Identification of novel genes associated with litter size of indigenous sheep population in Xinjiang, China by SLAF-seq technology

Haiyu Ma [1], **Chao Fang** [2], **Lingling Liu** [1], **Qiong Wang** [1], **Jueken Aniwashi** [1], **Yiming Sulaiman** [1], **Kezierkailedi Abudilaheman** [3], **Wujun Liu** [Corresp. 1]

[1] College of Animal Science, Xinjiang Agriculture University, Urumqi, Xinjiang, China

[2] Department of veterinary managment of animal resources,Faculty of Veterinary Medicine, University of Liège, Liège, Belgium

[3] People's Congress of Xinjiang Uygur Autonomous Region, Urumqi, Xinjiang, China

Corresponding Author: Wujun Liu
Email address: lwj_ws@163.com

Background. Xinjiang in China has a diverse ecological system. Abundant sheep breed resources exist, and among them, several high litter size sheep populations have been found. Previous studies have confirmed that the high prolificacy major gene cannot be used for the detection of high litter size. Our scientific research team found a resource group in Pishan County, southern Xinjiang, with a multi-breast rate of more than 60%. It showed high fertility, with an average litter size of 2–4 lambs in one birth, excellent breast development, and a high survival rate of lambs. In the present study, we considered it an ideal sample for studying the genetic mechanisms of high prolificacy in sheep.

Methods. Indigenous sheep populations from Xinjiang, with different litter size, were selected for the research, and specific-locus amplified fragment sequencing (SLAF-seq) technology was used to comprehensively screen single nucleotide polymorphisms (SNPs) from the whole genome that may cause differences in litter size . Novel genes associated with litter size of sheep were detected by genome-wide association studies, providing new clues for revealing the regulation mechanism of sheep fecundity. Candidate genes related to ovulation and litter size were selected for verification by Kompetitive Allele Specific PCR (KASP) cluster analysis.

Results. A total of 685,300 SNPs were identified by SLAF-seq technique for subsequent genome-wide analysis. In total, 155 SNPs were detected at the genome-wide level. Fourteen genes related to sheep reproduction were notated: *COIL*, *SLK*, *FSHR*, *Plxna3*, *Ddx24*, *CXCL12*, *Pla2g7*, *ATP5F1A*, *KERA*, *GUCY1A1*, *LOC101107541*, *LOC101107119*, *LOC101107809* and *BRAF*. Based on literature reports, 30 loci of 7 genes and candidate genes (*CXCL12*, *FSHR*, *SLK*, *GUCY1A1*, *COIL*, *LOC101107541*, *LOC101107119*) related to ovulation and litter size were selected for verification by KASP cluster analysis. Among them, 9 loci of 3 genes were successfully genotyped. Three loci of *FSHR* (GenBank ID: 443299, g. 75320741G>A site), GUCY1A1 (GenBank ID: 101110000, g. 43266624C>T site) and *COIL* (GenBank ID: 101123134, g. 7321466C>G site) were found to be significantly or extremely significantly associated with litter size. These three loci are expected to be used as molecular markers to determine differences in litter size in sheep.

Identification of novel genes associated with litter size of indigenous sheep population in Xinjiang, China by SLAF-seq technology

Haiyu Ma[1], Chao Fang[2], Lingling Liu[1], Qiong Wang[1], Jueken Aniwashi[1], Yiming Sulaiman[1], Kezierkailedi Abudilaheman[3], Wujun Liu[1]

[1] College of Animal Science, Xinjiang Agricultural University, Urumqi, Xinjiang, China
[2] Department of veterinary managment of animal resources,
Faculty of Veterinary Medicine, University of Liège, Liège, Belgium
[3] People's Congress of Xinjiang Uygur Autonomous Region, Urumqi, Xinjiang, China

Corresponding Author:
Wujun Liu[1]
Email address: Lwj_ws@163.com

Abstract

Background. Xinjiang in China has a diverse ecological system. Abundant sheep breed resources exist, and among them, several high litter size sheep populations have been found. Previous studies have confirmed that the high prolificacy major gene cannot be used for the detection of high litter size. Our scientific research team found a resource group in Pishan County, southern Xinjiang, with a multi-breast rate of more than 60%. It showed high fertility, with an average litter size of 2–4 lambs in one birth, excellent breast development, and a high survival rate of lambs. In the present study, we considered it an ideal sample for studying the genetic mechanisms of high prolificacy in sheep.

Methods. Indigenous sheep populations from Xinjiang, with different litter size, were selected for the research, and specific-locus amplified fragment sequencing (SLAF-seq) technology was used to comprehensively screen single nucleotide polymorphisms (SNPs) from the whole genome that may cause differences in litter size. Novel genes associated with litter size of sheep were detected by genome-wide association studies, providing new clues for revealing the regulation mechanism of sheep fecundity. Candidate genes related to ovulation and litter size were selected for verification by Kompetitive Allele Specific PCR (KASP) cluster analysis.

Results. A total of 685,300 SNPs were identified by SLAF-seq technique for subsequent genome-wide analysis. In total, 155 SNPs were detected at the genome-wide level. Fourteen genes related to sheep reproduction were notated: *COIL, SLK, FSHR, Plxna3, Ddx24, CXCL12, Pla2g7, ATP5F1A, KERA, GUCY1A1, LOC101107541, LOC101107119, LOC101107809* and *BRAF*. Based on literature reports, 30 loci of 7 genes and candidate genes (*CXCL12, FSHR, SLK, GUCY1A1, COIL, LOC101107541, LOC101107119*) related to ovulation and litter size were selected for verification by KASP cluster analysis. Among them, 9 loci of 3 genes were successfully genotyped. Three loci of *FSHR* (GenBank ID: 443299, g. 75320741G>A site), GUCY1A1 (GenBank ID: 101110000, g. 43266624C>T site) and *COIL* (GenBank ID: 101123134, g. 7321466C>G site) were found to be significantly or extremely significantly associated with litter size. These three loci are expected to be used as molecular markers to determine differences in litter size in sheep.

Introduction

Fertility is one of the most important economic traits in sheep. Sheep populations with high reproductive performance show 2 to 3 times higher production efficiency and economic benefit from lambs compared with those with low reproductive performance. Therefore, detection of molecular markers of sheep's high fecundity is of great significance in revealing the genetic basis of sheep reproductive traits, improving sheep breeds by molecular breeding, and establishing core groups or breeding new lines. Therefore, reproductive traits have become a research hotspot in sheep breeding.

Among the sheep breeds worldwide, few sheep breeds exhibit high litter size, early sexual maturity, and perennial oestrus. At present, few major genes affecting high fecundity have been found in Cambridge and Belclare (Hanrahan et al., 2004), Icelandic (Eiriksson JH et al., 2017), Romanov (Deniskova et al., 2017), Finnish (Mullen & Hanrahan 2014) and other high fecundity sheep breeds abroad, whereas domestic studies mostly focus on Hu (Yue, 1996) and Small Tail Han sheep (He et al., 2012). The results show that different sheep breeds possess different major genes affecting their litter size. As a result, the major genes identified from small-tailed Han sheep are selected as markers for domestic mutation detection for sheep litter size, which results in pseudoscience.

Xinjiang has a diverse ecological system with abundant sheep resources, and among them, several high litter size sheep populations are found. Previous studies have confirmed the existence of the high prolificacy major gene *FecB* mutation in the Cele Black Sheep (Jiang et al., 2017) and Duoliang sheep populations (Wang et al., 2017), but it cannot be used for the

detection of high litter size. Our scientific research team found a resource group in Pishan County, southern Xinjiang, for the indigenous prolificacy sheep breed in Xinjiang. Mainly due to natural mating, the ewes are oestrus all year, and, therefore, are bred throughout the year with a multi-breast rate of more than 60%. It showed high fertility, with an average litter size of 2–4 lambs in one birth, excellent breast development, and a high survival rate of lambs. Bashbai sheep are single-breasted, seasonally oestrus sheep found in the Tacheng area of Xinjiang. Breeding is mainly via artificial insemination, and the oestrus period is typically in November. The lambing rate is 103%, the milk yield is stable, and the lamb survival rate is 98%. They are very different varieties in the number of lambs. Herein, we considered it as an ideal sample for studying the genetic mechanisms of high prolificacy in sheep.

In this study, ewes with different litter size (one to four) in Xinjiang were selected as research materials. Genome-wide association studies (GWAS) based on specific-locus amplified fragment sequencing (SLAF-seq) technique were used to identify single nucleotide polymorphisms (SNPs) that might cause differences in litter size.

