in genome-wide association studies and genome selection, supporting efforts in molecular breeding and genetic improvement of sheep.

Key words: chip design, genotyping by target sequencing, GWAS, sheep breeding

Abbreviations: F_{ST}, fixation index; GBTS, genotyping by target sequencing; GS, genome selection; GWAS, genome-wide association studies; LD, linkage disequilibrium; MAF, minor allele frequency; MQ, mapping quality; mSNP, multiple single-nucleotide polymorphisms; PCR, polymerase chain reaction; QTL, quantitative trait loci; SDFR, SNP density in flanking region; SNP, single-nucleotide polymorphisms

Introduction

Single-nucleotide polymorphism (**SNP**) chips have been widely used in different molecular breeding programs because of their high accuracy (Gershoni et al., 2022) and low cost (Suratannon et al., 2020). Currently, the most popular Ovine SNP chips are the Illumina Ovine 50K BeadChip (Magee et al., 2010) and Ovine 600K Beadchip (Kranis et al., 2013), which were developed in the last ten years, mainly based on European sheep breeds. These chip data have been used for the analysis of genetic diversity (Mastrangelo et al., 2017), population genetics (Kijas et al., 2012), selective sweeps (Fariello et al., 2014), and genome-wide association studies (**GWAS**) (Kijas et al., 2013). A new genotyping technology, referred to as genotyping by target sequencing (**GBTS**), can target more sites than the 50K BeadChip, by using 40K probes (Guo et al., 2019). GBTS can be performed through multiplexing polymerase chain reaction (**PCR**) (GenoPlexs) and regular PCR (GenoBaits), depending on the number of

Received December 29, 2021 Accepted November 16, 2022.

markers. The latter captures DNA fraction in solution, which is also called a liquid chip. The principle of the liquid chip is based on designing probes in the vicinity of the target SNPs and hybridizing with the targeted regions of the genome. After elution, amplification, and library building, high-depth next-generation sequencing is performed to genotype SNPs on all captured sites. Capturing sites with high SNP density in their surrounding regions can provide a set of adjacent SNPs, also called multiple SNPs (**mSNPs**), which are more conducive to haplotype analysis and genotype imputation. Studies on maize have demonstrated that mSNPs can significantly increase the detection efficiency of GWAS signals compared to SNPs (Guo et al., 2019, 2021). In addition, liquid chips based on GBTS have the advantages of flexible design, sites that can be added at any time, and low price, which makes large-scale genotyping of sheep possible at a lower cost.

In the previous chip design, the marker selection process mainly considered the SNPs' frequency and distribution. However, some candidate loci and regions related with important traits in sheep have already been reported by multi-omics studies, which can provide more useful information for marker selection during chip design. First, the genotypic divergence of domestic and wild animals can reflect the phenotypic changes during domestication, such as presence of horns (Kardos et al., 2015), the number of vertebrae (Rubin et al., 2012), dietary intake (Axelsson et al., 2013), immunity (Zheng et al., 2020), and other important traits (Rochus et al., 2018). Therefore, genome screen for selection signatures is a powerful method to detect variation in adaptive response during domestication. Second, each breed has unique genomic characteristics, which may be related to environmental adaptation and breed-specific traits (Li et al., 2020b; Xu et al., 2021). Third, Animal QTL Database (Hu et al., 2013) contains many quantitative trait loci (**QTL**) regions related to the important economic traits of sheep. Fourth, SNPs known to be related to economic traits are also important for chip design. For example, heterozygote genotypes at the FecB locus can increase the average number of ovulations by 1.5 and the average number of lambs by 1, and homozygotes can increase the number of ovulations by 3 and the number of lambs by 1.5 (Davis, 2005). Fifth, a draft assembly of the ovine Y chromosome has been reported in our previous study (Li et al., 2020a). Including SNPs on the Y chromosome during chip design allows the study of paternal origin and sex identification. In conclusion, the chip design containing the above potential functional sites can improve the detection accuracy of subsequent analyses of GWAS and genome selection (**GS**).

Therefore, with the development of equipment and technique, we designed an accurate, high-throughput, and low-cost liquid chip for sheep genotyping. We further validated its performance by performing GWAS on 317 animals. The chip will be a powerful tool for sheep breeding and research.

Materials and Methods

All experimental procedures were approved by the Northwest A&F University Animal Care Committee (permit number: NWAFAC1019) and performed in accordance with the guidelines.

Data collection and detection of the candidate SNP markers

We collected and generated resequencing data from 36 wild sheep and 551 domestic animals belonging to 61 breeds from all over the world (Supplementary Table S1). All cleaned reads were mapped to the sheep reference assembly Oar4.0 (GCF_000298735.2) using BWA-MEM (0.7.13-r1126) with default parameters (Li, 2013). Duplicate reads were removed using Picard Tools (http://broadinstitute.github.io/picard/). SNPs were detected by the Genome Analysis Toolkit (GATK, version 3.8-0-ge9d806836) (Auwera et al., 2013).

To get high-quality SNPs for chip design, stringent filter values as the GATK website recommended were applied for all variants with quality by depth < 2.0, root mean square of mapping quality (**MQ**) < 40.0, Fisher strand > 60.0, symmetric odds ratio test > 3.0, MQRankSum < −12.5, and ReadPosRankSum < −8.0. After this, variants with reads depth (DP) < mean read depth/3 or > mean read depth × 3 were further excluded using VCFtools (Danecek et al., 2011). Finally, a total of 99,406,384 SNVs were obtained.

Development of the SNP panel

Sites identification in domestication and selection regions

We detected the selective sweep signals by searching the genome for regions with high fixation index (F_{ST}) values and high differences in genetic diversity (π log ratio). First, we calculated the F_{ST} and π log ratio in sliding 50-kb windows with 25-kb steps along the autosomes using VCFtools (Danecek et al., 2011) and in-house scripts for comparisons between wild and domestic sheep, wild and European sheep, wild and Chinese sheep. We then filtered out any windows that had fewer than 10 SNPs in the F_{ST} and π log-ratio results. At last, we selected the top 1% regions as the divergent regions in each comparison.

Then, the single-locus F_{ST} and single-locus π log ratio of each SNP in the divergent regions were calculated using default parameters by VCFtools and in-house scripts for comparisons between wild and domestic sheep, wild and European sheep, wild and Chinese sheep. The single-locus π values were obtained in domestic, European, and Chinese sheep, respectively.

