# Ten Genetic Loci Identified for Milk Yield, Fat, and Protein in Holstein Cattle

*Liyuan Liu[1,2#], Jinghang Zhou[1,2#], Chunpeng James Chen[2#], Juan Zhang[1], Wan Wen[3], Jia Tian[3], Zhiwu Zhang[2*], Yaling Gu[1*]*

*[1]School of Agriculture, Ningxia University, Yinchuan, Ningxia, China, [2]Department of Crop and Soil Sciences, Washington State University, Pullman, Washington, USA, [3]Animal Husbandry Workstation, Yinchuan, Ningxia, China.*

*[#]These authors contributed equally to this work.*

*[*]Correspondence should be addressed to* Zhiwu Zhang (email: zhiwu.zhang@wsu.edu) or Yaling Gu (email: guyl@nxu.edu.cn)

**Abstract**

High-yield and high-quality of milk are the primary goals of dairy production. Understanding the genetic architecture underlying these milk production traits is beneficial so that genetic variants can be targeted toward the genetic improvement. In this study, we measured five milk production traits in Holstein cattle population from China. These traits included milk yield, protein yield, fat yields; fat percentage and protein percentages. We used the estimated breeding values as dependent variables to conduct the genome-wide association studies (GWAS). Breeding values were estimated through pedigree relationships by using a mixed linear model for individuals with and without phenotypic data. Genotyping was carried out on the individuals with phenotypes by using the Illumina BovineSNP150 BeadChip. The association analyses were conducted by using the Fixed and random model Circulating Probability Unification (FarmCPU) method. A total of ten SNPs was detected above the genome-wide significant threshold, including six located in previously reported QTL regions. We found eight candidate genes within distances of 120 kb upstream or downstream to the associated SNPs. The most significant SNP is on *DGAT1* gene affecting milk fat and protein percentage. These genetic variants and candidate genes would be valuable resources to enhance dairy cattle breeding.

**Keywords:** Dairy, milk, fat, protein, *DGAT1*

## Introduction

Traditional dairy breeding mainly depends on phenotypic and pedigree information that is expensive, labor-intensive, and time-consuming. Dairy cows have a long generation interval and give birth only once a year. With the advances in molecular genetic techniques, genomic selection (GS) has been widely used in plant and animal breeding. It was demonstrated that implementation of genomic selection can increase genetic gain in dairy cattle [1]. Many statistical methods for GS have been proposed, and the method of integrating Genome-wide association studies (GWAS) and GS into one step has been developed, which account the genetic architecture of interest traits [2]. The conventional genomic best linear unbiased prediction (GBLUP) considering all markers that have the same effects, so thus ignores the markers that have large effect on target traits.

Research showed that the prediction accuracy was higher than that use of conventional GBLUP when fitting the most significant marker in the prediction model [3]. Also Resende et al suggested that the prediction accuracy reached maximum when constructed the genomic relationship matrix using causative quantities trait nucleotides (QTNs) [4]. Incorporating the public most significant markers from GWAS showed that it can improve the prediction accuracy in dairy cattle [5,6]. And the prediction accuracy were higher when accounting for variants from GWAS results [7]. Therefore, GWAS are very helpful for further GS. Since GWAS have proven to be a powerful method for identifying potential genetic variants, especially single-nucleotide polymorphisms (SNPs) associated with complex traits in humans and animals [8–10].

Milk production and milk composition are the most important economic traits in dairy industry. A large number of association studies reveals numerous QTLs for milk-related traits in dairy cattle population around the world over the past 20 years (CatttleQTLdb: https://www.animalgenome.org/cgi-bin/QTLdb/BT/index), and many researchers conducted meta-analysis based on GWAS results for milk-related traits in different cattle breeds [7,11,12]. Although early study revealed 105 SNPs associated with milk production traits in Chinese Holstein population, but they used a lower density marker which contains only about 50,000 SNPs [13].

Research showed that high-density genotype could provide markers close to the QTL and help in fine mapping of causative mutation[14]. Also Vanraden reported that high-density marker increased the precision of QTL detection in cattle population[15]. A maize study showed that the GWAS detection efficiency was improved when increased the marker density for husk traits [16]. In addition, several studies reported that the genomic prediction accuracy increased when increased the marker density in cattle breed[17] and shrimp breed[18],Therefore, it is necessary to use dense genotype to identify genetic variation and implementation of genomic prediction.

In this study, we conducted GWAS using Illumina BovineSNP150 BeadChip which contains about 150,000 SNPs. The population was from Holstein cows raised in the Ningxia area of northwest China. Our objectives were to identify new genetic variants associated with five milk traits, including milk yield, fat yield, protein yield, fat percentage, and protein percentage. We expect the newly identified genetic variants and potential candidate genes would become valuable resources for genomic evaluation.

## Material and methods

*Population and phenotypic data*

The study population is Holstein cows that were raised on 22 dairy farms in the Ningxia area of China. We used estimated breeding values (EBVs) as phenotypes implement association analysis, Milk yield (MY), fat yield (FY), protein yield (PY), fat percentage (FP), protein percentage (PP) measurements were recorded once a month for each cow after calving and the start of lactation. The milk yield was automatically recorded by the milking system in each farm, the milk components are tested by the Dairy Herd Improvement lab at Animal Husbandry of Extension Station in Ningxia, using spectrometers. FY was calculated as (FP*MY)/100; PY was calculated as (PP*MY)/100. The distribution of phenotypes and correlations between the different phenotypic traits are illustrated in **Figures** S1.

*Estimated breeding values*

The DMU package (Derivative-free approach to MUltivariate analysis ) [19] was used to estimate breeding values using Random Regression Test-Day Model [20,21]. Totally about

452,920 test-day records from 61,600 cows spanning a 9-year period (2011-2019) at their first lactation stage, all recorded data reflect the first lactation for each cow. We considered herd-test-day and calving year-season as fixed effects, calving month-age as a fixed regression effect, and individual additive effect and permanent environment effect as random regression effects. Fixed regression using a fourth order Legendre polynomial. The model equation is as follows:

$$y_{ijklm} = HTD_i + cays_j + \sum_{m=1}^{5} b_{km} X_{bm}(\omega) + \sum_{m=1}^{r} a_{lm} Z_{am}(\omega) + \sum_{m=1}^{s} p_{lm} W_{pm}(\omega) + e_{ijklm}$$

Where $y_{ijklm}$ is the test-day records m of cow ; $HTD_i$ is the fixed effect of the i th herd-test day; $cays_j$ is the fixed effect of the j th combined of calving years and season; $b_{km}$ is the m th fixed regression coefficient of the k th calving year-season , $X_{bm}$ is the Legendre polynomial m th covariable; $a_{lm}$ is the m th random regression coefficient under additive genetic effects, corresponds to the l th cow in the pedigree; $p_{lm}$ is the m th random regression coefficient under the permanent environment genetic effects, correspond to the l th cow in the data file; $Z_{am}$ and $W_{pm}$ are the m th covariance elements for additive effects and permanent environment effects; $\omega$ is the days of lactation after standardization; and $e_{ijklm}$ is the residual effects. A heatmap of estimated breeding values for milk production traits is illustrated in **Figure** S2.