Materials & Methods

Sample Collection and DNA Extraction

A total of 126 sheep from two populations were used in this study, including 62 Hetian sheep from Pishan county of Hetian city of Xinjiang (East longitude 77°31′~ 79°38′, latitude 35°22′~ 39°01′), and 64 Bashbay sheep from Yumin county of Tacheng state of Xinjiang (East longitude 82°12′～83°30′ , latitude 45°24′～46°3′). Whole blood samples (6 mL) were collected from the jugular veins of sheep and transferred into EDTA anticoagulant tubes. Genomic DNA was extracted by standard phenol chloroform extraction method (Köchl et al., 2005) for subsequent experiments. The study design was approved by the appropriate ethics review board. The University of Xinjiang Agricultural University approved the use of its facilities for the study (Animal protocol number: 2017010).

Construction of SLAF-seq Library and High Throughput Sequencing

The current sheep genome (Oar_v4.0) was selected as the reference genome to simulate the restriction enzyme digestion and identify the expected SLAF yield. Hpy166II + EcoRV-HF enzyme was selected for enzymatic digestion (Davey JW et al., 2013). At the same time, in order to evaluate the accuracy of the digestion strategy, *Oryza sativa indica* (http://rapdb. dna. affrc. go. JP/) was selected to evaluate the digestion efficiency (Li et al., 2009b). Subsequently, 126 individuals were subjected to SLAF library construction (Kozich et al., 2013) and sequencing

after quality inspection. In order to evaluate the accuracy of the enzyme test, Nippon Sunshine was selected as the control for sequencing.

Analysis of SLAF-seq data and identification of SNP loci

According to SLAF tags, SNP locus information was analysed and screening criteria was set at minor allele frequency (MAF)>0.05. The development of SNP markers was based on the sheep reference genome, using BWA (Li & Durbin, 2009) to compare the sequenced reads to the reference genome, and GATK (Mckenna et al., 2010) and SAMtools (Li et al., 2009a) to SNP calling. The intersection of SNP markers obtained by the two methods was used as the final reliable data set of SNP markers. Sequencing reads of the control were compared with their reference genomes by SOAP software. The double-end contrast efficiency was 92.46%, and the contrast efficiency was normal. The enzymatic cleavage efficiency of the control was 93.60%, indicating that the enzymatic cleavage reaction was normal.

Population structure

Principal component analysis (PCA) was performed using PLINK1.9, and the ggplot2 package in R (v3.4.4) was used to generate the PCA figure(Wickham, 2015). We removed the SNPs in linkage disequilibrium in PLINK 1.9 with the command(--indep-pairwise 50 5 0.2).

GWAS

Based on SNP analyses, the general linear model (GLM) and mixed linear model (MLM) of the TASSEL (Bradbury et al., 2007) software (http://www.maizegenetics.net) were used to obtain the correlation values. The Q matrix of the sample population structure was calculated by Admixture software, and the K matrix of the relationship between samples was calculated by SPAGeDi software. The general linear model uses the group structure information, while the mixed linear model uses the information of the population structure and the kinship relationship, the fixed effects of the MLM model are parity and population, and the random effects are related, and finally each SNP site gets an associated value. Manhattan chart analysis and Q-Q Plot were used to analyse the population structure. Manhattan Diagram and QQ-Plot Diagram were drawn using R language (Zhiwu et al., 2010).

Bioinformatics analyses

We combined several commonly used bioinformatics databases, such as NCBI (http://www.ncbi.nlm.nih.gov/), UCSC (http://genome.ucsc.deu/), and Ensembl (http://asia.ensembl.org/index.html) to locate significant SNPs from these databases.
The position of the significant SNPs was located, and for the SNP markers that were not within the gene, candidate genes 500 kb upstream and downstream of the significant SNP site were searched in order to determine the linkage disequilibrium between the markers. Candidate gene functions were identified and analysed by online gene enrichment software DAVID 6.7

(https://david.ncifcrf.gov/home.jsp). Venny2.1.0 (http://bioinfogp.cnb.csic.es/tools/venny/index.html) was used to draw a Venn diagram of the overlap sites of the two methods.

KASP typing verification

KASP typing was performed at 30 loci of the seven candidate differential genes (*FSHR, COIL, GUCY1A1, CXCL12, SLK, LOC101107541, LOC101107119*), which were annotated. Nine pairs of primers with successful typing are listed in Table 1. One-way ANOVA and *t*-test in SPSS 19.0 were used to analyse the association between litter size and mutant locus genotypes.

Results

Establishment of database and sequencing evaluation

In order to obtain the actual SLAF tags used in this study, 62 Hetian sheep and 64 Bashbai sheep were subjected to SLAF-seq using the same enzyme combinations as those used in computer restriction analysis. A total of 854.88 Mb reads data were obtained for all individuals, the average Q30 and GC contents were 91.74% and 42.14%, respectively, indicating that the SLAF-seq process was normal and available. After genome comparison and SNP mining, 5,218,278 population SNPs were found using all individuals. The completeness was 0.5 times, and the genomic frequency was 0.05 filtered to 685,300 SNPs sites, and 685,300 SNPs were identified for subsequent analysis.

Population stratification assessment

Results of the PCA are shown in Fig. 1. PC1 had 3.02% variance, and PC2 had 2.15% variance. The results showed that Hetian and Bashbay sheep were separated by PC1 and there was no mixing between the two populations, which provided a foundation for the subsequent GWAS.

GWAS

In this study, GLM and MLM were used to analyse the genome-wide association of litter size traits in Hetian and Bashbai sheep. A Bonferroni correction of α=0.1% was applied for genome-wise thresholds (significance threshold = $-\log_{10}(\alpha/\text{number of independent SNPs})$) . The SNPs with p-values below 1.459e-07 (= 0.1/685300) were considered to be significantly associated with the phenotype. GWAS results showed that 111 and 44 SNPs were significantly associated with the genome in the GLM and MLM, respectively, and 155 significant loci were identified. Manhattan charts of litter size traits are shown in Fig. 2 and 3. A total of 25 SNPs were detected by both methods (Fig. 4). According to the QQ-plot diagram (Fig. 5 and 6), it can be seen that there is a large deviation of the SNP point, and it is considered that the deviation of the

observation value of this SNP site is caused by the genetic action produced by this SNP mutation.

Gene function annotation

Using Oar_v4.0 sequence information of sheep genome and common database information such as NCBI, SNP loci with significant GWAS results were analysed and annotated. A total of 133 genes were annotated in the two models, including *FRS2, RGS3, MDH1, IMPA1* and *KCNE3*, which are involved in the differentiation and survival of nerve cells. Some of them are new genes that have not been clearly labelled, and their functions need to be further studied; 14 of them are related to reproduction. *COIL, SLK, Plxna3* and *Ddx24* genes affect the development of ovaries and follicles in sheep; *CXCL12, Pla2g7, ATP5F1A,* and *KERA* genes affect the early development of the placenta and placental membrane; *GUCY1A1, LOC101107541, LOC101107119,* and *LOC101107809* participate in the oxytocin signalling pathway and indirectly regulate the production of ovarian steroids; *BRAF* and *Ddx24* are the most expressed in the uterus.

Enrichment analysis results

Genecards and DAVID6.7 online websites were used to analyse the functions of the candidate genes, and Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) online websites were used to search for candidate genes. Enrichment analysis revealed that these differential genes are mainly involved in oxytocin signalling pathway, amino acid biosynthesis, neurotrophic signalling pathway, pentose phosphate pathway, Wnt signalling pathway, etc. In addition to the *FSHR, SLK, GUCY1A1* and *LOC1001107541* genes, a few other genes are involved in the oxytocin signalling pathway, which regulates the litter size of sheep by regulating follicle stimulating hormone receptors, and ultimately forms a multi-foetal sheep population.

Functional validation of genes related to litter size regulation

Based on the results of previous studies and comprehensive analysis of gene function, 30 missense mutation sites of 7 genes (*CXCL12, FSHR, SLK, GUCY1A1, Coil, LOC101107541, LOC101107119*) closely related to ovulation and litter size were screened, and 126 individuals with different litter size were typed by KASP. Nine sites of the final three genes (*FSHR, GUCY1A1, COIL*) were successfully typed. Only one genotype was identified at 21 loci of the remaining four genes, which could not be used to determine the genotype.

Association analysis between different genotypes and litter size

The association analysis between genotypes with successful typing and individuals with different litter size was carried out (using SPSS 19.0 software). The results showed that there were significant differences or significant correlations in litter size among different genotypes of

FSHR (g.75320741G>A), *GUCY1A1* (g.43266624C>T) and *COIL* genes (g.7321466C>G) (see Fig.7-9).

At the g.75320741 locus of *FSHR* gene, the average litter size of G/G genotype individuals was significantly higher than that of A/A genotype individuals (P = 0.004). At g.43266624 locus of GUCY1A1 gene, the average litter size of C/C genotype individuals was significantly higher than that of T/T genotype individuals (P = 0.038). At g.7321466 locus of COIL gene, the average litter size of C/C genotype individuals was significantly higher than that of T/T genotype individuals (P = 0.042). (see Fig.10).