Finally, the top 50 sites with the highest single-locus F_{ST}, highest single-locus π log ratio, and lowest single-locus π value among the divergent regions were selected separately, and the intersection using scripts written in-house was taken as the candidate sites for each interval.

Sites selection in breed-specific and animal QTL database-related regions

Genome-wide screening and functional annotation studies have identified genomic regions related to important traits in different sheep breeds, such as reproduction (Hu sheep vs. Tan Sheep), milk yielding (East Friesian milk sheep vs. Finn sheep; Li et al., 2020b). Besides, some of the QTL regions which have been previously documented in the animal QTL database (Hu et al., 2013), associated with major economic traits (such as growth and development, reproduction, and milk production) in sheep, were also used as candidate intervals for chip design. Breed-specific regions and the regions in the sheep QTL database mentioned above were combined into larger regions using the "merge" subcommand of BEDtools (Quinlan and Hall, 2010).

For each region described above, tagSNPs were found as the representative sites in each region. We used PLINK v1.9 (Purcell et al., 2007) to filter SNPs with options: "--maf

0.1 --geno 0.1 --hwe 0.001." Then, pairwise tagging was performed by Haploview v4.2 (Barrett et al., 2005). We regarded all SNPs in the blocks assigned by Haploview as the candidate SNPs (rather than the tagSNPs recommended by the software).

For the two kinds of candidate regions above, we counted the total number of SNPs within 100 bp upstream and downstream of all tagSNPs in these regions (including itself). This value was defined as SDFR (SNP Density in Flanking Region). After that, the sites with SNP density between 3 and 7 were retained as candidate SNPs. For each candidate region, we selected SNPs from regions with the highest SDFR as candidates for every 50 kb fragment.

Targeting potential functional sites for genotyping

Many SNPs reported by GWAS studies are associated with some economic traits, including growth, reproduction, milk traits, appearance, etc. (Table 1). The SNPs located on can help to improve the accuracy of GWAS and GS studies. A total of 209 SNPs from 48 studies (Table S9) were directly added to the chip panel candidate sites dataset.

We selected 26 SNPs in the male-specific region of the sheep Y chromosome according to our previous study (Li et al., 2020a) and added these loci to the chip sites dataset (see detail in Supplementary Table S2) to enable the identification and verification of the sex of the samples to be tested. In addition, we selected conserved loci on two strains of *Brucella melitensis* (*Brucella melitensis* str. M1981 and *Brucella melitensis* str. RM57), including the molecular marker VirB12 (Rolan et al., 2008) for detection of brucellosis in sheep. Since the loci on the Y chromosome and the sites on the sheep brucella genome are not on the sheep reference genome (Oar4.0), they were not taken into account in the next filling stage, although they were also used as backbone sites.

Table 1. The number of SNPs in important functional genes in the chip

Genes	Sites number	Traits	References
BMPR1B	19	Litter size	(Souza et al., 2001)
BMP15	5	Litter size	(Dixit et al., 2006)
GDF9	7	Litter size	(Hanrahan et al., 2004)
LEMD3	3	Ear size	(Zhang et al., 2014)
MSRB3	4	Ear size	(Paris et al., 2020)
MSTN	14	Muscle	(Hickford et al., 2010)
RXFP2	13	Horns	(Luhken et al., 2016)
CSN1S1	10	Milk composition	(Calvo et al., 2013)
DGAT1	3	Milk composition	(Martin et al., 2017)
FBXL3	3	Circadian rhythm	(Godinho et al., 2007)
PDGFD	17	Tail fat	(Dong et al., 2020)
TBXT	8	Tail length	(Han et al., 2019)
IRF2BP2	20	Fleece	(Demars et al., 2017)
KRT36	7	Fleece	(Sulayman et al., 2018)
BCO2	10	Fat deposition	(Våge and Boman, 2010)
NRIP1	1	Fat deposition	(Xu et al., 2017)
VRTN	5	Vertebral number	(Li et al., 2019)

Gap filling

We first calculated the SDFR values of SNPs on the GGP Ovine HD (Zhou, 2019) with minor allele frequency (**MAF**) > 0.05, and then screened out all sites with SDFR values between 3 and 7 as the source of filling. Next, the interval between adjacent backbone sites with spacing greater than 100 Kb was filled by SNPs with SDFR values between 3 and 7 on the GGP Ovine HD. Then, we selected SNPs using the same method described above according to the SDFR values.

After combination and de-redundance, all sites were considered for the adjacent spacing, and the intervals with spacing greater than 100 Kb were filled again using the same method.

Probe production and sequencing

The liquid chip is based on GBTS technology, which relies on target capture by the complementary combination of the probes and the target sequences. To ensure capture efficiency, GenoBaits Probe Designer (Jianan Zhang, Molbreeding Biotech.) was used for designing two probes with 60% to 70% overlap, both of which should cover the target SNP sites. The length of a designed probe is 110 bp, with the GC content between 30% and 70%, excluding the ones with non-specific amplification, and without simple sequence repeat or gaps.

Genomic DNA was extracted according to the standard phenol-chloroform method from whole blood. The libraries were constructed using the GenoBaits DNA-seq Library Prep Kit (MolBreeding Biotechnology Co., Shijiazhuang, Hebei, China) according to the manufacturer's protocol. Next, probes and hybridization buffer were mixed and hybridized at 65 °C for 16 h. Then, Dynabeads MyOne Streptavidin C1 and binding buffer were put in the solution to enrich the target DNA fragments but remove the non-target ones. The target fragments were amplified by library amplification primer and DNA polymerase. Next, two rounds of purification were conducted using Beckman AMPure Beads. Finally, Qubit 2.0 Fluorometer (Thermo Fisher Scientific, CA) and qPCR were used to quantify the library concentration and sequencing was done with PE150 on the MGISEQ-2000 platform (MGI, Shenzhen, China).

Evaluation of the liquid chip

Samples, genotyping, and imputation

Samples were collected from a crossbred generated by backcrossing F1 East-Friesian × Hu dams to East-Friesian sires bred by Gansu Yuansheng Agriculture and Animal Husbandry Technology Co., Ltd. (Jinchang, Gansu, China). A total of 323 samples were collected. Phenotypes include fleece types in the tail and teat number. DNA was extracted from blood and sequenced by the 40K liquid chip. We combined the sheep reference genome Oar1.0, Y chromosome, and sheep brucella genome as the new reference genome, called Oar1.0ForChip, and used the above methods for mapping, calling, and filtering SNPs. Additionally, we used BCFtools (Danecek et al., 2021) to filter SNPs again with options: "-i 'F_MISSING < 0.1 & MAF > 0.05' -v snps -m2 -M2," remaining 317 sheep and 209,625 SNPs.