*Genotypic data*

Blood samples from the 1,220 cows were collected by cattle farm staff in this study. DNA was extracted and Genotyping was carried out by Compass Biotechnology (http://www.kangpusen.com/) using the Illumina BovineSNP150 BeadChip. We used genome reference the Bos_taurus_UMD_3.1. In total, there were 124,743 variants for the association analysis after conducted quality control by Plink software [22]. Markers were removed if (1) the call rate of an individual genotype was less than 95%, (2) the call rate of a single SNP genotype was less than 90%, and (3) if the MAF of a SNP was less than 0.05 and deviated from Hardy-Weinberg equilibrium ($p < 1.0e-6$). We calculated marker intervals and linkage disequilibrium (LD) to estimate R square for all markers and plotted the marker distribution in **Figure** 1.

*Principle component analysis*

*P*rinciple component analysis (PCA) was conducted on 1,220 cows genotyped with 124,743 markers covering the whole genome to detect the population structure. The 1,220 cows are the

progeny of the frozen semen imported from multiple countries. The of origins of the sires were plotted against the principal components to demonstrate how the population structure was formed.

*Association analysis*

We performed the single-SNP analysis for the markers using the FarmCPU (Fixed and random model Circuitous Probability Unification) [23]. The genome wise threshold corresponding type I error of 1% was 4.0e-07 after Bonferroni multiple test correction (5%/124,743). FarmCPU can consider a population structure matrix as a covariate variable to correct false positive results caused by population stratification [23]. Considering the stratification in the dairy cow population, we fit the first three principal components as covariate variables in the GWAS models.

*Annotation of candidate gene and pathway analysis*

The genome reference Bos_taurus_UMD_3.1 was used to search candidate gene. The candidate genes were found within distances of 120 kb upstream or downstream to the associated SNPs. The online website "https://oct2018.archive.ensembl.org/Bos_taurus/Info/Index", "https://www.ncbi.nlm.nih.gov/gene/ ", "https://www.genome.jp/kegg/pathway.html" were used for functional analysis and pathway analysis of the candidate gene by GWAS.

## Results

*Phenotypic and estimated genetic parameters*

Phenotype distributions and correlations among phenotypic traits, estimated breeding values, and residuals are shown in the supplementary material (**Figure** S1). There were strong positive phenotypic correlations between "yield" type of traits, including MY, PY, and FY. Their phenotypic correlations were 0.90 (MY and PY), 0.70 (MY and FY), and 0.74 (PY and FY). We also found strong positive genetic correlations between MY and PY ($r_g$ = 0.92), MY and FY ($r_g$ = 0.84), and PY and FY ($r_g$ = 0.88). In contrast, there were weak negative phenotypic and genetic correlations between MY and FP ($r_p$ = -0.15, $r_g$ = -0.32), MY and PP ($r_p$ = -0.20, $r_g$ = -0.44).

In this study, we used EBVs as the dependent variables for GWAS, **Figure** S2 showed that the heatmap of EBVs for five milk traits. A test-day model to estimate the heritability for each trait and breeding values for individuals. The heritability estimates for MY, FY, PY, FP and PP were 0.12, 0.18, 0.27, 0.30 and 0.32, respectively (**Table** 1).

*Marker information*

We conducted GWAS analyses with 1,220 Holstein dairy cows and 124,743 markers after quality control (QC). Markers covered all 29 autosomes plus the X sex chromosome (**Figure 1a**). After the QC filtering, we re-calculated the minor allele frequency (MAF) for all SNPs. The minimum MAF was 3.8%. There were only 0.1% of markers with MAF below 5% (**Figure 1c**). Marker density was high. Majority of markers (56%) are within 20 kb distances to their adjacent markers (**Figure 1b**). Within such distance, the LD was strong (average $R^2 = 0.56$) (**Figure 1d**).

*Population structure*

To determine the level of population stratification, we plotted the population structure by principal component analysis (PCA). The population stratified into two unevenly sized groups (**Figure 2**). We also made scatter plot of bull's country source (**Figure** S3). To adjust for the population stratification, we fit the first three principal components (PCs) as covariate variables in the association analysis. We also made a scatter plot between the first three PCs and the five traits, there are weak correlation observed between the PCs and these phenotypes (**Figure** S4).

*Results of the genome-wide associations*

By drawing the Quantile-Quantile (QQ) plots, we found that the model for GWAS analysis in this study was reasonable, and the point at the upper right corner also shown that some significant markers were found that associated with four milk traits (**Figure** 3). We used $P <$ 4.0e-07 as threshold, which was corresponding to 1% of type I error after Bonferroni multiple test correction. We did not find threshold significant SNPs associated with MY, a total of ten highly significant SNPs associated with four traits (**Table** 2, **Figure** 3). Three SNPs (rs42295213, rs136949224 and rs109421300) associated with FP are located on BTA1, 8 and 14, four SNPs (rs43526055, rs137676276, rs109528658 and rs135780687) associated with FY are located on BTA7, 11, 17, and X, respectively. Three SNPs (rs109875012, rs109421300 and

rs108996837) associated with PP are located on BTA5, 14, and 21, respectively. One SNP associated with PY is located on BTA5. Four of these ten significant SNPs are located inside genes *EPHA6* (EPH receptor A6), *SLCO1A2* (solute carrier organic anion transporter family member 1A2), *DGAT1* (diacylglycerol O-acyltransferase 1) *and EP400* (E1A binding protein p400). The SNP (rs109875012) on BTA5 located close to the *ZNF384* (zinc finger protein 384). The SNP (rs10705865) on BTA8 located close to *SCARA5* (scavenger receptor class A member 5) gene. The SNP (rs137676276) on BTA 11 located close to *VIT* (vitrin). The SNP (rs108996837) on BTA 21 located close to *EXOC3L4* (exocyst complex component 3 like 4), and the SNP (rs135780687) on X chromosome located close to *GRPR* (gastrin releasing peptide receptor). The most significant SNP (rs109421300) associated with both FP and PP located in the *DGAT1* gene. Two SNPs (rs137676276, rs108996837) exhibit notably smaller MAFs compared to other SNPs, 0.11 and 0.12, respectively (**Table** 2).