Discussion

Evaluation of the reliability of SLAF-seq Technology

In this study, SLAF-seq method was used to identify SNPs located in the genome of Chinese indigenous sheep populations. More than 685,300 SNPs were detected. In addition, 133 genes were annotated by comparing and analysing the loci of SNPs with significant GWAS results for reproductive traits.

SLAF-seq is a simplified deep genome sequencing technique, which can be used for obtaining a large number of molecular markers and accurately typing by means of bioinformatics. This technology has been successfully implemented in many species, such as cotton (Li et al., 2017) (IF: 7.44) and soybean (Han et al., 2016) (IF: 7.33). In this study, a total of 1192168 SLAF tags from sheep reference genome were predicted by using the Hpy166II + EcoRV-HF® enzyme combination. The average sequencing depth was 13.06x, and 685,300 SNPs were found. Therefore, the SLAF-seq method used in this experiment provides comprehensive genomic variation information. Considering the enormous differences in genome sequences between the European and Chinese sheep breeds, the experimental data show that SLAF-seq is a powerful method and has great potential to study more breeds. Therefore, SLAF-seq method can be considered as a highly efficient option for sheep genome research.

Molecular markers and candidate genes for litter size in sheep

Genetic markers for the number of lambs in sheep have been studied, but the reports are limited. They are only concerns about ovulation rate, low attention to litter size. Current studies have shown that *GDF9, BMP15*, and *BMPR-IB* genes, and 13 mutation loci (*FecBB, FecXB, FecXG, FecXGR, FecXH, FecXI, FecXL, FecXO, FecXR, FecGH, FecGT, V371M*) (Mullen & Hanrahan, 2014) are the major genes affecting litter size or ovulation rate in sheep. However, these markers are not stable in other sheep breeds and are not the major genes affecting their litter size, indicating that there are other genes responsible for litter size and that there are interspecies differences. The high fecundity of sheep breeds is very rare all over the world.

Xinjiang has very rich sheep resources. The Hetian sheep breeds used in this experiment often produce many lambs, and it is rare to be able to use them as an experimental material.

In this study, 30 loci of 7 ovulation-related genes (*FSHR, COIL, GUCY1A1, CXCL12, SLK, LOC101107541, LOC101107119*) were selected for KASP validation. Results showed that nine out of 30 loci from three genes (*CXCL12, FSHR,* and *COIL*) were successfully genotyped. The other four genes are seldom reported in livestock and are mostly studied in human and mouse. Among them, the *SLK* gene affects the development of the ovary and follicle (An, 2012); *CXCL12* gene affects the early development of the placenta and placental membrane (Quinn et al., 2016; Sanchez et al., 2017); *LOC101107541* and *LOC101107119* participate in the oxytocin signalling pathway and indirectly regulate the production of ovarian steroids. The results need to be further verified.

The relationship between *FSHR* gene and sheep reproduction

This study concluded that the *FSHR* gene may be one of the important genes affecting litter size of Hetian sheep. Follicular stimulating hormone receptor (FSHR) is a member of the glycoprotein family of G protein-coupled receptor superfamily and plays an important role in follicular development in animals. FSHR is mainly expressed in granulosa cells of follicles in super ovulated and normal lambs, and there are positive signals in primordial follicles. No positive signals are observed in primordial follicles in normal adult sheep, and the expression of FSHR is decreased in large dominant follicles (Chu et al., 2012). Some studies have reported that the full-length coding sequences of six FSHR splicing forms in sheep have been obtained. The open reading frames are 695aa, 694aa, 648aa, 633aa, 595aa, and 533aa. 533aa is not found in the ovaries of lambs, whereas 694aa and 648aa are not found in adult sheep (Jiang, 2014). In this study, FSHR was found to be associated with follicular development.

Chu et al. detected 50 SNPs in the regulatory region of the *FSHR* gene in two high-reproductive sheep breeds (small-tailed Han and Hu sheep) and two low-reproductive sheep breeds (Kaolidai and Chinese Merino sheep). Four mutations in the sheep *FSHR* gene were detected by PCR-single-strand conformation polymorphism (PCR-SSCP) technology, suggesting that *FSHR* gene may significantly affect the litter size of sheep (Chu et al, 2012). Pan et al. cloned 50 flanking regions of the sheep *FSHR* gene and analysed its genomic structure. RT-qPCR showed that *FSHR* was widely expressed in the tissues detected by sheep. In addition, the homologous mutation of the *FSHR* gene was found to be significantly correlated with litter size ($P < 0.1$; Pan et al., 2014). Some of them observations are consistent with previous studies.

In this study, it was found that the 75320741 sites of the *FSHR* gene was successfully classified into three genotypes, namely GG, GA, and AA, in the breeds of Xinjiang Tianyang

and Bashbai sheep. Correlation analysis between the different genotypes and litter size showed that the GG genotype was highly correlated with litter size, which further verified the reliability of SLAF-seq and verified that *FSHR* gene is a candidate gene that affects the litter size in sheep.

Relationship between Coilin Gene and Sheep Reproduction

Coilin is a characteristic protein of Cajal body. Cajal body is a conserved nuclear organelle, which participates in many aspects of small ribonucleoprotein (RNP) biogenesis (Gall et al., 1999). Coilin is necessary for Cajal bodies to form and recruit spliced nucleoproteins (snRNPs) to modify protein complexes that guide RNA and motor neurons (SMN). When coilin is deleted, the remaining Cajal bodies lose contact with SMN complexes. Because SMN is considered a necessary condition for snRNPs assembly and internal circulation, the lack of interaction between SMN and coilin in the nucleus may lead to a decline in RNP assembly capacity, which may have downstream effects on development and gametes, thus affecting litter size. Michael et al. studied the effect of coilin removal on the overall viability and reproductive success of mice. The results showed that the number of oocytes that could fertilize was significantly reduced after coilin knockout, and the number of litters per foetus was low, showing obvious fertility and reproductive deficiencies, (Walker et al., 2009). Coilin gene was found to be closely related to litter size. In this study, the association between genotype and litter size revealed that CC genotype was significantly associated with litter size, and it was verified that the *COIL* gene is one of the candidate genes affecting the litter size in sheep.

The relationship between *GUCY1A1* gene and sheep reproduction

Guanylic acid cyclase is an enzyme that converts guanylic-5'-triphosphate into cycloguangan-3', 5'-monophosphate (Pyriochou & Papapetropoulos, 2005). As a membrane-binding molecule, guanylic acid cyclase exists in the form of membrane-binding and cytoplasm. Soluble guanylate cyclase (sGCX), an isodimer enzyme composed of alpha and beta subunits, is the only receptor of nitric oxide (NO) in biological messenger identified so far and is closely involved in various signal transduction pathways (Peter et al., 2010). Soluble guanylic cyclase collector (GUCY1A1) was found to be involved in hormone regulation, oxytocin signalling pathway, and reproductive capacity. In this study, we found that the mutation site of the *GUCY1A1* gene (g. 43266624C>T) was significantly correlated with litter size of Hetian sheep, thus verifying the correctness of the sequencing results.

The results of KASP typing showed that nine of the three genes (*FSHR, GUCY1A1, COIL*) were successfully typed. The genotypes of mutation loci were correlated with high fecundity and average litter size of field sheep. The results showed that the genotypes of *FSHR* (g.75320741G>A) were extremely significantly correlated with high fecundity and average litter

size of field sheep. *GUCY1A1* (g.43266624C>T), and *COIL* (g.7321466C>G) genes were significantly correlated with high fecundity and average litter size of field sheep. The results of this study confirm that mutations in these three genes cause changes in litter size of sheep and, in addition, lead to validation of the SLAF sequencing technology.

Molecular marker screening is a major step in molecular breeding. SNP markers are third-generation molecular markers and are currently the mainstream molecular markers. KASP genotypes SNPs and indels at specific sites for precise genetic determination. Compared to other detection methods, KASP has higher throughput, is faster, more cost-effective, and produces more accurate results.

Breeding sheep with high fertility is important for mutton breeding. Molecular marker screening is used to improve early molecular selection. In this study, the KASP technology was used to verify three sites with significant or extremely significant effects on litter size, to make breakthroughs in the development of sheep multi-lamb genes and markers and to accelerate the efficiency of molecular breeding of sheep with high fecundity. It can be used to improve other low-community groups and improve the efficiency of meat production in sheep. The study showed that SLAF-seq and KASP are cost-effective tools for selecting the desirable genotypes in sheep breeding programs.