Then, we first used conform-gt (https://faculty.washington.edu/browning/conform-gt.html) to assign alleles in our VCF file consistent with the reference VCF file. We used Beagle5.0 (Browning et al., 2018) to impute missing SNPs using a reference panel of 43 East Friesen sheep and 8 Hu sheep from the same farm with default parameters. To ensure a high imputation accuracy, we used the criterion of dosage R-squared > 0.8 according to a previous study (Pook et al., 2020) to filter the imputed VCF, resulting in a total of 647,471 SNPs. Finally, we converted the VCF file to a PLINK file.

Genome-wide association study and variants annotation

GWAS was conducted by GEMMA (0.98.3) using the following mixed linear model (Zhou and Stephens, 2012):

$$y = W\alpha + x\beta + u + \varepsilon,$$

where y is the phenotypes of $n \times 1$ vector; W denotes the $n \times c$ matrix of covariates (fixed effects); α is the $c \times 1$ vector of the corresponding coefficients including the intercept; x is the $n \times 1$ vector of markers; β is the effect size of the markers; u is $n \times 1$ vector of the random effect with u ~ N $(0, KV_g)$; K represents the known $n \times n$ relatedness matrix calculated by SNP markers, and V_g means polygenic additive variance; ε is the random error vector with ε ~ N $(0, IV_e)$, and I denotes $n \times n$ identity matrix, V_e means polygenic residual component.

In this study, the fixed effects were the first three principal components, and the relatedness matrix was used as the random effect. The number of independent SNPs was estimated based on the linkage disequilibrium (**LD**) analysis performed with PLINK v.1.9 (Purcell et al., 2007), with the option "--indep-pairwise 50 5 0.4." PLINK calculated the LD between each pair of SNPs in a window of 50 SNPs with a step size of 5 SNPs and removed one of a pair of SNPs if the LD is greater than 0.4. The threshold of genome-wide significance (P genome < 0.05/number of independent SNPs) was determined by the Bonferroni correction at the empirical level of 0.05 (Marees et al., 2018).

The functional annotation of significant variant SNPs in GWAS results was performed using ANNOVAR (1 February 2016) (Wang et al., 2010).

Results

Identifying SNP sources from global sheep breeds

To provide a reliable dataset for panel design, we collected re-sequencing data from 587 sheep, including 551 domestic individuals belonging to 64 breeds from all over the world (Supplementary Table S1). There were also 36 wild samples, including Mouflon, Urial, Argali, Bighorn sheep, and Thinhorn sheep (Supplementary Table S1). All the above data were mapped to the sheep reference genome Oar4.0. All variants were subjected to stringent filtering and a total of 99,406,384 SNPs were finally obtained as the source for the following chip marker selection.

Development of the chip panel

Regions and sites with potential functions were selected as the skeleton of the chip structure. The genomic regions included domestication-related regions (143; type I in Figure 1A), selective sweeps regions of Chinese (224; type II in Figure 1A) and European sheep (252; type III in Figure 1A), breeds-specific regions (575; type IV in Figure 1A), and sheep QTL regions (353; type V in Figure 1A).

Besides the potential functional regions, 149 SNPs associated with economic traits from previous studies (Table 1) and 60 SNPs from an additional GWAS (Li et al., 2020a) were added to the skeleton sites set without any filtering (type VI in Figure 1A). A total of 26 SNPs came from the non-homologous regions of the Y chromosome also used as candidate sites for panel design (typeVII in Figure 1A), to identify the paternal origin. Six SNPs of the conserved regions on the *Brucella melitensis* genome were selected to detect whether the samples were infected by Brucella (type VIII in Figure 1A).

All mentioned above SNPs were merged and then duplicates were removed. A total of 7,594 SNPs were selected and considered as the skeleton sites. After that, 40,579 SNPs were added in the large intervals between two adjacent sites to ensure even distribution across the genome (type IX in Figure 1A). A total of 48,173 candidate SNPs were used to assess the ability of probe designs.

After evaluating all candidate SNPs, a total of 40,156 sites were finally used for probe design, including 5,741 potential functional sites and 34,415 filler sites (Table 2; Supplementary Tables S2 and S8). A strong correlation was found between the length of chromosomes and the number of sites selected on each chromosome (Figure 1B; Supplementary Table S3; Spearman correlation; $r = 0.988$; P-value = 2.2×10^{-16}), indicating that all sites in the panel were evenly distributed on the sheep genome. The average spacing between adjacent sites was about 64.4K, and the spacing was mostly between 50K and 70K (Figure 1C).

Genotyping performance of the chip

To validate this chip, 317 genomes of F2 hybrids of East Friesian and Hu Sheep were genotyped. The average capture ratio of the 40K probes was 99.7%, and 98.7% of the individuals had a capture ratio of 99.0% or higher (Figure 2A, Supplementary Table S4). At least 126 Mb and 17Mb of the genome were covered above 10× and 5× (Supplementary Table S5), respectively.

A total of 210K high-quality SNPs were identified in the 40K regions, suggesting more than five SNPs per region can be genotyped on average (Supplementary Table S6). Among the 210K SNPs, 99.71% of them had a MAF > 5% (Figure 2B), which was the basis for performing GWAS. Using the standard capture and sequencing pipeline, the average depth of the 40K target sites for probe design was about 70×. Within 100 bp and 200 bp nearby the target sites, the average depth can still reach more than about 30× and 5×, respectively (Figure 2C; Supplementary Table S7). The depth of autosomal loci in rams and ewes (from another dataset) was roughly the same, while the average depth of Y chromosome loci in rams was significantly higher than that in ewes (Figure 2D). The latter one was found only 0.03×. Therefore, the average depth of Y chromosome sites can be used to identify the sex or perform sex verification of the sample.

High-resolution GWAS for fleece and teat number

To evaluate the performance of this chip, we conducted GWAS on fleece and teat number traits using 317 F2 hybrids mentioned above.