*Pleiotropic QTLs for milk production traits*

We used the SNPs with $P < 0.0005$ to make heatmap to look for markers associated with two or more milk production traits (**Figure** S5). We expected to find pleiotropic QTLs because these milk traits are moderately or highly correlated (**Figure** S1, S2). QTL on BTA1 at 0.4Mb, on BTA2 at 14.8Mb, on BTA2 at 134.3Mb, on BTA3 at 64.4Mb, on BTA11 at 39 Mb, on BTA17 at 29−32.9Mb, on BTA20 at 67.8 Mb, BTA22 at 54.6Mb associated with three milk production traits (MY, PY, and FY). QTL on BTA1 at 40.3-41.1Mb, on BTA6 at 37.7Mb, on BTA14 at 73-77Mb, on BTA17 at 29-32Mb associated with MY and PY. QTL on BTA20 at 24Mb associated with FY and FY.

## Discussion

*Population structure*

Population stratification is an important confounding factor due to systematic ancestry differences that can cause false positives in GWAS [24]. Partially in cattle population, not all the registered cattle are the full blood. This point can be explained by looking at the cattle breed registration requirements in different countries. Here we take the Holstein cattle as an example, one of the requirements to register a Holstein cattle in China is that the cattle at least has 87.5% blood of Holstein (Chinese Holstein, GB/T 3157 2008), all the animals with Holstein genetics

can be registered in Canadian (https://www.holstein.ca/Public/en/Services/Registration/Registration_Eligibilities) and there are the similar clause in USA (http://www.holsteinusa.com/animal_id/register.html). According to the above standards we can see the common ground to register a Holstein cattle between different countries is that not all the Holstein cattle are the full blood. Even most of the Holstein cattle are pure breed, some of registered cattle could still contains a little other blood in the long-term breeding progress. That's why the population structure analysis is necessary in this study. By the principle component analysis, the PCA scatter plot showed that there are two groups in the population in this study; that is, a small number of individuals divided into the other group (**Figure** 2, **Figure** S3). There are probably two reasons caused the deviation of the population structure, the first reason is the semen the cattle farm used are different, as we know most of the farms in this study participated in a dairy breeding project that introduce of some Holstein frozen semen overseas annually from 2013 to 2018. Another reason is that some cows are introduced from different countries there might has some various background.

As we observed the population structure, the principle components were fitted as covariance do association analysis to correcting population stratification. After adjusted PCA factors, there are four SNPs are overlapped with association model not fitted PCs, will discuss that later.

*GWAS for milk production traits*

Milk yield and composition are very important economic traits in dairy cattle because good milk production performance can bring greater economic benefits. As early as 1994, a research identified a QTL significantly associated with FY was linked to kappa-casein and a QTL for PY was linked to beta-lactoglobulin [25]. Subsequently, a growing number of studies detected thousands and tens of QTLs through the 30 chromosomes associated with 646 different traits in cattle (Cattle QTLdb). In this study, there were no significant SNPs passed the Bonferroni correction threshold for MY, but ten SNPs were significantly associated with the other four milk traits (FP, FY, PP and PY).

In this study, we found three SNPs on BTA1, 8, 14 were associated with FP, they are within reported QTLs [26–29]. By search the candidate genes within distances of 120 kb upstream or

downstream to the associated SNPs, the genome reference is Bos_taurus_UMD_3.1. The SNP on BTA1 located in the *EPHA6* gene, which have functions of ATP binding (GO:0005524), protein binding (GO: 0005515), protein tyrosine kinase activity (GO: 0004713), and has been proposed to participate in Axon guidance pathway (KEGG: bta04360). The SNP on BTA8 located close to *SCARA5* gene, this gene is a member of the scavenger receptor (SR) family, which is broad expression in fat tissue in human individuals, research showed that scavenger receptors involved in lipid accumulation and inflammation [30], some studies reported the gene plays a critical role in progression and metastasis of breast cancer [31], and involved in breast carcinogenesis [32]. *SCARA5* gene may be a potential candidate gene involved in the biosynthesis of milk fat. The SNP on BTA14 associated both with FP and PP, which is in the *DGAT1* gene. As we know, *DGAT1* gene is widely reported associated with milk yield and composition, especially K232A polymorphism affecting on milk fat and protein [27,33,34]. In this study, the SNP (rs109421300) we identified is within an intron of *DGAT1* gene is the most significantly affecting FP ($P$ = 9.92e-25 ), and also affecting PP ($P$ = 7.05e-9 ), and a lot of studies have reported for milk yield and composition in Holstein, Jersey and Holstein-Friesian cattle [35–37]. *DGAT1* gene encodes an multipass transmembrane protein that functions as a key metabolic enzyme, involved in triacylglycerol biosynthesis, and glycerolipid metabolism (KEGG: bta00561), retinol metabolism (KEGG: bta00830), metabolic pathways (KEGG: bta01100) and fat digestion and absorption (KEGG: bta04975). The PP-associated SNP on BTA5 is within previously reported milk fatty acid content QTL [28], this SNP is close to the *ZNF384*, which is encodes a C2H2-type zinc finger protein, it may function as a transcription factor and have DNA binding and metal ion binding functions. Another PP-associated SNP on BTA21. Four SNPs on BTA7, 11, 17 and X chromosome are associated with FY. The SNP on BTA11 at 19Mb located near the VIT, this gene encodes an extracellular matrix (ECM) protein, which has function of glycosaminoglycan binding (GO:0005539), is has been proposed participate in growth plate cartilage chondrocyte morphogenesis pathway (GO:0003429) and nervous system development pathway (GO:0007399). The SNP on BTA17 is within previously reported QTL for milk protein composition [38], and it also located in the *EP400*, this gene has been participate in histone H2A acetylation, histone H4 acetylation, and has the function of ATP binding, DNA binding, protein binding and helicase activity. The FY-associated SNP on chromosome X was located nearby the *GRPR*, this gene is gastrin releasing peptide receptor, gastrin-releasing peptide regulates

numerous functions of gastrointestinal and central nervous system, which involved in calcium signaling pathway and neuroactive ligand-receptor interaction. Song reviewed a lot of studies on calcium signaling pathway participated in the effect of the sympathetic never in regulating adipose metabolism [39]. We suspect that *GRPR* gene may affect milk fat through the calcium signaling pathway. Because of the complexity of the biological mechanisms of quantitative traits, the study is only based on SNP data, many studies showed that copy number variation and DNA methylation are also associated with milk performance[40–42], The PY-associated SNP on BTA5 is within previously reported QTL region, which is associated with milk fatty acid content, this SNP is in the *SLCO1A2* gene, which is encoding solute carrier anion transporter family, member 1A2, the gene participate in digestive system, organic anion transport process and bile secretion pathway (KEGG: 04976). Furthermore, when the association model not fitted PCs as covariance, four SNPs are overlapped with model with fitted PCs：rs109875012 (*ZNF384*), rs136949224 (*SCARA5*), rs109421300 (*DGAT1*), and rs109528658 (*EP400*), it is suggested that these significant SNPs could more stable and reliable.