Conclusions

In this study, DNA from the Chinese indigenous sheep from southern Xinjiang with different litter size fecundity were sequenced by the SLAF-seq technique. GWAS of SNPs that may cause differences in litter size was carried out. In total, 155 genes with significant mutations were obtained by SNP loci, gene annotation, and pathway analysis, of which 17 genes were related to reproductive traits. Seven candidate genes closely related to litter size and ovulation were obtained through literature search and comprehensive analysis. KASP technique was used to verify the role of the seven genes. Among the seven selected genes, only *FSHR* (GenBank ID: 443299, g. 75320741G>A site), *GUCY1A1* (GenBank ID: 101110000, g. 43266624C>T site), and *COIL* (GenID: 101123134, g. 7321466C>G site) loci are significant among different genotypes as molecular markers to detect the differences in litter size of sheep population from Xinjiang. This finding provides new clues for explaining the regulatory mechanism underlying sheep fecundity and identifying molecular markers for litter size traits. The sub-markers are of great significance for breeding of high-fecundity sheep breeds.

Acknowledgements

We would like to thank all the participants of the study. We would also like to thank Editage [www.editage.cn] for English language editing.

References

An N. 2012. Localization dynamics and preliminary research Function of Ste20-like kinase(*SLK*) during Mouse Oocytes maturation. *Jilin Agricultural University*.

Bradbury P, Zhang Z, Kroon D, Casstevens T, Y, and Buckler E. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633-2635 DOI: 10.1093/bioinformatics/btm308.

Chu MX, Guo XH, Feng CJ, Li Y, Huang DW, Feng T, Cao GL, Fang L, Di R, and Tang QQ. 2012. Polymorphism of 5' regulatory region of ovine FSHR gene and its association with litter size in Small Tail Han sheep. *Molecular Biology Reports* 39:3721-3725 DOI: 10.1007/s11033-011-1147-x.

Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAD Sequencing data: implications for genotyping. Molecular ecology 22: 3151-3164 DOI：10.1111/mec.12084

Deniskova T, Dotsev AV, Selionova M, Wimmers K, Reyer H, Kharzinova VR. 2017. Whole-genome single nucleotide polymorphism study of Romanov sheep. *J Anim Sci* 95:339-340 DOI:10.2527/asasann.2017.696.

Eiriksson JH, Sigurdsson A. 2017. Sources of bias, genetic trend and changes in genetic correlation in carcass and ultrasound traits in the Icelandic sheep population. *Iceland Agric Sci* 30:3-12 DOI: 10.16886/IAS.2017.01.

Gall JG, Bellini M, Wu Z, and Murphy C. 1999. Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Molecular Biology of the Cell* 10:4385 DOI: 10.1091/mbc.10.12.4385.

Han Y, Zhao X, Liu D, Li Y, Lightfoot DA, Yang Z, Zhao L, Zhou G, Wang Z, and Huang L. 2016. Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist* 209:871-884 DOI: 10.1111/nph.13626

Hanrahan JP, Gregan SM, Philippe M, Michael M, Davis GH, Richard P, and Galloway SM. 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). *Biology of Reproduction* 70:900-909 DOI: 10.1095/biolreprod.103.023093.

He JN, Zhang BY, Chu MX, Feng T, Cao GL, Di R, Fang L, Huang DW, Tang QQ, and Li N. 2012. Polymorphism of insulin-like growth factor 1 gene and its association with litter size in Small Tail Han sheep. *Molecular Biology Reports* 39:9801-9807 DOI: 10.1007/s11033-012-1846-y.

Jiang WD, Zhang J, Song XM, Ai DX, Li LR, and Shi CQ. 2017. Extraction of total RNA from the overy of oira black sheep. *Heilongjiang Animal Science and Veterinary Medicine*:62-63.

Jiang XJ. 2014. Expression analysis of AMH、FSHR、LHR in ovine ovarian follicles, *Shihezi University*.

Köchl S, Niederstätter H, and Parson W. 2005. DNA Extraction and Quantitation of Forensic Samples Using the Phenol-Chloroform Method and Real-Time PCR. *Methods in Molecular Biology* 297:13 DOI: 10.1385/1-59259-867-6:013.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, and Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applenvironmicrobiol* 79:5112-5120 DOI: 10.1128/AEM.01043-13.

Li H, and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754-1760 DOI: 10.1093/bioinformatics/btp324.

Li H, Handsaker B, Wysoker A, Fennell T, and Ruan J. 2009a. The Sequence Alignment-Map format and SAMtools. *Bioinformatics* 25:2078-2079 DOI: *10.1093/bioinformatics/btp352.*

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, and Wang J. 2009b. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25:1966-1967 DOI: 10.1093/bioinformatics/btp336.

Li T, Ma X, Li N, Zhou L, Liu Z, Han H, Gui Y, Bao Y, Chen J, and Dai X. 2017. Genome-wide association study discovered candidate genes of Verticillium wilt resistance in upland cotton (Gossypium hirsutum L.). *Plant Biotechnology Journal* 15:1520-1532 DOI: 10.1111/pbi.12734.

Mckenna A, Hanna ME, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, and Depristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297-1303 DOI: 10.1101/gr.107524.110.20.

Mullen MP, and Hanrahan JP. 2014. Direct Evidence on the Contribution of a Missense Mutation in GDF9 to Variation in Ovulation Rate of Finnsheep. *Plos One* 9:e95251 DOI: 10.1371/journal.pone.0095251.

Pan X, Liu S, Li F, Wang W, Li C, Ma Y, and Li T. 2014. Molecular characterization, expression profiles of the ovine FSHR gene and its association with litter size. *Molecular Biology Reports* 41:7749-7754 DOI: 10.1007/s11033-014-3666-8.

Peter O, Shruti S, Albert G, Anthony A, Cynthia H, Johengen MJ, Jong-Hau H, Sohrab F, Black SM, and Fineman JR. 2010. Alterations in cGMP, soluble guanylate cyclase, phosphodiesterase 5, and B-type natriuretic peptide induced by chronic increased pulmonary blood flow in lambs. *Pediatric Pulmonology* 42:1057-1071 DOI: 10.1002/ppul.20696.

Pyriochou, and Papapetropoulos. 2005. Soluble guanylyl cyclase: more secrets revealed. *Cellular Signalling* 17:407-413 DOI: 10.1016/j.cellsig.2004.09.008.

Quinn KE, Reynolds LP, Grazul-Bilska AT, Borowicz PP, and Ashley RL. 2016. Placental development during early pregnancy: Effects of embryo origin on expression of chemokine ligand twelve (CXCL12). *Placenta* 43:77-80 DOI: 10.1016/j.placenta.2016.05.008.

Sanchez NS, Quinn KE, Ashley AK, and Ashley RL. 2017. In the ovine pituitary, CXCR4 is localized in gonadotropes and somatotropes and increases with elevated serum progesterone. *Domest Anim Endocrinol* 62:88-97.

Walker MP, Liping T, and A Gregory M. 2009. Reduced viability, fertility and fecundity in mice lacking the cajal body marker protein, coilin. *Plos One* 62:88-97 DOI: 10.1016/j.domaniend.2017.10.003.

Wang S, Li XL, Niu ZG, and Shi HC. 2017. Whole genomic association analysis of polyembryonic gene in Duolong sheep. *Jiangxi Agricultural Journal* 29:77-81.

Wickham, H., 2015. ggplot2. Wiley Interdisciplinary Reviews Computational Statistics 3(2),180-185.

Yue GH. 1996. Reproductive characteristics of Chinese Hu sheep. *Animal Reproduction Science* 44:220-230 DOI:10.1016/0378-4320(96)01562-x.

Zhiwu Z, Elhan E, Chao-Qiang L, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Jianming Y, Arnett DK, and Ordovas JM. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* 42:355-360 DOI: 10.1038/ng.546.

Table 1(on next page)

Information of Four Genes

1 Table.1 Information of Four Genes

Gene	Position	Primer Sequence
GUCY1A1	43266624	F1: GAAGGTCGGAGTCAACGGATTGGAGTGGGCCAGCAGCTAC
		F2: GAAGGTGACCAAGTTCATGCTGGAGTGGGCCAGCAGCTAT
		R1: GTTCTTGTCAGGGACACCTGG
SLK	23608558	F1:GAAGGTCGGAGTCAACGGATTCTTGCGAGATGAAGCCAAGC
		F2:GAAGGTGACCAAGTTCATGCTCTTGCGAGATGAAGCCAAGT
		R1: ACATTCTGAAATTTGGACAGCTC
COIL	7314134	F1: GAAGGTCGGAGTCAACGGATTCCATGAAAGAACCTGGGAAA
		F2: GAAGGTGACCAAGTTCATGCTCCATGAAAGAACCTGGGAAC
		R1: CCTCAGCTCCATTTTCGTTG
COIL	7321466	F1: GAAGGTCGGAGTCAACGGATTGACTCCGAGGAGGAATCGC
		F2: GAAGGTGACCAAGTTCATGCTGACTCCGAGGAGGAATCGG
		R1: GTGGCATGGTCGTCCGTAC
COIL	7321563	F1:GAAGGTCGGAGTCAACGGATTGCACAGTCTGTGAAAGAGTGGA
		F2: GAAGGTGACCAAGTTCATGCTGCACAGTCTGTGAAAGAGTGGG
		R1: TCTAGCAGGAAGAGCTTTAGGG
FSHR	75132817	F1: GAAGGTCGGAGTCAACGGATTAGCCCAAGCTCAGGAATGC
		F2: GAAGGTGACCAAGTTCATGCTGAGCCCAAGCTCAGGAATGT
		R1: GGTGGATGGATAAGTAAACATGG
FSHR	75320579	F1: GAAGGTCGGAGTCAACGGATTGGACAGGGAAGACTCACTCACA
		F2: GAAGGTGACCAAGTTCATGCTGACAGGGAAGACTCACTCACG
		R1: CTCACCTACCCCAGCCACT
FSHR	75320741	F1: GAAGGTCGGAGTCAACGGATTGATATTTCAAGAACCAGGATCCA
		F2: GAAGGTGACCAAGTTCATGCTATATTTCAAGAACCAGGATCCG
		R1: CAGCTTCTTAAGATTTTCTAAGCC
FSHR	75132820	F1: GAAGGTCGGAGTCAACGGATTATGATGCTGGCAGCATGGT
		F2: GAAGGTGACCAAGTTCATGCTATGATGCTGGCAGCATGGC
		R1: CATCACCCACGCCATGCAG

