Genome-wide association analysis of milk production, somatic cell score, and body conformation traits in Holstein cows

Peng Wang1, Xue Li2,3†, Yihao Zhu1, Jiani Wei4, Chaoxin Zhang2,3, Qingfang Kong1, Xu Nie1, Qi Zhang5 and Zhipeng Wang2,3,*

1Heilongjiang Animal Husbandry Service, Harbin, China, 2College of Animal Science and Technology, Northeast Agricultural University, Harbin, China, 3Bioinformatics Center, Northeast Agricultural University, Harbin, China, 4School of mathematics, University of Edinburgh, Edinburgh, United Kingdom, 5College of Animal Science and Technology, China Agricultural University, Beijing, China

Milk production and body conformation traits are critical economic traits for dairy cows. To understand the basic genetic structure for those traits, a genome-wide association study was performed on milk yield, milk fat yield, milk fat percentage, milk protein yield, milk protein percentage, somatic cell score, body form composite index, daily capacity composite index, feed, and leg conformation traits, based on the Illumina Bovine HD100k BeadChip. A total of 57, 12 and 26 SNPs were found to be related to the milk production, somatic cell score and body conformation traits in the Holstein cattle. Genes with pleiotropic effect were also found in this study. Seven significant SNPs were associated with multi-traits and were located on the PLEC, PLEKHA5, TONSL, PTGER4, and LCORL genes. In addition, some important candidate genes, like GPAT3, CEBPB, AGO2, SLC37A1, and FNDC3B, were found to participate in fat metabolism or mammary gland development. These results can be used as candidate genes for milk production, somatic cell score, and body conformation traits of Holstein cows, and are helpful for further gene function analysis to improve milk production and quality.

KEYWORDS
milk production traits, body conformation traits, pleiotropic effect, genome-wide association study, Holstein cattle

Introduction

Milk is a source of nutrients essential for human growth and development. The milk production traits are important for the dairy industry. Body conformation traits have been applied in several countries with the development of dairy cattle breeding since they are closely related to the health (1), productivity (2), lifetime (3), and calving ease (4) of cows. Some studies have identified the genetic correlation between body conformation traits and first lactation milk yield to be between 0.48 and 0.54 (5). These correlations are therefore very important for the dairy industry to improve the milk production traits and body conformation traits.

OPEN ACCESS

EDITED BY
Xiangdong Ding,
China Agricultural University, China

REVIEWED BY
Yao Jiang,
China Agricultural University, China
Yuchun Pan,
Shanghai Jiao Tong University, China

CORRESPONDENCE
Zhipeng Wang
wangzhipeng@neau.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Veterinary Science

RECEIVED 29 April 2022
ACCEPTED 09 September 2022
PUBLISHED 04 October 2022

CITATION
Wang P, Li X, Zhu Y, Wei J, Zhang C, Kong Q, Nie X, Zhang Q and Wang Z (2022) Genome-wide association analysis of milk production, somatic cell score, and body conformation traits in Holstein cows. Front. Vet. Sci. 9:932034. doi: 10.3389/fvets.2022.932034

COPYRIGHT
© 2022 Wang, Li, Zhu, Wei, Zhang, Kong, Nie, Zhang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
The rapid development of sequencing technology has revealed the cause variants of complex traits using genome-wide association analysis (GWAS). A study by Schennink et al. (6) has revealed DGAT1 and SCD1 to be highly associated with the composition of milk-fat (long-chain fatty acid). Kiser et al. (7) verified the TFAP2A gene to be related to the production of colostrum in Jersey cattle. It reported the genes CDH2 and GABRG2 to be related to the milk fat percentage and milk protein traits, respectively, in dual-purpose Xinjiang brown cattle (8). Bouwman et al. (9) and Vanvanhossou et al. (10) have reported the VEPH1 gene to be associated with conformation. However, the identified genes have not explained all genetic variances. There is a need to continue the search for novel genes related to some quantitative traits.

This study conducted GWAS using the Illumina Bovine HD100k BeadChip, for identifying important candidate genes or variants related to milk production, somatic cell score, and body conformation traits. There was an expectation for discovering novel genetic variations or candidate genes.

Materials and methods

Animal population

This experiment involved 1,313 cows from 7 different pastures in Heilongjiang Province. The use and care of the animals in this study were approved by the Animal Care Advisory Committee, Northeast Agricultural University (Harbin, China), and all the experimental procedures were according to the university’s guidelines for animal research.

Genotypes data

The samples were collected from the tail roots near the hips of the cows. The DNA in the hair was extracted and genotyped using Illumina Bovine HD100k BeadChip, containing 95,256 SNPs. The markers with minor allele frequencies < 0.05 and call rates < 0.90 were filtered out and individuals with a call rate of 0.80 or greater were selected. These SNPs were distributed across 29 chromosomes.

Population stratification

The SNP genotypes of these individuals were used to estimate the population stratification based on principal component analysis (PCA), and Plink (version 1.9) (11) was used to analyze a total of 1,310 cows with 86,645 markers covering the whole genome to study the population structure (12). The software uses the default matrix construction method to construct G matrix and get the PCA results. We used R language (version 4.1.2)—ggplot 2 package to draw pictures. The PCA scatterplots (Figure 1) illustrate a clear population structure for the 1,310 individuals in the seven pastures cattle herds that comprised our study population.