*Correlations among milk traits and pleiotropic QTLs*

Estimates of heritability and genetic correlations are essential population genetic parameters in animal breeding research and application of animal breeding programs [43]. Genetic correlation can be useful for indirect selection, selection in different environment and selecting multiple traits simultaneously. We found there were relatively high genetic correlation between milk yield and protein yield, is consistent with the results in Holsteins [44] and the results in Jersey dairy cattle [43]. We expect that there are overlapping region of genetic variation between different traits that are relatively high correlation. A heatmap using *P*-values from GWAS results helped identify pleiotropic QTLs (**Figure** S5). The QTL on BTA2 at 134Mb associated with MY, PY and FY, and also reported associated with the milk composition in another Holstein study [45]. The MY- and PY-associated QTL is adjacent to the *ABCG2* gene on BTA 6, previously studies reported a missense mutations in the *ABCG2* gene is associated with milk yield and composition in Holstein, Braunvieh, and Fleckvieh cattle [11,46] , a intron variant affecting milk fatty acids in Chinese Holstein [47], and also reported for body weight, calving ease direct in US cattle breed [48,49]. The QTL on BTA14 in the *DGAT1* gene associated with milk yield, fat and protein percentage. The QTL on BTA17 associated with MY, PY and FY, some study have pointed to

QTL on this region related to milk fatty acid and body weight [45,50]. A previous study reported a QTL on BTA 20 is associated with PP and body weight[51,52], in this study, we found this QTL is associated with MY, PY and FY. The QTL on BTA24 is pleiotropic and is associated with MY and FY, Ibeagha-Awemu and Peter reported for milk fatty acid [53], Kiser and Clancey reported for conception rate in Holstein cows [54].

## Conclusions

We performed genome-wide association studies using test-day records for five milk production traits (milk yield, fat yield, protein yield, fat percentage, and protein percentage) in Holstein cows. In total of 10 significantly SNPs associated with fat and protein percentage, fat and protein yield, five SNPs are located within previously reported QTLs. An important top variant in the *DGAT1* gene affecting milk fat and protein, and we also found *SCARA5* and *GRPR* gene may be potential candidate gene involved in the biosynthesis of milk fat. In addition, some pleiotropic QTLs on BTA1, 2, 3, 6, 11, 14, 17, 20, and 24 associated with more than two milk traits, these results could provide some basis for molecular breeding and targets for gene cloning.

## Acknowledgements

This study was partially supported by the High-Level Academic Papers in Ningxia University, the project of High-yield and High-quality Dairy Cattle Breeding (2019NYYZ05), the National Science Foundation (Award # DBI 1661348), and the USDA National Institute of Food and Agriculture Hatch project (1014919). The authors would like to thank the DHI measurement center of Ningxia Animal Husbandry of Extension Station and 21 dairy farms for their phenotypic data. We also thank Dr. Linda R. Klein for valuable writing advice and editing the manuscript.

## Author contributions

Conceived experiment: Zhiwu Zhang and Yaling Gu;

Data analyses: Liyuan Liu, Jinghang Zhou, and James Chen;

Data collection: Juan Zhang, Jia Tian, and Wan Wen;

Wrote manuscript: Liyuan Liu, Jinghang Zhou, James Chen, Zhiwu Zhang, and Yaling Gu

## Disclosures

The authors declare no conflict of interest.

## Ethics Statement

No applicable. This study only did statistical analysis based on the former database which come from the Holstein breeding project in Ningxia province, China. All the authors in this study didn't participated in any sample collection process. All the farms involved were consent and agree to take part in this research.

## Literature Cited


1. Hayes, B.J.; Bowman, P.J.; Chamberlain, A.J.; Goddard, M.E. Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* **2009**, *92*, 433–443, doi:10.3168/jds.2008-1646.
2. Wang, J.; Zhou, Z.; Zhang, Z.; Li, H.; Liu, D.; Zhang, Q.; Bradbury, P.J.; Buckler, E.S.; Zhang, Z. Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. *Heredity* **2018**, *121*, 648–662, doi:10.1038/s41437-018-0075-0.
3. Lopes, M.S.; Bovenhuis, H.; van Son, M.; Nordbø, Ø.; Grindflek, E.H.; Knol, E.F.; Bastiaansen, J.W.M. Using markers with large effect in genetic and genomic predictions. *Journal of Animal Science* **2017**, *95*, 59, doi:10.2527/jas2016.0754.
4. Fragomeni, B.O.; Lourenco, D.A.L.; Masuda, Y.; Legarra, A.; Misztal, I. Incorporation of causative quantitative trait nucleotides in single-step GBLUP. *Genet Sel Evol* **2017**, *49*, 59, doi:10.1186/s12711-017-0335-0.
5. Zhang, Z.; Ober, U.; Erbe, M.; Zhang, H.; Gao, N.; He, J.; Li, J.; Simianer, H. Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies. *PLoS ONE* **2014**, *9*, e93017, doi:10.1371/journal.pone.0093017.
6. Raymond, B.; Bouwman, A.C.; Schrooten, C.; Houwing-Duistermaat, J.; Veerkamp, R.F. Utility of whole-genome sequence data for across-breed genomic prediction. *Genet Sel Evol* **2018**, *50*, 27, doi:10.1186/s12711-018-0396-8.
7. Teissier, M.; Sanchez, M.P.; Boussaha, M.; Barbat, A.; Hoze, C.; Robert-Granie, C.; Croiseau, P. Use of meta-analyses and joint analyses to select variants in whole genome sequences for genomic evaluation: An application in milk production of French dairy cattle breeds. *Journal of Dairy Science* **2018**, *101*, 3126–3139, doi:10.3168/jds.2017-13587.
8. Kristensen, V.N.; Børresen-Dale, A.-L. SNPs associated with molecular subtypes of breast cancer: On the usefulness of stratified Genome-wide Association Studies (GWAS) in the identification of novel susceptibility loci. *Mol Oncol* **2008**, *2*, 12–15, doi:10.1016/j.molonc.2008.02.003.
9. Yang, W.; Gu, C.C. A whole-genome simulator capable of modeling high-order epistasis for complex disease. *Genet. Epidemiol.* **2013**, *37*, 686–694, doi:10.1002/gepi.21761.
10. Hayes, B.; Goddard, M. Genome-wide association and genomic selection in animal breedingThis article is one of a selection of papers from the conference "Exploiting Genome-wide Association in Oilseed Brassicas: a model for genetic improvement of major OECD crops for sustainable farming". *Genome* **2010**, *53*, 876–883, doi:10.1139/G10-076.
11. Pausch, H.; Emmerling, R.; Gredler-Grandl, B.; Fries, R.; Daetwyler, H.D.; Goddard, M.E. Meta-analysis of sequence-based association studies across three cattle breeds reveals