2

Figure 1

Principal component analysis (PCA) of Hetian(Mulitiple) and Bashbay(Single)sheep breeds

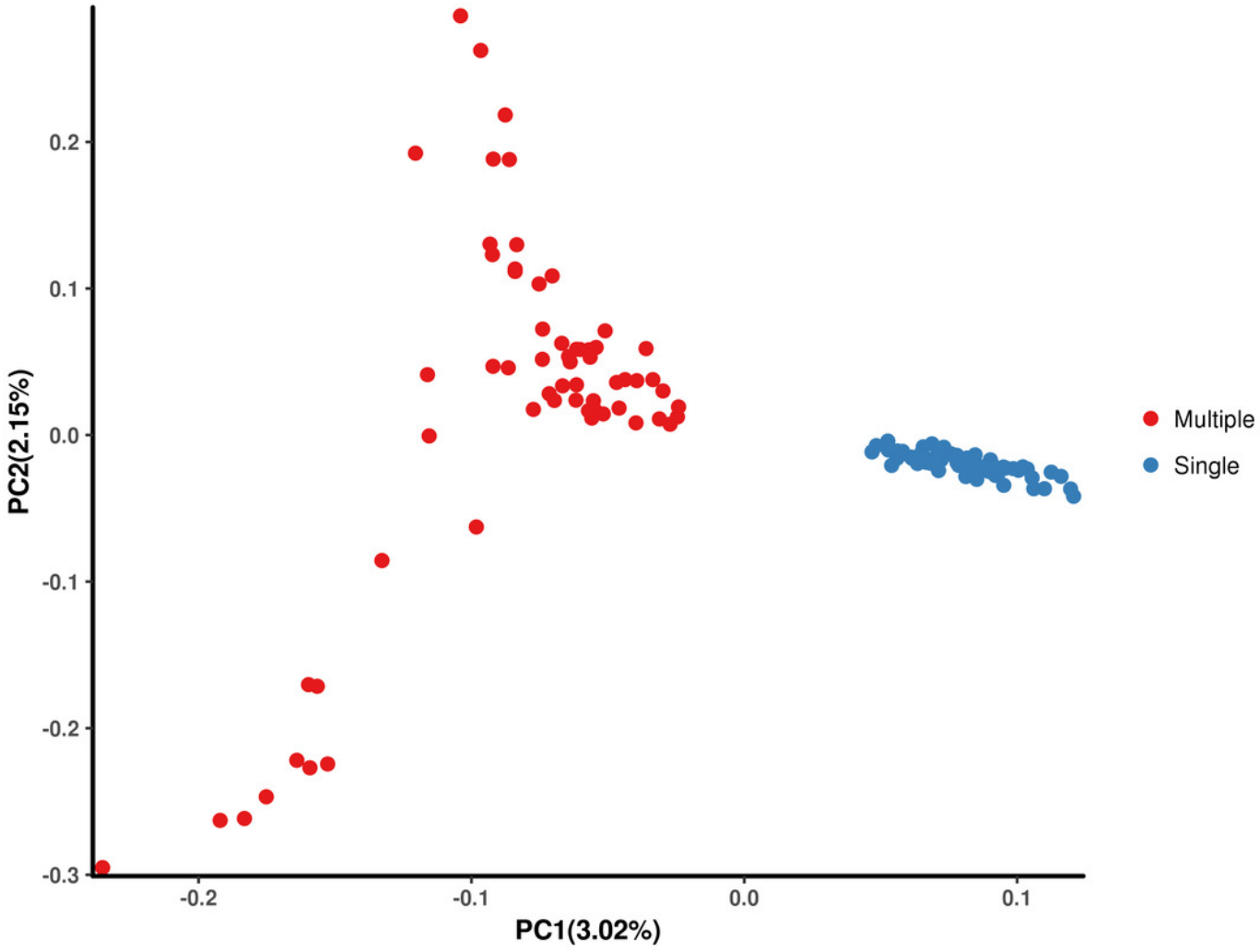

Figure 2

Manhattan plot for genome-wide association study on GLM model

Note: the scale on the X-axis represents ID of chromosomes, NC_019458.2-NC_019483.2represents ID 1-26 of chromosomes, X chromosome is represented by NC_019484.2. The scale on the Y-axis is the –log10 □P-value□score of association analysis. The rad dashed line indicates genome-wide significance of suggestive association.

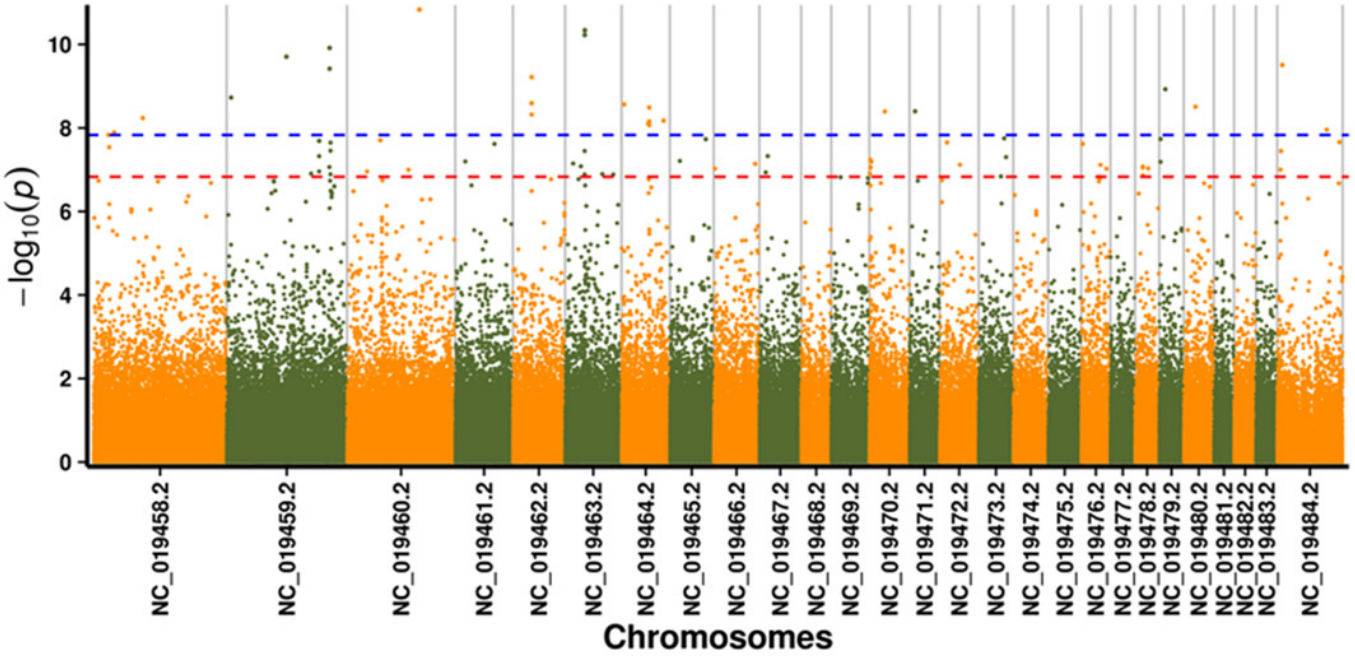

Figure 3

Manhattan plot for genome-wide association study on MLM model

Note: the scale on the X-axis represents ID of chromosomes, NC_019458.2-NC_019483.2represents ID 1-26 of chromosomes, X chromosome is represented by NC_019484.2. The scale on the Y-axis is the –log10 □P-value□score of association analysis. The rad dashed line indicates genome-wide significance of suggestive association.