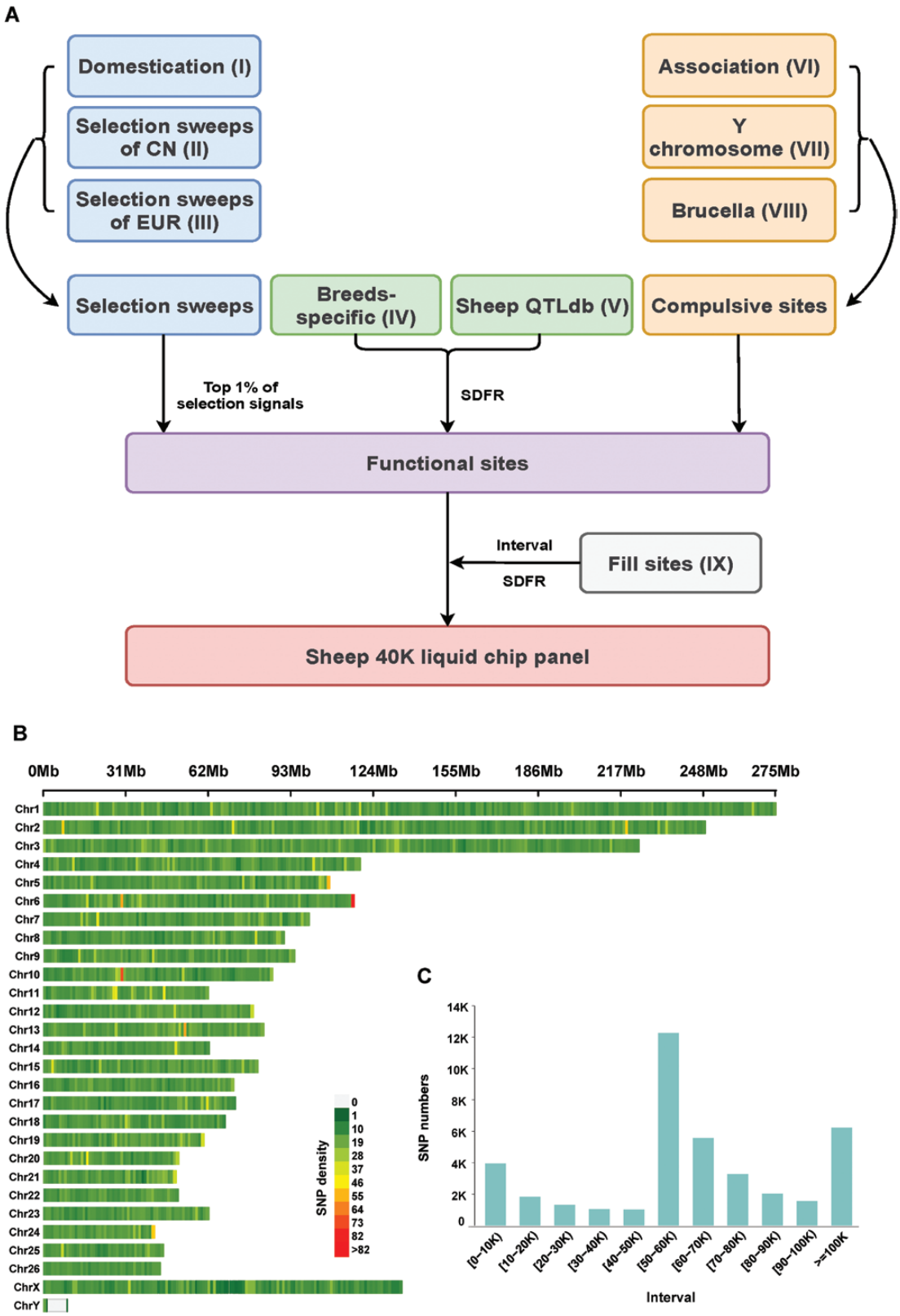

Figure 1. Roadmap and characterization of the SNPs on the chip. (A) Roadmap for the design of sheep liquid chip. (B) Distribution of the designed sites on the genome with 1Mb window size. (C) The frequency distribution of the spaces between adjacent SNPs.

There are two types of hair in our samples: long hairy fleece and short wooly fleece. GWAS for this trait detected 163 associated SNPs with a genome-wide significance (8.01E-7, 0.05/62422). These SNPs spanned from 6.7Mb to 7.1Mb on OAR25 (Supplementary Figure S1). There are eight genes in this region, including *SLC35F3*, *LOC105604913*, *LOC101113195*, *TARBP1*, *LOC105604914*, *LOC101111733*, *IRF2BP2*, and *LOC114110878*.

Table 2. The number of target SNPs in each term of SNP[1] source

Type	Category	Number of SNPs
I	SNPs in domestic regions	1,351
II	SNPs in the selective region of Chinese sheep	1,134
III	SNPs in the selective region of European sheep	547
IV	SNPs of breeds-specific regions	885
V	SNPs of Sheep QTL[2] regions	1,583
VI	Sites in genes about important traits	209
VII	Sites in the Y chromosome	26
VIII	Sites in the brucella genome	6
IX	Filled sites	34,415
	Total	40,156

[1]SNP, single-nucleotide polymorphism
[2]QTL, quantitative trait loci

To get more precise results, we imputed the chip data at sequence level. The GWAS result after imputation revealed the same association signals (genomic region located at 6.7 Mb to 7.1 Mb on OAR25) as before imputation (Figure 3A,B). Previous publications identified similar associations in the same region (Demars et al., 2017; Lv et al., 2021). Therefore, a high probe density was designed in this region and showed no significant improvement for associations compared to those before imputation.

Normally sheep have two teats, but a few ewes have one or two extra teats. These extra teats usually do not produce milk and may become a bacterial reservoir, negatively affecting machine milking (Pausch et al., 2012). The results of GWAS for teat number suggested an association signal on OAR13: 26,401,124 (Supplementary Figure S2), even though only one SNP passed the threshold.

To exclude false positives, we conducted another GWAS using the dataset after imputation. The results showed that the region associated with the teat number was located around 26.4 Mb on OAR13, and the positions of the three most significant SNPs were OAR13: 26398953 (rs425911101), OAR13: 26401124 (rs430301061), and OAR13: 26403722 (rs416056176) (Figure 3C,D). All three sites were located between the *ARHGAP21* and *KIAA1217* genes. The haplotype heatmap of this region showed different haplotype patterns between these three sites and other surrounding sites (Supplementary Figure S3). On these three sites, the individuals with two teats had a proportion of 71.5% of the reference allele (the blue grids), while the samples with four teats had a proportion of 30.2% of the reference allele (the blue grids) but 69.8% of the alternative allele (the red grids).