25 QTL for fat and protein percentages in milk at nucleotide resolution. *BMC Genomics* **2017**, *18*, 853, doi:10.1186/s12864-017-4263-8.
12. Marete, A.G.; Guldbrandtsen, B.; Lund, M.S.; Fritz, S.; Sahana, G.; Boichard, D. A Meta-Analysis Including Pre-selected Sequence Variants Associated With Seven Traits in Three French Dairy Cattle Populations. *Front. Genet.* **2018**, *9*, 522, doi:10.3389/fgene.2018.00522.
13. Jiang, L.; Liu, J.; Sun, D.; Ma, P.; Ding, X.; Yu, Y.; Zhang, Q. Genome Wide Association Studies for Milk Production Traits in Chinese Holstein Population. *PLoS ONE* **2010**, *5*, e13661, doi:10.1371/journal.pone.0013661.
14. Pryce, J.E.; Johnston, J.; Hayes, B.J.; Sahana, G.; Weigel, K.A.; McParland, S.; Spurlock, D.; Krattenmacher, N.; Spelman, R.J.; Wall, E.; et al. Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. *Journal of Dairy Science* **2014**, *97*, 1799–1811, doi:10.3168/jds.2013-7368.
15. VanRaden, P.M.; Null, D.J.; Sargolzaei, M.; Wiggans, G.R.; Tooker, M.E.; Cole, J.B.; Sonstegard, T.S.; Connor, E.E.; Winters, M.; van Kaam, J.B.C.H.M.; et al. Genomic imputation and evaluation using high-density Holstein genotypes. *Journal of Dairy Science* **2013**, *96*, 668–678, doi:10.3168/jds.2012-5702.
16. Cui, Z.; Dong, H.; Zhang, A.; Ruan, Y.; Jiang, S.; He, Y.; Zhang, Z. Denser Markers and Advanced Statistical Method Identified More Genetic Loci Associated with Husk Traits in Maize. *Sci Rep* **2020**, *10*, 8165, doi:10.1038/s41598-020-65164-0.
17. Pryce, J.E.; Arias, J.; Bowman, P.J.; Davis, S.R.; Macdonald, K.A.; Waghorn, G.C.; Wales, W.J.; Williams, Y.J.; Spelman, R.J.; Hayes, B.J. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. *Journal of Dairy Science* **2012**, *95*, 2108–2119, doi:10.3168/jds.2011-4628.
18. Wang, Q.; Yu, Y.; Yuan, J.; Zhang, X.; Huang, H.; Li, F.; Xiang, J. Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp Litopenaeus vannamei. *BMC Genet* **2017**, *18*, 45, doi:10.1186/s12863-017-0507-5.
19. Madsen, P.; Jensen, J. A Package for Analysing Multivariate Mixed Models. Version 6, release 5. 31.
20. Schaeffer, L.R. Application of random regression models in animal breeding. *Livestock Production Science* **2004**, *86*, 35–45, doi:10.1016/S0301-6226(03)00151-9.
21. Cobuci, J.A.; Euclydes, R.F.; Lopes, P.S.; Costa, C.N.; Torres, R. de A.; Pereira, C.S. Estimation of genetic parameters for test-day milk yield in Holstein cows using a random regression model. *Genetics and Molecular Biology* **2005**, *28*, 75–83, doi:10.1590/S1415-47572005000100013.
22. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575, doi:10.1086/519795.
23. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S.; Zhang, Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLOS Genetics* **2016**, *12*, e1005767, doi:10.1371/journal.pgen.1005767.
24. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **2006**, *38*, 904–909, doi:10.1038/ng1847.
25. Bovenhuis, H.; Weller, J.I. Mapping and Analysis of Dairy Cattle Quantitative Trait Loci by Maximum Likelihood Methodology Using Milk Protein Genes as Genetic Markers. *Genetics* **1994**, *137*, 267–280.

26. Thaller, G.; Krämer, W.; Winter, A.; Kaupe, B.; Erhardt, G.; Fries, R. Effects of DGAT1 variants on milk production traits in German cattle breeds1. *Journal of Animal Science* **2003**, *81*, 1911–1918, doi:10.2527/2003.8181911x.
27. Kühn, C.; Thaller, G.; Winter, A.; Bininda-Emonds, O.R.P.; Kaupe, B.; Erhardt, G.; Bennewitz, J.; Schwerin, M.; Fries, R. Evidence for Multiple Alleles at the *DGAT1* Locus Better Explains a Quantitative Trait Locus With Major Effect on Milk Fat Content in Cattle. *Genetics* **2004**, *167*, 1873–1881, doi:10.1534/genetics.103.022749.
28. Bouwman, A.C.; Bovenhuis, H.; Visker, M.H.; van Arendonk, J.A. Genome-wide association of milk fatty acids in Dutch dairy cattle. *BMC Genet* **2011**, *12*, 43, doi:10.1186/1471-2156-12-43.
29. Sanchez, M.P.; Govignon-Gion, A.; Ferrand, M.; Gelé, M.; Pourchet, D.; Amigues, Y.; Fritz, S.; Boussaha, M.; Capitan, A.; Rocha, D.; et al. Whole-genome scan to detect quantitative trait loci associated with milk protein composition in 3 French dairy cattle breeds. *Journal of Dairy Science* **2016**, *99*, 8203–8215, doi:10.3168/jds.2016-11437.
30. Nayeri, S.; Sargolzaei, M.; Abo-Ismail, M.K.; May, N.; Miller, S.P.; Schenkel, F.; Moore, S.S.; Stothard, P. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* **2016**, *17*, 75, doi:10.1186/s12863-016-0386-1.
31. You, K.; Su, F.; Liu, L.; Lv, X.; Zhang, J.; Zhang, Y.; Liu, B. SCARA5 plays a critical role in the progression and metastasis of breast cancer by inactivating the ERK1/2, STAT3, and AKT signaling pathways. *Mol Cell Biochem* **2017**, *435*, 47–58, doi:10.1007/s11010-017-3055-4.
32. Ulker, D.; Ersoy, Y.E.; Gucin, Z.; Muslumanoglu, M.; Buyru, N. Downregulation of SCARA5 may contribute to breast cancer via promoter hypermethylation. *Gene* **2018**, *673*, 102–106, doi:10.1016/j.gene.2018.06.036.
33. Grisart, B.; Farnir, F.; Karim, L.; Cambisano, N.; Kim, J.-J.; Kvasz, A.; Mni, M.; Simon, P.; Frere, J.-M.; Coppieters, W.; et al. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences* **2004**, *101*, 2398–2403, doi:10.1073/pnas.0308518100.
34. Bovenhuis, H.; Visker, M.H.P.W.; Poulsen, N.A.; Sehested, J.; van Valenberg, H.J.F.; van Arendonk, J.A.M.; Larsen, L.B.; Buitenhuis, A.J. Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. *Journal of Dairy Science* **2016**, *99*, 3113–3123, doi:10.3168/jds.2015-10462.
35. Meredith, B.K.; Kearney, F.J.; Finlay, E.K.; Bradley, D.G.; Fahey, A.G.; Berry, D.P.; Lynn, D.J. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genet* **2012**, *13*, 21, doi:10.1186/1471-2156-13-21.
36. Aliloo, H.; Pryce, J.E.; González-Recio, O.; Cocks, B.G.; Hayes, B.J. Validation of markers with non-additive effects on milk yield and fertility in Holstein and Jersey cows. *BMC Genet* **2015**, *16*, 89, doi:10.1186/s12863-015-0241-9.
37. Frąszczak, M.; Szyda, J. Comparison of significant single nucleotide polymorphisms selections in GWAS for complex traits. *J Appl Genetics* **2016**, *57*, 207–213, doi:10.1007/s13353-015-0305-6.
38. Schopen, G.C.B.; Koks, P.D.; van Arendonk, J.A.M.; Bovenhuis, H.; Visker, M.H.P.W. Whole genome scan to detect quantitative trait loci for bovine milk protein composition. *Animal Genetics* **2009**, *40*, 524–537, doi:10.1111/j.1365-2052.2009.01880.x.
39. Song, Z.; Wang, Y.; Zhang, F.; Yao, F.; Sun, C. Calcium Signaling Pathways: Key Pathways in the Regulation of Obesity. *IJMS* **2019**, *20*, 2768, doi:10.3390/ijms20112768.
40. Wen, Y.; He, H.; Liu, H.; An, Q.; Wang, D.; Ding, X.; Shi, Q.; Feng, Y.; Wang, E.; Lei, C.; et al. Copy number variation of the *USP16* gene and its association with milk traits in Chinese Holstein cattle. *Animal Biotechnology* **2020**, 1–6, doi:10.1080/10495398.2020.1777148.