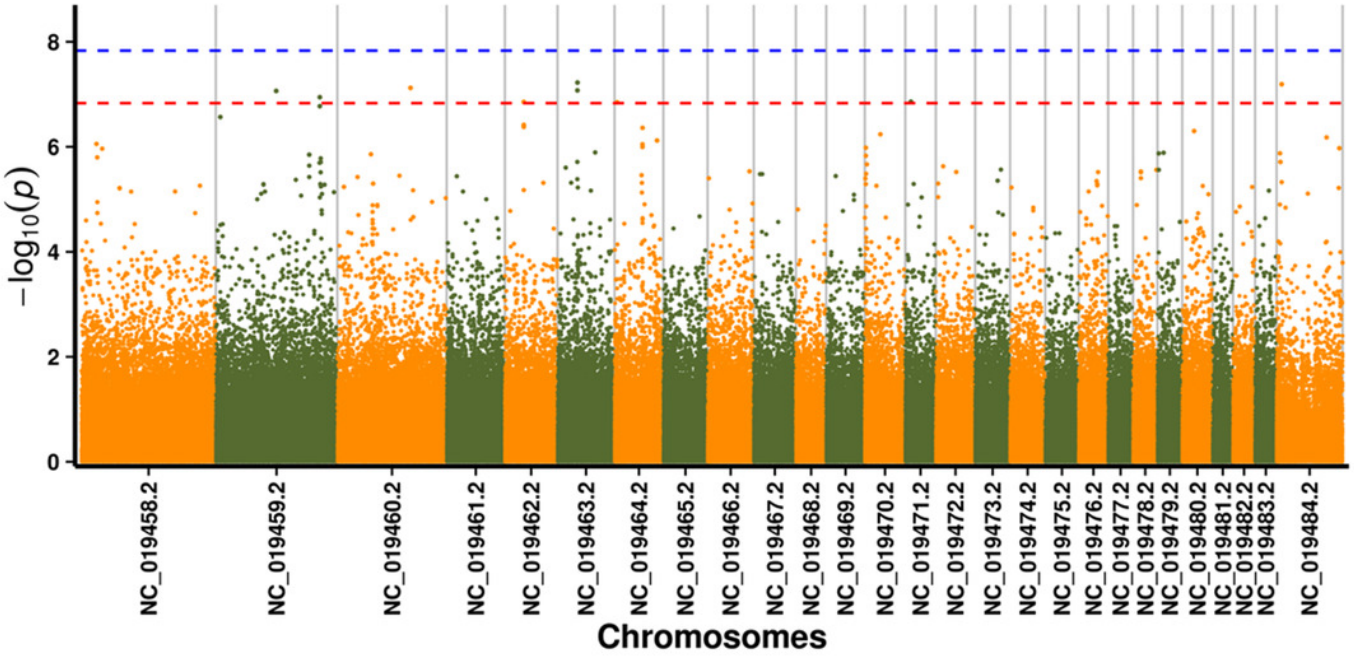

Figure 4

Venn diagram of two methods

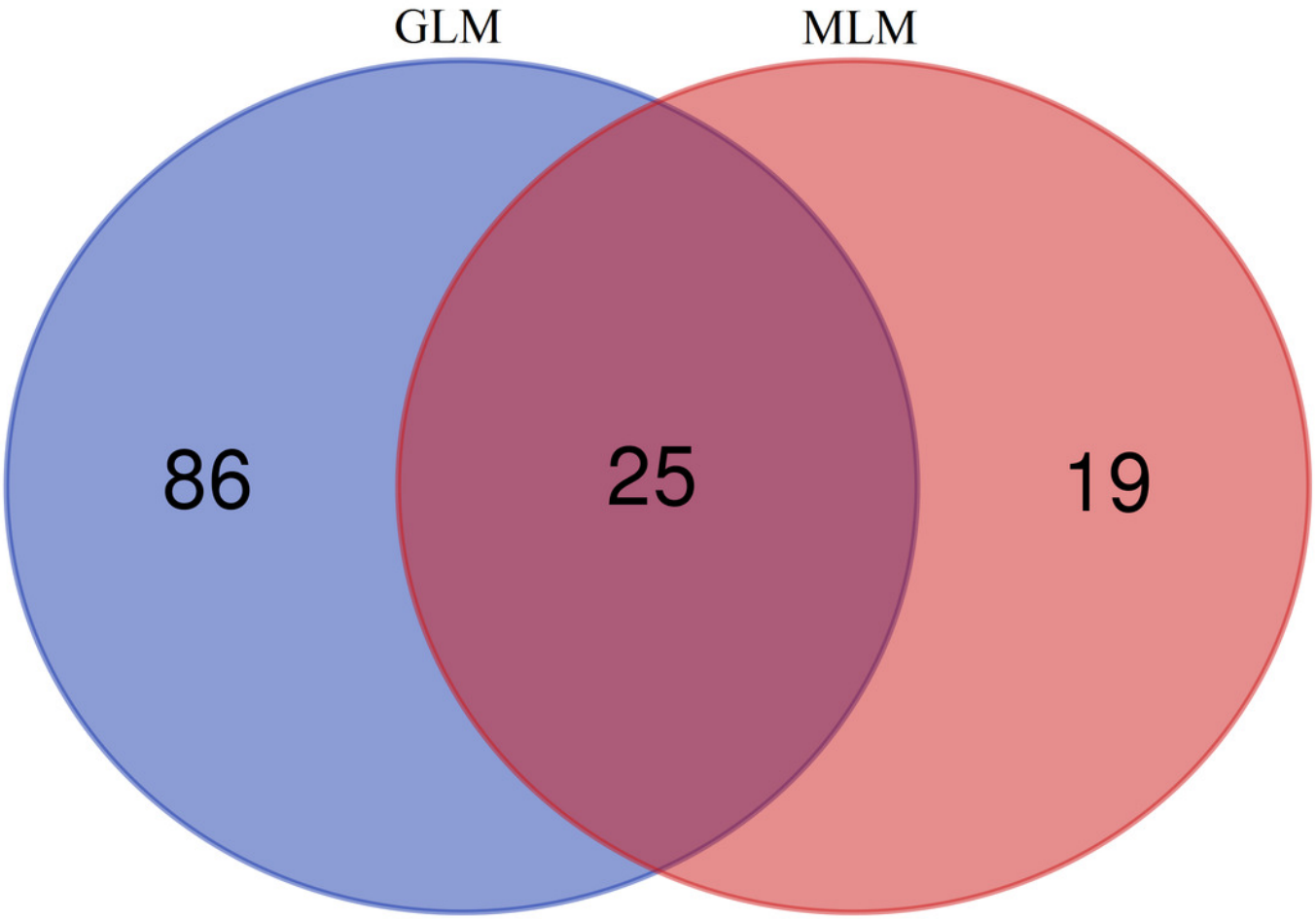

Figure 5

The results of quantile-quantile (Q-Q) plot for Litter size trait in GLM model

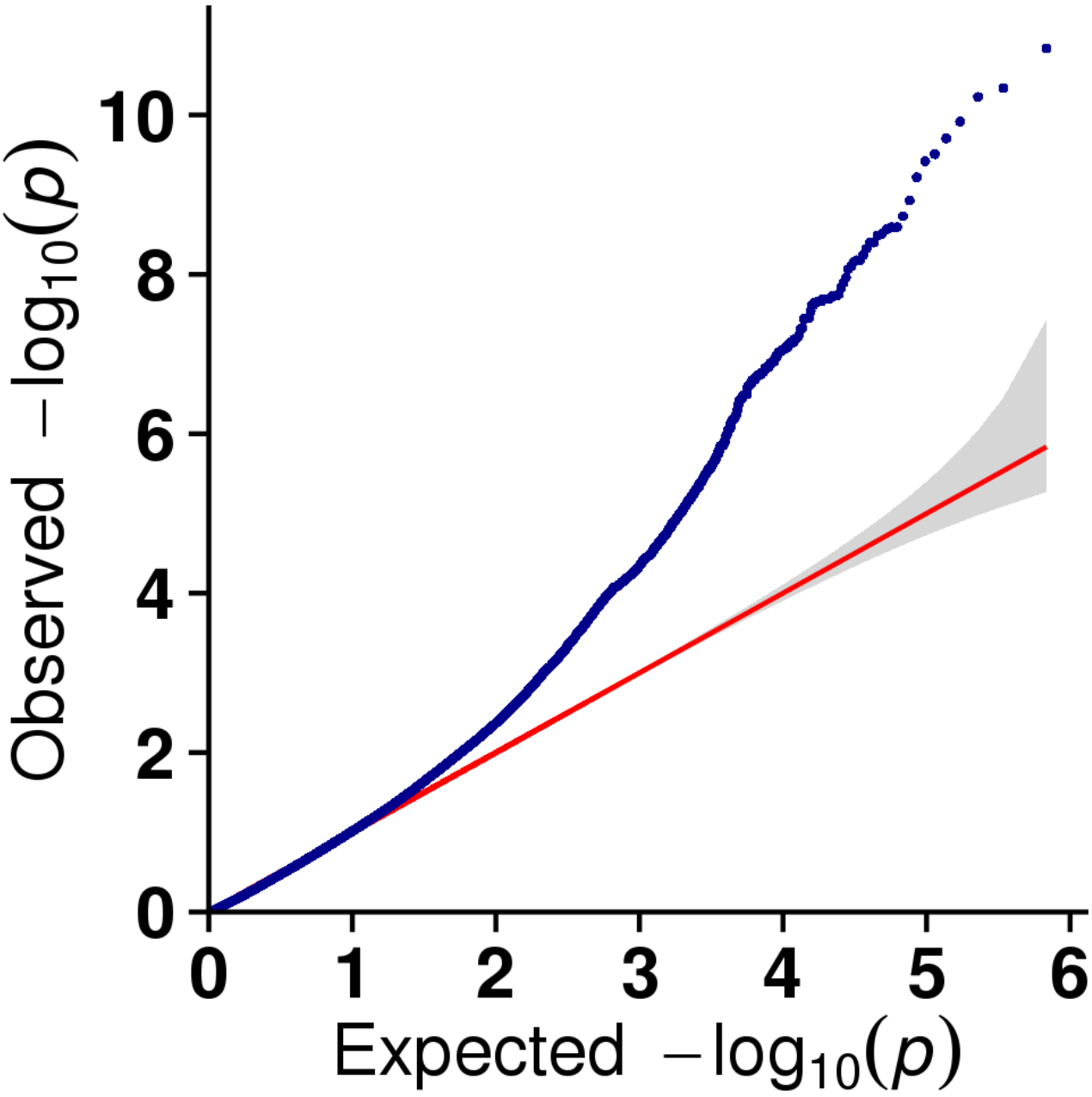