Discussion

Obtaining accurate genotypes at a lower cost is one of the urgent issues that needs to be solved in sheep breeding. In this study, a sheep liquid chip based on genotyping by target sequencing was developed from large-scale resequencing data by using some new chip design methods. The chip contains 5,741 potential functional SNPs related to domestication and selective sweeps, breeds-specific regions, QTLs, known functional SNPs/genes, sites on the Y chromosome, and sites on the Brucella genome. Besides, we selected 34,415 SNPs as filter sites with a high density of SNPs in their surrounding sequences. Finally, a total of 40,156 sites were used for chip production.

The chip's performance was assessed by genotyping 317 individuals. We observed a high capture rate due to probe design and strict filtration. The captured fragments were re-sequenced with a high depth, in which the target sites could be sequenced at a depth of 70X, which resulted in high genotyping accuracy. In addition, not only the target sites but also the variants around the target sites could be detected, including SNPs and indels. Besides, this chip could detect more than 200K SNPs and other variants with 40K probes, even novel variants, or variants with a very low frequency, which was beyond the Illumina Ovine 50K BeadChip's capability. Moreover, the price of this chip is comparable to the Illumina Ovine 50K BeadChip. In other words, by combining the advantages of re-sequencing and bead chips, the newly designed chip can obtain much more variants than Illumina Ovine 50K BeadChip at a comparable price, which is very economical for GWAS and other studies.

We performed GWAS on 317 F2 hybrids genotyped by this chip and found regions associated with traits of fleece and teat number. The result of GWAS for fleece type indicated an association signal on OAR25, including the *SLC35F3*, *TARBP1*, and *IRF2BP2* genes. A previous study has shown an insertion of the 3'-UTR of the *IRF2BP2* gene leads to a change between a long hairy fleece and a short wooly fleece (Demars et al., 2017). One recent study has shown that a novel mutation in the 3′UTR of *IRF2BP2* could be the causal variant for fleece types (Lv et al., 2021). However, since our samples belonged to the F2 population and the signal was localized to a relatively broad region, a more refined post-GWAS analysis might clarify the causal region more clearly. The genomic region association with teat number is located between the *ARHGAP21* and *KIAA1217* genes on OAR25. However, a similar study on sheep has reported genes related to teat number (*BBX* and *CD47*) on OAR1 (Peng et al., 2017), possibly because the mechanisms that determine sheep teat number traits are complex and still need to be demonstrated by GWAS or more accurate experiments with larger populations.

For the fleece type GWAS, the results after imputation did not improve much compared with those before imputation. The candidate region showed a high probe density because we screened out this region during the process of chip design by four data sources: sheep domestic regions (type I in Table 2), European sheep selective regions (type III in Table 2), sheep QTL regions (type V in Table 2), and important functional genes (type VI in Table 2). Additionally, there was a higher linkage disequilibrium between markers on this specific chromosomal region than that of the whole chromosome. The mean r^2 of a pair of SNPs in the region with an average distance of 5Kb was 0.8, while the mean r^2 of a pair of SNPs of the whole chromosome with the same distance was 0.6. Due to these two reasons, even without imputation, a significant association signal could be obtained (Supplementary Figure S1). This strategy of selecting more markers in the important region reduced the reliance on imputation due to low marker density. In contrast, the region associated with teat number

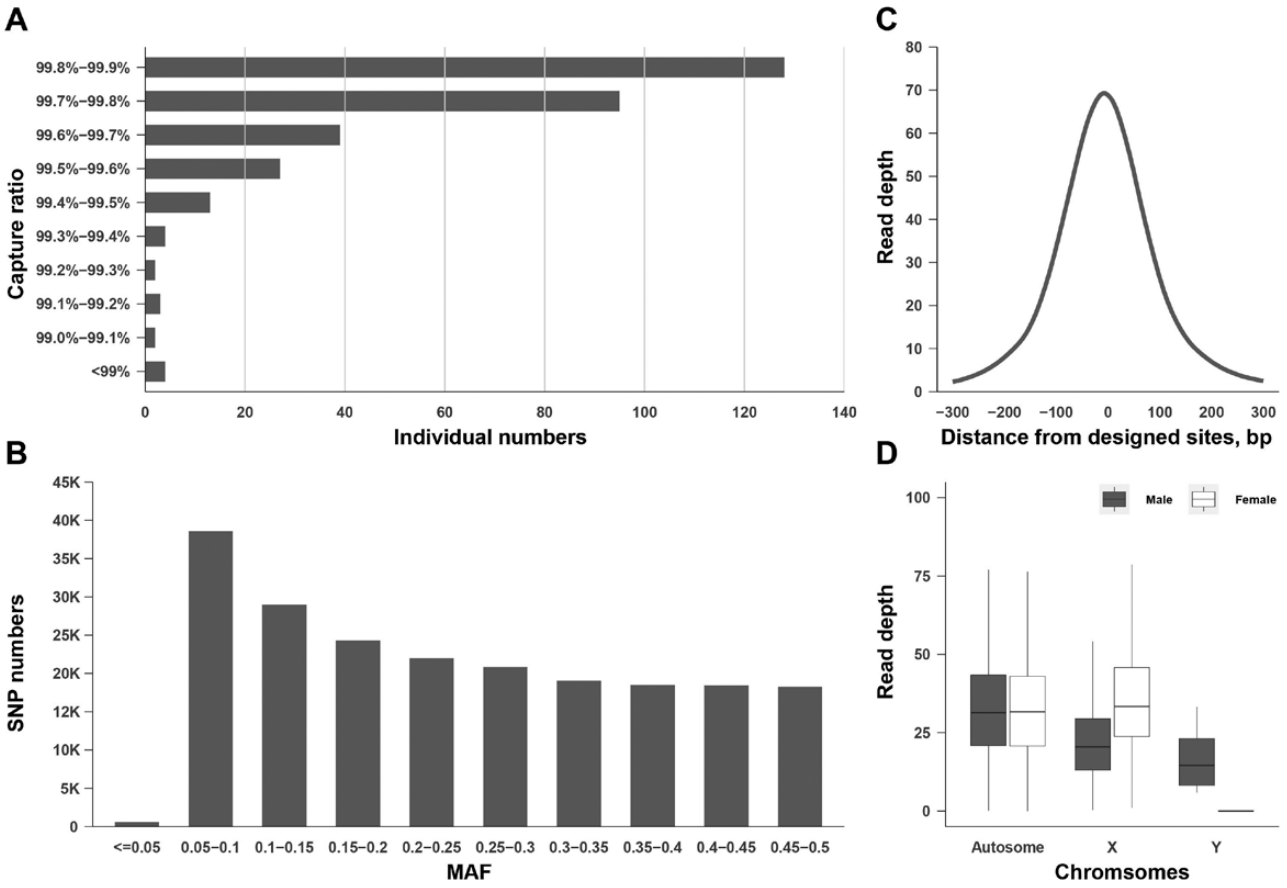

Figure 2. Validation of the sheep liquid chip. (A) Distribution of the capture ratio of the 317 sheep. (B) Distribution of the MAF of the SNPs on the chip and their flanking regions. (C) Read depth of designed SNPs and their flanking regions. (D) Depth of SNPs on the autosome, X chromosome, and Y chromosome, respectively.