41. Di Gerlando, R.; Sutera, A.M.; Mastrangelo, S.; Tolone, M.; Portolano, B.; Sottile, G.; Bagnato, A.; Strillacci, M.G.; Sardina, M.T. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PLoS ONE* **2019**, *14*, e0215204, doi:10.1371/journal.pone.0215204.
42. Zhao, H.; Zhang, S.; Wu, X.; Pan, C.; Li, X.; Lei, C.; Chen, H.; Lan, X. DNA methylation pattern of the goat <i>PITX1</i> gene and its effects on milk performance. *Arch. Anim. Breed.* **2019**, *62*, 59–68, doi:10.5194/aab-62-59-2019.
43. Missanjo, E.; Imbayarwo-Chikosi, V.; Halimani, T. Estimation of Genetic and Phenotypic Parameters for Production Traits and Somatic Cell Count for Jersey Dairy Cattle in Zimbabwe. *ISRN Veterinary Science* **2013**, *2013*, 1–5, doi:10.1155/2013/470585.
44. Zaabza, H.B.; Gara, A.B.; Rekik, B. Genetic analysis of milk production traits of Tunisian Holsteins using random regression test-day model with Legendre polynomials. *ajas* **2018**, *31*, 636–642, doi:https://doi.org/10.5713/ajas.17.0332.
45. Gebreyesus, G.; Buitenhuis, A.J.; Poulsen, N.A.; Visker, M.H.P.W.; Zhang, Q.; van Valenberg, H.J.F.; Sun, D.; Bovenhuis, H. Multi-population GWAS and enrichment analyses reveal novel genomic regions and promising candidate genes underlying bovine milk fatty acid composition. *BMC Genomics* **2019**, *20*, 178, doi:10.1186/s12864-019-5573-9.
46. Cohen-Zinder, M. 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 Research* **2005**, *15*, 936–944, doi:10.1101/gr.3806705.
47. Li, C.; Sun, D.; Zhang, S.; Yang, S.; Alim, M.A.; Zhang, Q.; Li, Y.; Liu, L. Genetic effects of FASN, PPARGC1A, ABCG2 and IGF1 revealing the association with milk fatty acids in a Chinese Holstein cattle population based on a post genome-wide association study. *BMC Genet* **2016**, *17*, 110, doi:10.1186/s12863-016-0418-x.
48. Saatchi, M.; Schnabel, R.D.; Taylor, J.F.; Garrick, D.J. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* **2014**, *15*, 442, doi:10.1186/1471-2164-15-442.
49. Smith, J.L.; Wilson, M.L.; Nilson, S.M.; Rowan, T.N.; Oldeschulte, D.L.; Schnabel, R.D.; Decker, J.E.; Seabury, C.M. Genome-wide association and genotype by environment interactions for growth traits in U.S. Gelbvieh cattle. *BMC Genomics* **2019**, *20*, 926, doi:10.1186/s12864-019-6231-y.
50. McClure, M.C.; Morsci, N.S.; Schnabel, R.D.; Kim, J.W.; Yao, P.; Rolf, M.M.; McKay, S.D.; Gregg, S.J.; Chapple, R.H.; Northcutt, S.L.; et al. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Anim. Genet.* **2010**, *41*, 597–607, doi:10.1111/j.1365-2052.2010.02063.x.
51. Snelling, W.M.; Allan, M.F.; Keele, J.W.; Kuehn, L.A.; McDaneld, T.; Smith, T.P.L.; Sonstegard, T.S.; Thallman, R.M.; Bennett, G.L. Genome-wide association study of growth in crossbred beef cattle12. *Journal of Animal Science* **2010**, *88*, 837–848, doi:10.2527/jas.2009-2257.
52. Jiang, J.; Ma, L.; Prakapenka, D.; VanRaden, P.M.; Cole, J.B.; Da, Y. A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. *Front. Genet.* **2019**, *10*, 412, doi:10.3389/fgene.2019.00412.
53. Ibeagha-Awemu, E.M.; Peters, S.O.; Akwanji, K.A.; Imumorin, I.G.; Zhao, X. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Sci Rep* **2016**, *6*, 31109, doi:10.1038/srep31109.
54. Kiser, J.N.; Clancey, E.; Moraes, J.G.N.; Dalton, J.; Burns, G.W.; Spencer, T.E.; Neibergs, H.L. Identification of loci associated with conception rate in primiparous Holstein cows. *BMC Genomics* **2019**, *20*, 840, doi:10.1186/s12864-019-6203-2.