Figure 6

The results of quantile-quantile (Q-Q) plot for Litter size trait in MLM model

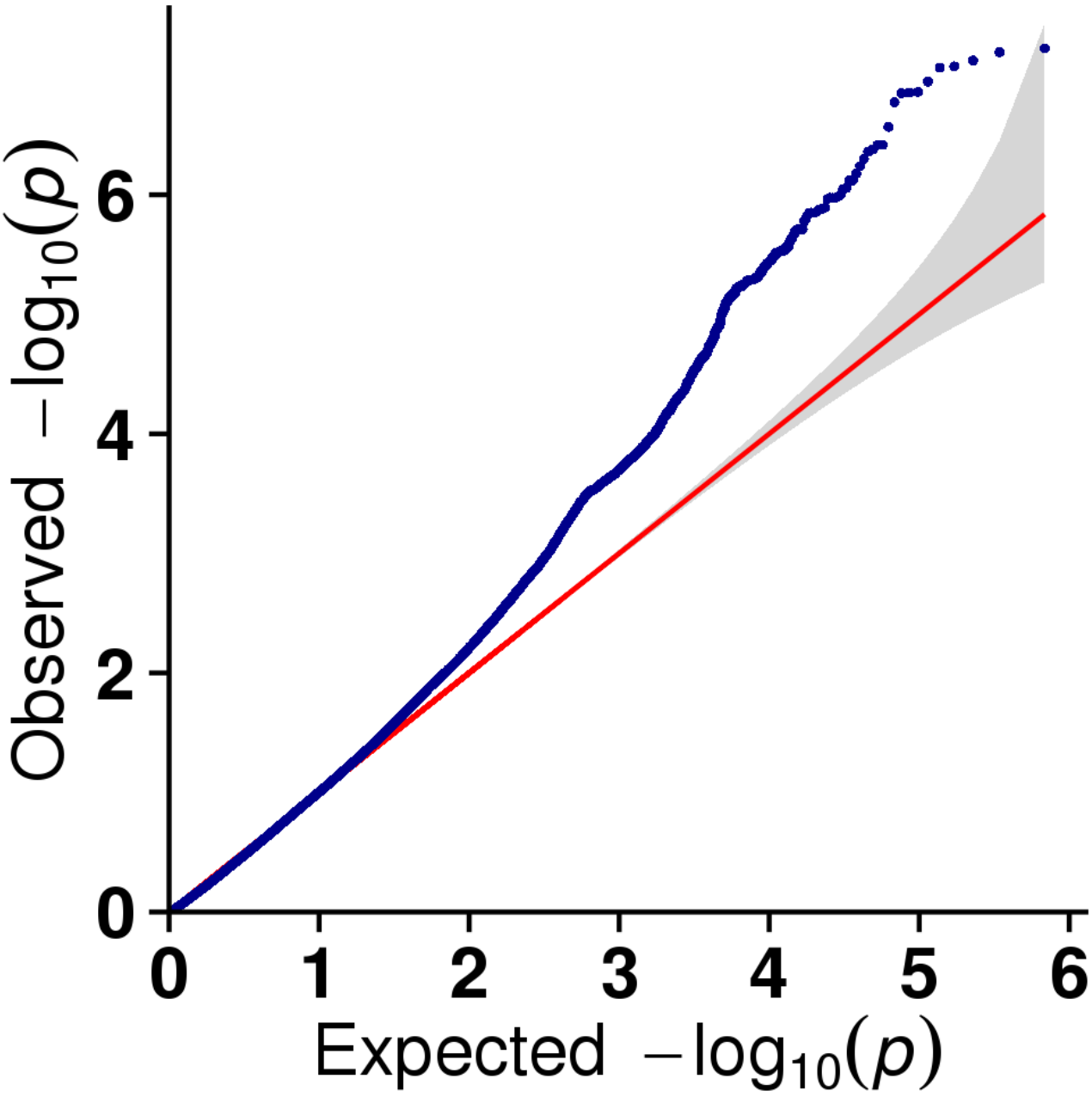

Figure 7

Genotyping results of GUCY1A1 gene

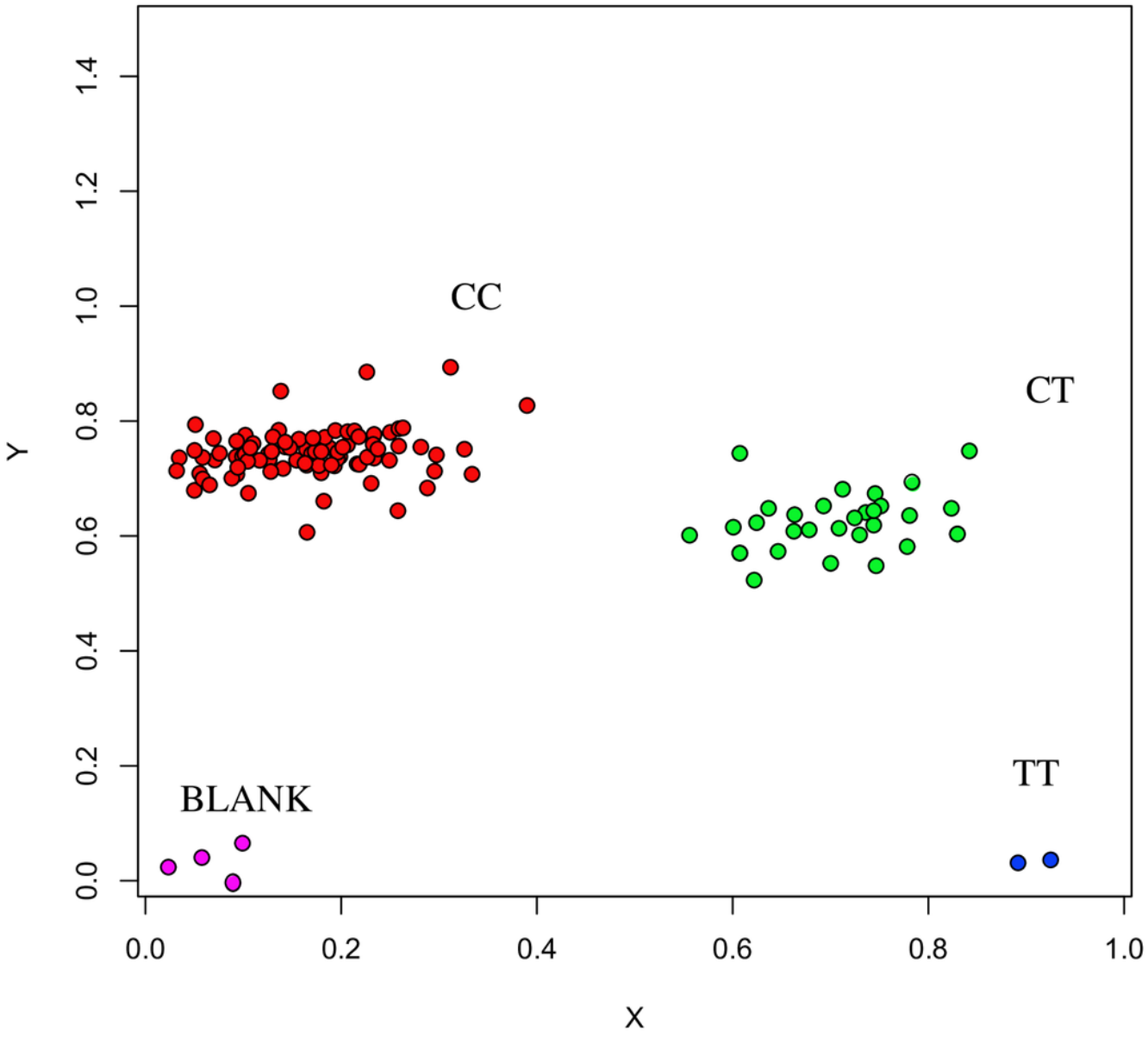