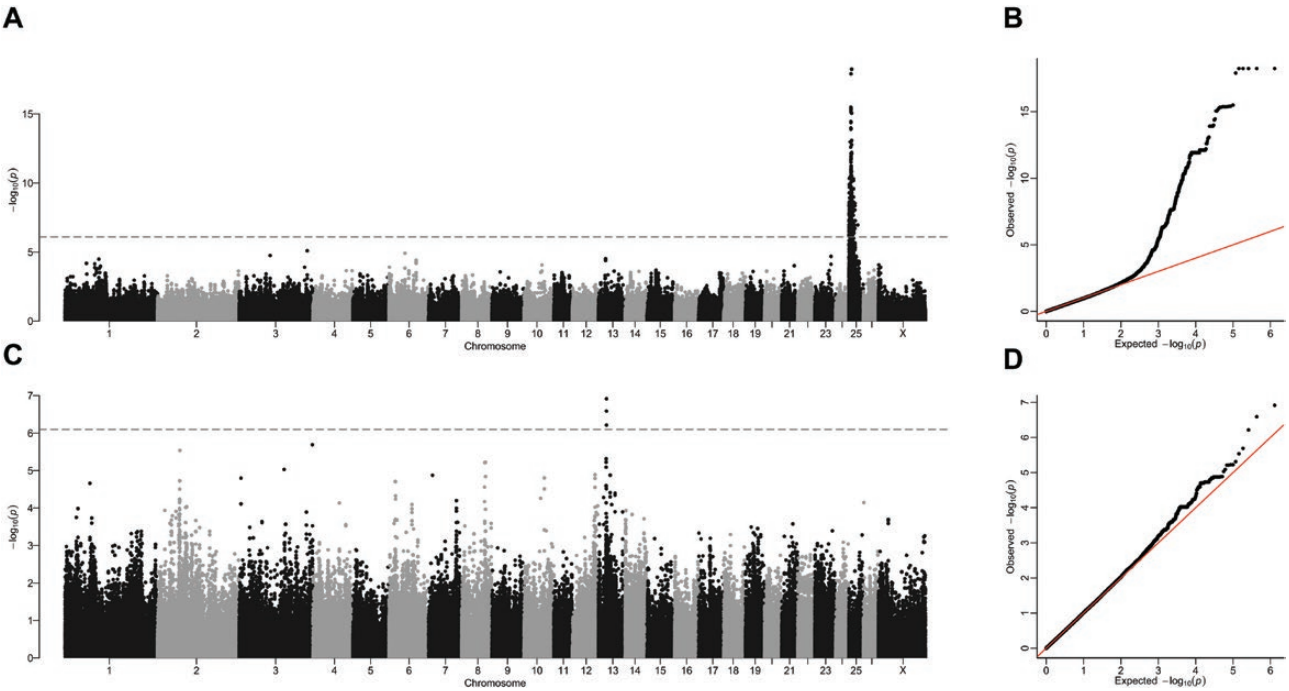

Figure 3. Manhattan and quantile-quantile plot of GWAS for hair types and teat number. (A, B) Manhattan and quantile-quantile plot for hair types. (C, D) Manhattan and quantile-quantile plot for teat number.

was a novel region, which had not been reported as a QTL in previous studies. During chip design for the regions like this, high-quality SNPs were evenly filled. Therefore, a significant association signal was obtained as well (Figure 3C,D; Supplementary Figure S2).

In conclusion, a low-cost, high-accuracy sheep whole-genome liquid chip based on re-sequencing data has been developed. This chip can be widely used in GWAS and GS for large-scale genotyping, offering a new tool for molecular breeding and genetic improvement of sheep.

Supplementary Data

Supplementary data are available at *Journal of Animal Science* online.

Acknowledgments

This project was supported by grants from the National Natural Science Foundation of China (U21A20247, 31822052). We thank the High-Performance Computing (HPC) of Northwest A&F University (NWAFU) for providing computing resources.

Conflict of Interest Statement

The authors declare no real or perceived conflicts of interest.

Literature Cited

Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, et al. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:11.10.1–11.19.33. doi:10.1002/0471250953.bi1110s43

Axelsson, E., A. Ratnakumar, M. L. Arendt, K. Maqbool, M. T. Webster, M. Perloski, O. Liberg, J. M. Arnemo, A. Hedhammar, and K. Lindblad-Toh. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495:360–364. doi:10.1038/nature11837

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265. doi:10.1093/bioinformatics/bth457

Browning, B. L., Y. Zhou, and S. R. Browning. 2018. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103:338–348. doi:10.1016/j.ajhg.2018.07.015

Calvo, J. H., E. Dervishi, P. Sarto, L. González-Calvo, B. Berzal-Herranz, F. Molino, M. Serrano, and M. Joy. 2013. Structural and functional characterisation of the αS1-casein (CSN1S1) gene and association studies with milk traits in Assaf sheep breed. *Livest. Sci.* 157:1–8. doi:10.1016/j.livsci.2013.06.014

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, et al; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. doi:10.1093/bioinformatics/btr330

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. Mccarthy, R. M. Davies, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. doi:10.1093/gigascience/giab008

Davis, G. 2005. Major genes affecting ovulation rate in sheep. *Genet. Sel. Evol.* 37:S1–S11. doi:10.1186/1297-9686-37-s1-s11

Demars, J., M. Cano, L. Drouilhet, F. Plisson-Petit, P. Bardou, S. Fabre, B. Servin, J. Sarry, F. Woloszyn, P. Mulsant, et al. 2017. Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep. *Mol. Biol. Evol.* 34:1722–1729. doi:10.1093/molbev/msx114