**Figures and Tables**

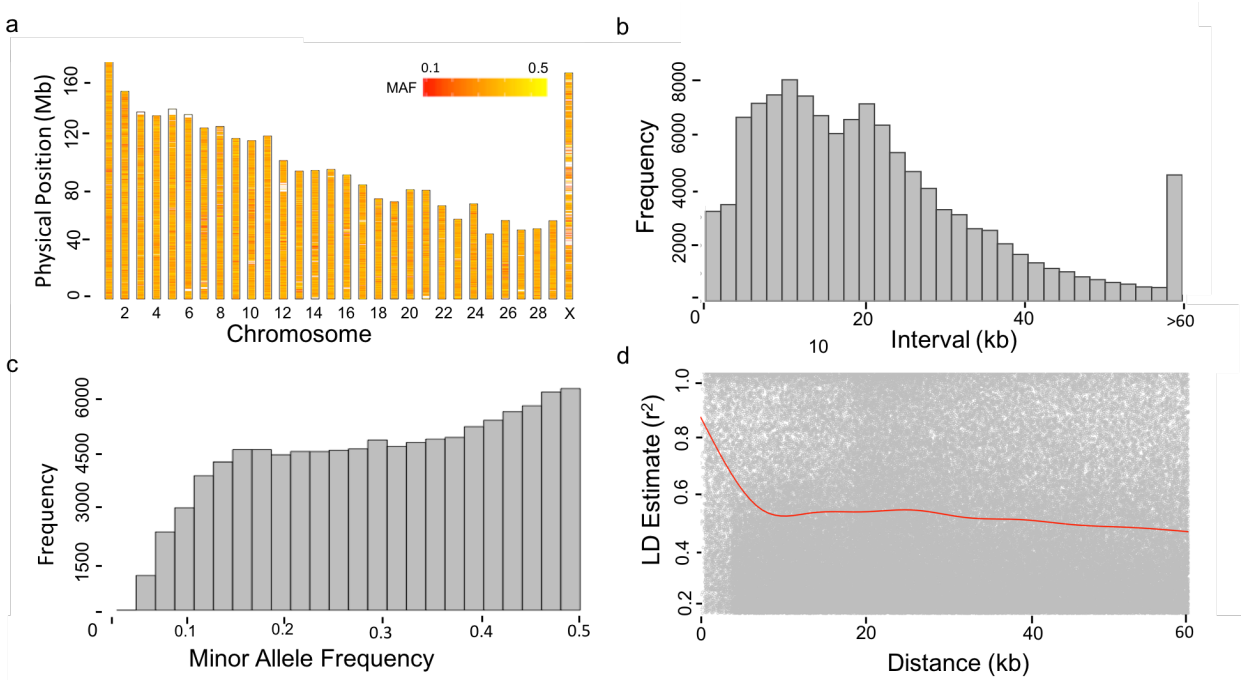


**Figure 1. Properties of Single Nucleotide Polymorphisms (SNPs)**. A total of 1,220 cows were genotyped by the Illumina Bovine 150 k BeadChip. After conducting quality control on both minor allele frequency (MAF, above 5%) and missing rate (< 10%), 1,220 individuals and 124,743 SNP remained. The distribution of the filtered SNPs is displayed over the 30 bovine chromosomes except for Y chromosome (a). The MAFs of SNPs were re-calculated after the filtering and were displayed by a heat map. Consequently, there SNPs with MAF < 5% remained, as demonstrated by the histogram (c). The density of SNPs is displayed by the frequency of the distance between adjacent SNPs (b). The distances over 60 kb clustered into one group. The maximum distance was 100.06 kb. Pairwise Linkage Disequilibrium (LD) was calculated as the R square for SNPs within the 100 kb window. The decay of LD over distance (red line) is displayed by the pairwise LD and moving average (d).

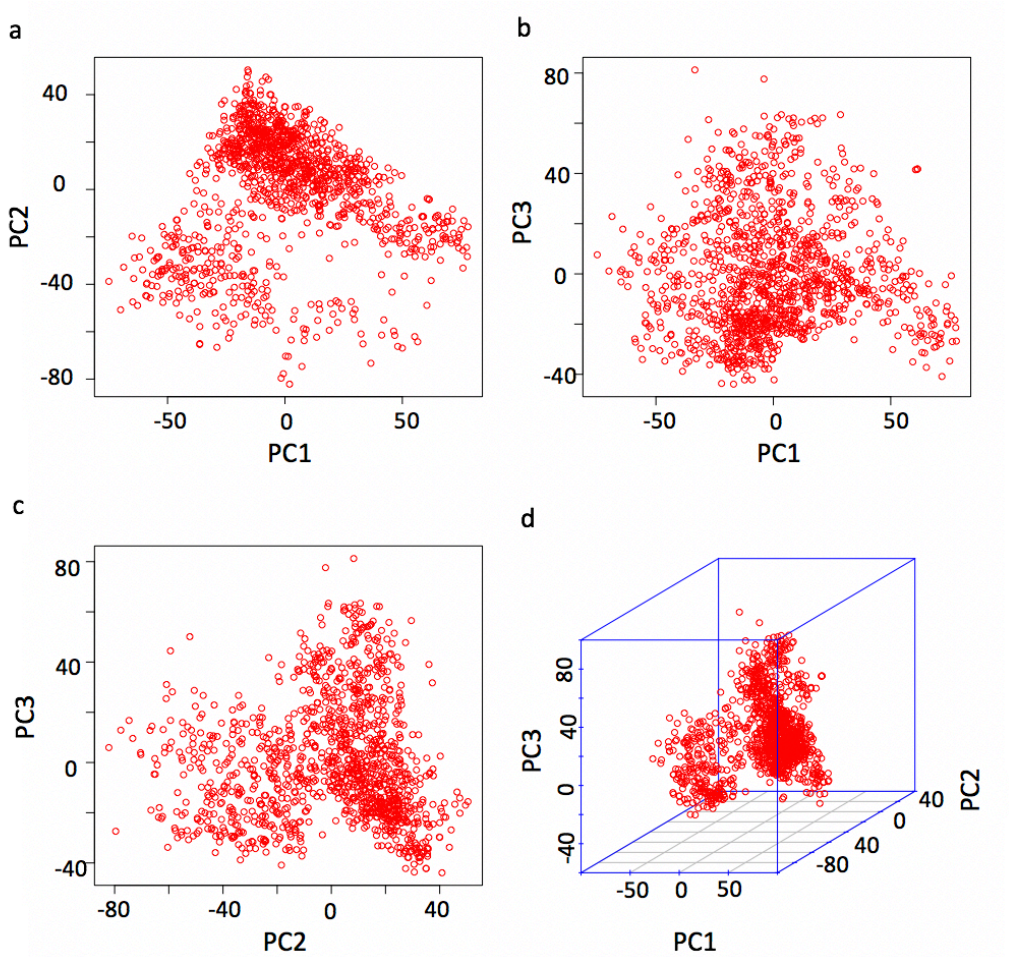


**Figure 2. Population structure demonstrated by principal component analysis**. Principal Component Analysis (PCA) was conducted with the 124,743 SNPs for the 1,220 cows. The population structure is demonstrated by the pairwise scatter plots (a, b, and c) and the 3D plot (d) of the first three principal components (PCs).