Figure 8

Genotyping results of FSHR gene

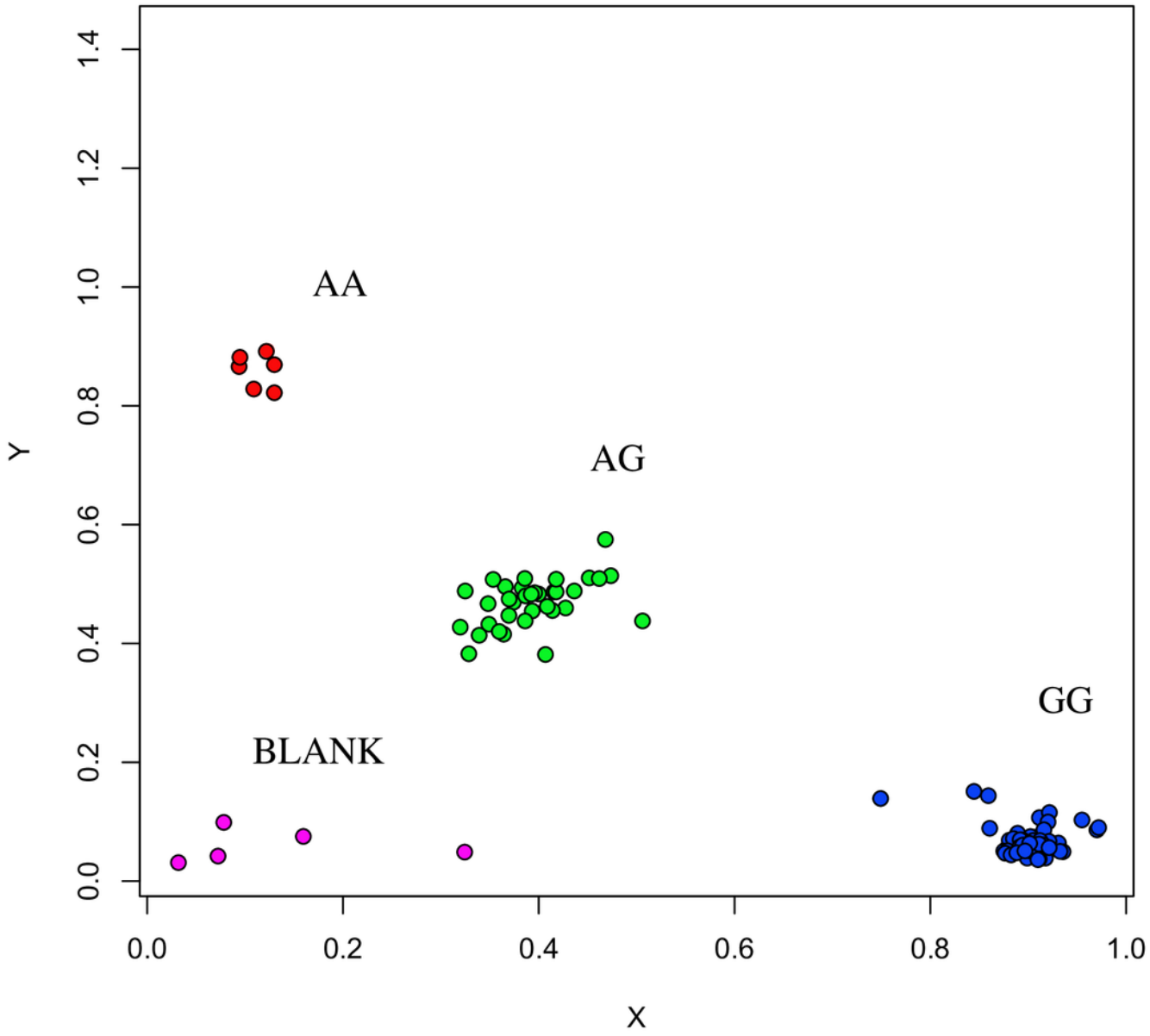

Figure 9

Genotyping results of COIL gene

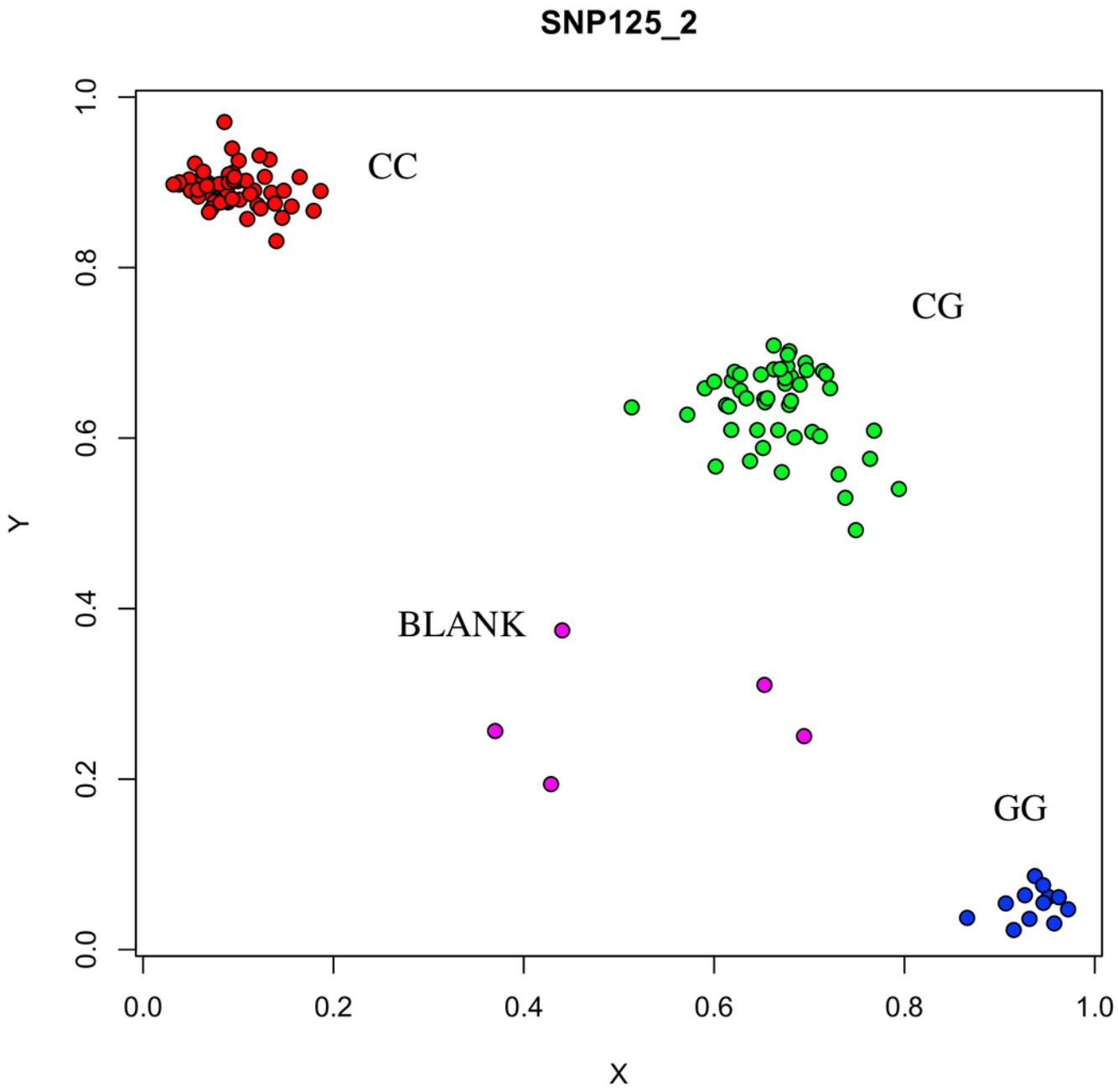

Figure 10

Correlation analysis between different genes and litter size