Dixit, H., L. K. Rao, V. V. Padmalatha, M. Kanakavalli, M. Deenadayal, N. Gupta, B. Chakrabarty, and L. Singh. 2006. Missense mutations in the BMP15 gene are associated with ovarian failure. *Hum. Genet.* 119:408–415. doi:10.1007/s00439-006-0150-0

Dong, K., M. Yang, J. Han, Q. Ma, J. Han, Z. Song, C. Luosang, N. A. Gorkhali, B. Yang, X. He, et al. 2020. Genomic analysis of worldwide sheep breeds reveals PDGFD as a major target of fat-tail selection in sheep. *BMC Genomics* 219:800. doi:10.1186/s12864-020-07210-9

Fariello, M. -I., B. Servin, G. Tosser-Klopp, R. Rupp, C. Moreno, M. S. Cristobal, and S. Boitard. 2014. Selection signatures in Worldwide sheep populations. *PLoS One* 9:e103813. doi:10.1371/journal.pone.0103813

Gershoni, M., A. Shirak, R. Raz, and E. Seroussi. 2022. Comparing BeadChip and WGS genotyping: non-technical failed calling is attributable to additional variation within the probe target sequence. *Genes* 13:485. doi:10.3390/genes13030485

Godinho, S. I. H., E. S. Maywood, L. Shaw, V. Tucci, A. R. Barnard, L. Busino, M. Pagano, R. Kendall, M. M. Quwailid, M. R. Romero, et al. 2007. The after-hours mutant reveals a role for Fbxl3 in determining mammalian circadian period. *Science* 316:897–900. doi:10.1126/science.1141138

Guo, Z., H. Wang, J. Tao, Y. Ren, C. Xu, K. Wu, C. Zou, J. Zhang, and Y. Xu. 2019. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. *Mol. Breed.* 39:37. doi:10.1007/s11032-019-0940-4

Guo, Z., Q. Yang, F. Huang, H. Zheng, Z. Sang, Y. Xu, C. Zhang, K. Wu, J. Tao, B. M. Prasanna, et al. 2021. Development of high-resolution multiple-SNP arrays for genetic analyses and molecular breeding through genotyping by target sequencing and liquid chip. *Plant Commun.* 2:100230. doi:10.1016/j.xplc.2021.100230

Han, J., M. Yang, T. Guo, C. Niu, J. Liu, Y. Yue, C. Yuan, and B. Yang. 2019. Two linked TBXT (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. *Anim. Genet.* 50:772–777. doi:10.1111/age.12852.

Hanrahan, J. P., S. M. Gregan, P. Mulsant, M. Mullen, G. H. Davis, R. Powell, and S. M. Galloway. 2004. Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belclare sheep (*Ovis aries*). *Biol. Reprod.* 70:900–909. doi:10.1095/biolreprod.103.023093

Hickford, J. G. H., R. H. Forrest, H. Zhou, Q. Fang, J. Han, C. M. Frampton, and A. L. Horrell. 2010. Polymorphisms in the ovine myostatin gene (MSTN) and their association with growth and carcass traits in New Zealand Romney sheep. *Anim. Genet.* 41:64–72. doi:10.1111/j.1365-2052.2009.01965.x

Hu, Z. -L., C. A. Park, X. -L. Wu, and J. M. Reecy. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 41:D871–D879. doi:10.1093/nar/gks1150

Kardos, M., G. Luikart, R. Bunch, S. Dewey, W. Edwards, S. Mcwilliam, J. Stephenson, F. W. Allendorf, J. T. Hogg, and J. Kijas. 2015. Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Mol. Ecol.* 24:5616–5632. doi:10.1111/mec.13415

Kijas, J. W., J. A. Lenstra, B. Hayes, S. Boitard, L. R. Porto Neto, M. San Cristobal, B. Servin, R. McCulloch, V. Whan, K. Gietzen, et al; M. International Sheep Genomics Consortium. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10:e1001258. doi:10.1371/journal.pbio.1001258

Kijas, J. W., M. Serrano, R. McCulloch, Y. Li, J. Salces Ortiz, J. H. Calvo, and M. D. Perez-Guzman; C. International Sheep Genomics. 2013. Genomewide association for a dominant pigmentation gene in sheep. *J. Anim. Breed. Genet.* 130:468–475. doi:10.1111/jbg.12048

Kranis, A., A. A. Gheyas, C. Boschiero, F. Turner, L. Yu, S. Smith, R. Talbot, A. Pirani, F. Brew, P. Kaiser, et al. 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14:59. doi:10.1186/1471-2164-14-59

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li, C., M. Li, X. Li, W. Ni, Y. Xu, R. Yao, B. Wei, M. Zhang, H. Li, Y. Zhao, et al. 2019. Whole-genome resequencing reveals loci associated with thoracic vertebrae number in sheep. *Front. Genet.* 10:674. doi:10.3389/fgene.2019.00674

Li, R., P. Yang, M. Li, W. Fang, X. Yue, H. A. Nanaei, S. Gan, D. Du, Y. Cai, X. Dai, et al. 2020a. A Hu sheep genome with the first ovine Y chromosome reveal introgression history after sheep domestication. *Sci. China Life Sci.* 64:1116–1130. doi:10.1007/s11427-020-1807-0

Li, X., J. Yang, M. Shen, X. L. Xie, G. J. Liu, Y. X. Xu, F. H. Lv, H. Yang, Y. L. Yang, C. B. Liu, et al. 2020b. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat. Commun.* 11:2815. doi:10.1038/s41467-020-16485-1

Luhken, G., S. Krebs, S. Rothammer, J. Kupper, B. Mioc, I. Russ, and I. Medugorac. 2016. The 1.78-kb insertion in the 3'-untranslated region of RXFP2 does not segregate with horn status in sheep breeds with variable horn status. *Genet. Sel. Evol.* 48:78. doi:10.1186/s12711-016-0256-3

Lv, F. H., Y. H. Cao, G. J. Liu, L. Y. Luo, R. Lu, M. J. Liu, W. R. Li, P. Zhou, X. H. Wang, M. Shen, et al. 2021. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression and agronomically important loci. *Mol. Biol. Evol.* 39:msab353. doi:10.1093/molbev/msab353