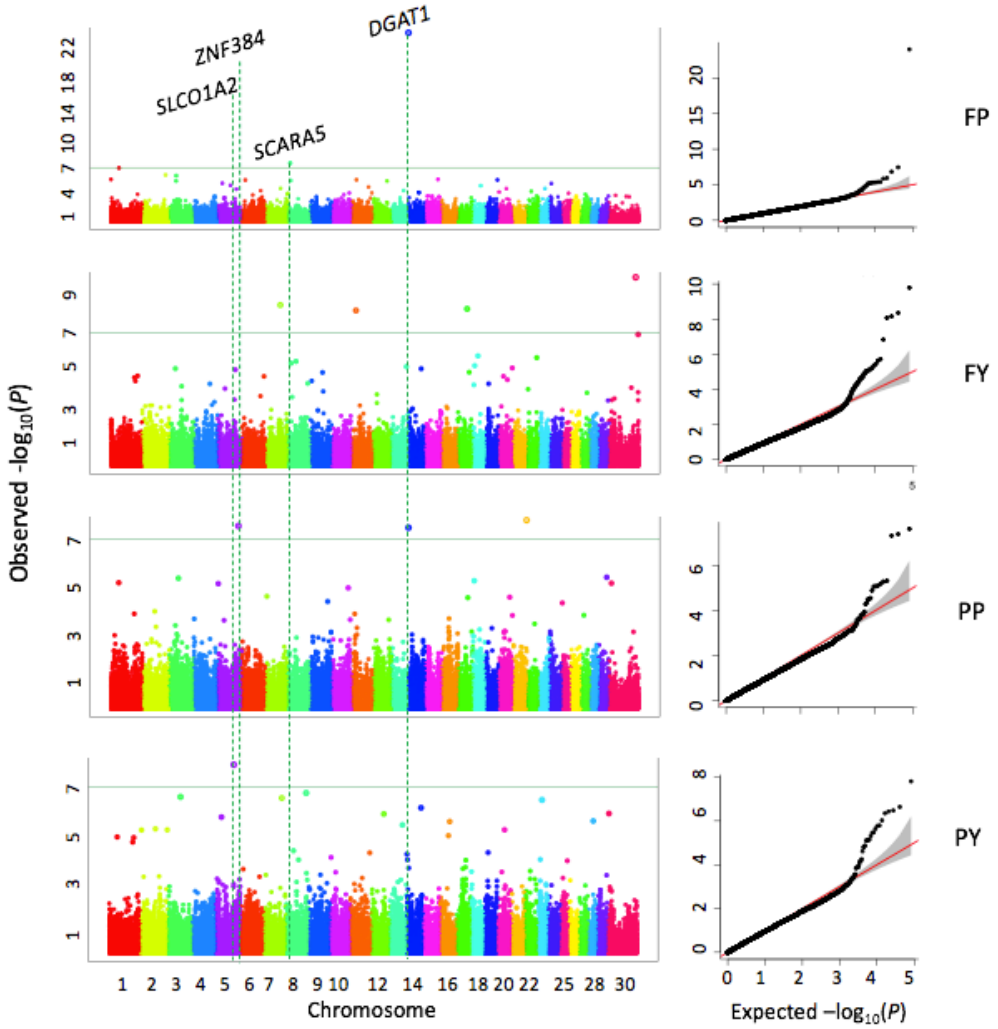


**Figure 3. Associations between 124,743 SNPs and milk traits.** Milk traits include fat yield (FY), protein yield (PY), fat percentage (FP), protein percentage (PP). The association analyses were conducted by the FarmCPU R package. Manhattan plots display the negative logarithms of the observed *P* values for SNPs across 30 bovine chromosomes (left panel). The green line indicates the Bonferroni multiple test threshold at *P* = 4.0e-07. The Quantile-Quantile (QQ) plots represent the negative logarithms of the expected *P* values (X-axis) and observed *P* values (Y-axis) (right panel).

1 **Table 1. Variance components and heritability of milk traits***

| Variance component | MY | FP | FY | PP | PY |
|---|---|---|---|---|---|
| Genetic | 1592.62 | 4.76 | 1.48 | 1.65 | 1.20 |
| Permanent environmental | 4267.93 | 10.58 | 5.46 | 3.45 | 3.96 |
| Residual | 7413.63 | 0.49 | 0.08 | 0.07 | 0.03 |
| $h^2$ | 0.12 | 0.30 | 0.21 | 0.32 | 0.23 |

2 *MY, milk yield; FP, fat percentage; FY, fat yield; PP, protein percentage, PY, protein yield.
3
4 **Table 2. Genome-wide significant SNPs associated with milk traits***

| Traits | SNP | CHR | Position (bp) | MAF | Nearest gene | Distance (kb) | *P*-value |
|---|---|---|---|---|---|---|---|
| FP | rs42295213 | 1 | 41,061,715 | 0.36 | *EPHA6* | within | 1.50E-07 |
| PP | rs109875012 | 5 | 104,120,905 | 0.45 | *ZNF384* | 2.6 | 4.03E-08 |
| PY | rs134480235 | 5 | 89,267,320 | 0.49 | *SLCO1A2* | within | 1.57E-08 |
| FY | rs43526055 | 7 | 73,431,219 | 0.31 | *ADRA1B* | 179.8 | 4.48E-09 |
| FP | rs136949224 | 8 | 10,705,865 | 0.14 | *SCARA5* | 43.5 | 3.57E-08 |
| FY | rs137676276 | 11 | 19,277,448 | 0.11 | *VIT* | 25 | 8.58E-09 |
| FP, PP | rs109421300 | 14 | 1,801,116 | 0.23 | *DGAT1* | within | 9.92E-25 |
| FY | rs109528658 | 17 | 46,090,458 | 0.40 | *EP400* | within | 7.05E-09 |
| PP | rs108996837 | 21 | 69,386,346 | 0.12 | *EXOC3L4* | 32.6 | 2.36E-08 |
| FY | rs135780687 | X | 134,726,985 | 0.42 | *GRPR* | 83 | 1.63E-10 |

5 *FP, fat percentage; PP, protein percentage; PY, protein yield; FY, fat yield.
6
7