Magee, D. A., S. D. E. Park, E. Scraggs, A. M. Murphy, M. L. Doherty, J. W. Kijas, and D. E. Machugh; International Sheep Genomics Consortium. 2010. Technical note: High fidelity of whole-genome amplified sheep (*Ovis aries*) deoxyribonucleic acid using a high-density single nucleotide polymorphism array-based genotyping platform1. *J. Anim. Sci.* 88:3183–3186. doi:10.2527/jas.2009-2723

Marees, A. T., H. De Kluiver, S. Stringer, F. Vorspan, E. Curis, C. Marie-Claire, and E. M. Derks. 2018. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 27:e1608. doi:10.1002/mpr.1608

Martin, P., I. Palhière, C. Maroteau, P. Bardou, K. Canale-Tabet, J. Sarry, F. Woloszyn, J. Bertrand-Michel, I. Racke, H. Besir, et al. 2017. A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. *Sci. Rep.* 7:1872. doi:10.1038/s41598-017-02052-0

Mastrangelo, S., B. Portolano, R. Di Gerlando, R. Ciampolini, M. Tolone, and M. T. Sardina; International Sheep Genomics Consortium. 2017. Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. *Animal* 11:1107–1116. doi:10.1017/S1751731116002780

Paris, J. M., A. Letko, I. M. Häfliger, P. Ammann, and C. Drögemüller. 2020. Ear type in sheep is associated with the MSRB3 locus. *Anim. Genet.* 51:968–972. doi:10.1111/age.12994

Pausch, H., S. Jung, C. Edel, R. Emmerling, D. Krogmeier, K. -U. Götz, and R. Fries. 2012. Genome-wide association study uncovers four QTL predisposing to supernumerary teats in cattle. *Anim. Genet.* 43:689–695. doi:10.1111/j.1365-2052.2012.02340.x

Peng, W. F., S. S. Xu, X. Ren, F. H. Lv, X. L. Xie, Y. X. Zhao, M. Zhang, Z. Q. Shen, Y. L. Ren, L. Gao, et al. 2017. A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (*Ovis aries*). *Anim. Genet.* 48:570–579. doi:10.1111/age.12575

Pook, T., M. Mayer, J. Geibel, S. Weigend, D. Cavero, C. C. Schoen, and H. Simianer. 2020. Improving imputation quality in BEAGLE for crop and livestock data. *G3 Genes|Genomes|Genetics* 10:177–188. doi:10.1534/g3.119.400798

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. De Bakker, M. J. Daly, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. doi:10.1086/519795

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. doi:10.1093/bioinformatics/btq033

Rochus, C. M., F. Tortereau, F. Plisson-Petit, G. Restoux, C. Moreno-Romieux, G. Tosser-Klopp, and B. Servin. 2018. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. *BMC Genomics* 19:71. doi:10.1186/s12864-018-4447-x

Rolan, H. G., A. B. den Hartigh, M. Kahl-McDonagh, T. Ficht, L. G. Adams, and R. M. Tsolis. 2008. VirB12 is a serological marker of Brucella infection in experimental and natural hosts. *Clin. Vaccine Immunol.* 15:208–214. doi:10.1128/CVI.00374-07

Rubin, C. J., H. J. Megens, A. Martinez Barrio, K. Maqbool, S. Sayyab, D. Schwochow, C. Wang, O. Carlborg, P. Jern, C. B. Jorgensen, et al. 2012. Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. USA.* 109:19529–19536. doi:10.1073/pnas.1217149109

Souza, C., C. MacDougall, B. Campbell, A. McNeilly, and D. Baird. 2001. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. *J. Endocrinol.* 169:R1–R6.

Sulayman, A., M. Tursun, Y. Sulaiman, X. Huang, K. Tian, Y. Tian, X. Xu, X. Fu, A. Mamat, and H. Tulafu. 2018. Association analysis of polymorphisms in six keratin genes with wool traits in sheep. *Asian-Australas. J. Anim. Sci.* 31:775–783. doi:10.5713/ajas.17.0349

Suratannon, N., R. T. A. van Wijck, L. Broer, L. Xue, J. B. J. van Meurs, B. H. Barendregt, M. van der Burg, W. A. Dik, P. Chatchatee, A. W. Langerak, et al; C. South East Asia Primary Immunodeficiencies. 2020. Rapid low-cost microarray-based genotyping for genetic screening in primary immunodeficiency. *Front. Immunol.* 11:614. doi:10.3389/fimmu.2020.00614

Våge, D. I., and I. A. Boman. 2010. A nonsense mutation in the beta-carotene oxygenase 2 (BCO2) gene is tightly associated with accumulation of carotenoids in adipose tissue in sheep (Ovis aries). *BMC Genet.* 11:10. doi:10.1186/1471-2156-11-10

Wang, K., M. Li, and H. Hakonarson. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164–e164. doi:10.1093/nar/gkq603

Xu, S. -S., X. Ren, G. -L. Yang, X. -L. Xie, Y. -X. Zhao, M. Zhang, Z. -Q. Shen, Y. -L. Ren, L. Gao, M. Shen, et al. 2017. Genome-wide association analysis identifies the genetic basis of fat deposition in the tails of sheep (Ovis aries). *Anim. Genet.* 48:560–569. doi:10.1111/age.12572

Xu, N. Y., W. Si, M. Li, M. Gong, J. M. Lariviere, H. A. Nanaei, P. P. Bian, Y. Jiang, and X. Zhao. 2021. Genome-wide scan for selective footprints and genes related to cold tolerance in Chantecler chickens. *Zool. Res.* 42:710–720. doi:10.24272/j.issn.2095-8137.2021.189

Zhang, L., J. Liang, W. Luo, X. Liu, H. Yan, K. Zhao, H. Shi, Y. Zhang, L. Wang, and L. Wang. 2014. Genome-wide scan reveals LEMD3 and WIF1 on SSC5 as the candidates for porcine ear size. *PLoS One* 9:e102085. doi:10.1371/journal.pone.0102085

Zheng, Z., X. Wang, M. Li, Y. Li, Z. Yang, X. Wang, X. Pan, M. Gong, Y. Zhang, Y. Guo, et al. 2020. The origin of domestication genes in goats. *Sci. Adv.* 6:eaaz5216. doi:10.1126/sciadv.aaz5216

Zhou, D. 2019. Design of an HD SNP chip for sheep named GGP OvineHD. *Master*, Northwest A & F University.

Zhou, X., and M. Stephens. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44:821–824. doi:10.1038/ng.2310