Genome-wide association analysis

Combination with dairy herd improvement data of National Holstein cows in China, this study estimated the genomic estimated breeding values (GEBVs) of all animal milk production traits, somatic cell score, and body conformation traits, using single-step genomic best linear unbiased prediction (ssGBLUP). The ssGBLUP was developed to integrate all the information including genotypes, phenotypes, and pedigree information in one step, and each SNP effect was calculated using the FarmCPU method (13) based on the predicted GEBVs. The ssGBLUP method is an improvement of BLUP, in which the pedigree relationship matrix $A^{-1}$ matrix must be replaced by $H^{-1}$ (14). The specific model is as follows:

\[ y = Xb + Zu + e \]

Where $y$ was each phenotypic value vector; $b$ is the fixed effect of the field and the PCA effect to explain the population stratification, and $u$ is a vector of animal effects. The $e$ was a vector of random residual effects with $e \sim N(0,I)$, and $X$, $Z$ were...


incidence matrices for b and u, respectively.

\[
H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ \tau(1 - w)\alpha + b & 0 \end{bmatrix} \begin{bmatrix} 1 - \omega A_{22}^{-1} - \omega A_{22}^{-1} \\ w \end{bmatrix}
\]

Where, the A matrix is pedigree relationship matrix, A_{22} is a numerator relationship matrix for genotyped animals, and G is a genomic relationship matrix (15). G^{-1} was obtained as the inverse of a combination of the G matrix and the corresponding A matrix. The w is the weight of A_{22} in the matrix, the default value is 0.05. The \( \tau \) and \( \omega \) are 1. We use DMU software to calculate the GEBV value. Both G and H matrices were derived using software default parameter setting by DMU software. G was calculated as:

\[
G = \frac{WDW'}{2\sum_{i=1}^{n}p_i(1 - p_i)}
\]

Where \( p_i \) is the allele frequency at locus i in all genotyped animals, is a normalizing constant (16) that sums expected variances across markers scaling G toward the A matrix (17). D is weight for each locus (I if same variance assumed), W is a design matrix as follows:

\[
\begin{cases}
0 - 2p_i, \text{ homozygous} \\
1 - 2p_i, \text{ heterozygous} \\
2 - 2p_i, \text{ homozygous}
\end{cases}
\]

Each SNP effect was calculated using the FarmCPU method (13) based on the predicted GEBVs. The FarmCPU method (13) in this study can be written as two models.

\[
y = SNP_i + K + e \\
y = pseudQTN + SNP_i + e
\]

The y is the GEBV value. The pseudQTN is significant marker from previous loops that is null when the model begins. SNP is testing marker in each loop. The K is the kinship between each individuals. The e is residual vector.

For each trait, the threshold P-value for genome-wide significance was 5.99 \times 10^{-7} = 0.05/83446 using the Bonferroni multiple test method.

QTLs annotation analysis

The cattle QTL data were downloaded from the Cattle QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index) referred to as the ARS-UCD1.2 assembly. The square of the correlation coefficient (\( r^2 \)) between the two loci is used to evaluate the range of LD measurement, because \( r^2 \) is considered to be more robust and not affected by changes in allele frequency and population size (18). Haploview software was used to calculate the genotype correlation coefficient (\( r^2 \)) between all SNP pairs in the cow population to estimate the LD of the whole genome, and the LD decay map with distance of the cow population was visualized.

Results

Population stratification

The phenomenon of group stratification is an important research problem in the study of group association (19). In order to determine the population stratification level, we drew the population structure by principal component analysis (PCA). The PCA scatterplots shows the population structure of a 1,300 individual composed of seven pastures (Figure 1). Different colors represent different pastures. It can be seen that it is mainly divided into three clusters, but most of the cows in the seven pastures are gathered together, and only a few cows are separated. These clusters indicate that, although individuals may come from different ranches, they still retain close genetic relationships.

The genome-wide association study

Basic descriptive statistics of milk production traits, somatic cell score and body conformation traits (see Table 1). A total of 86,645 SNPs were retained after quality control for the GWAS (Table 2). The average physical distance between the adjacent SNP markers was approximately 29.58 kb, ranging between 26.37 kb (BTA19) and 32.02 kb (BTA8).

The \( p \)-value profiles of all the SNP markers associated with each trait are represented in Figures 3, 4 and included the Manhattan and Quantile-Quantile plots. In total, 95 genome-wide significant SNPs were detected for the milk production traits, such as milk yield (MY), milk fat yield (FY), milk fat percentage (FP), milk protein yield (PY), milk protein percentage (PP), somatic cell score (SCS), and body conformation traits (body form composite index, BFCI; daily capacity composite index, DCCI; feed and leg conformation, FTLEG). There were 57, 12 and 26 SNPs related to milk production, somatic cell score and body conformation traits, respectively. Among them, we mainly focused on the first few significant SNPs in each trait. In addition, we also found seven SNPs that overlap with multiple traits, such as PLEC is related to MY, FP and PP, PLEKH5 is related to FP and FY, TONSL is connected with FY and SCS, LECORL is correlated with DCCI and FTLEG, PYGB is related to BFCI and FTLEG, and PTGER4 is related to BFCI, and DCCI (see Table 3).

As shown in Tables 4, 5, 12, 11, 15, 11, 12, and 17 genome-wide significant SNPs were detected for MY, FY, FP, PY, PP, and
TABLE 1 Descriptive statistics of milk production traits and body conformation traits.

| Statistic       | MY (kg)   | FP (%)  | PP (%)  | SCS  | BFCI  | FTLEG | DCCI  |
|-----------------|-----------|---------|---------|------|-------|-------|-------|
| Mean            | 8382.99   | 3.85    | 3.31    | 4.02 | 85.68 | 85.46 | 85.91 |
| Standard Deviation | 1950.68  | 0.50    | 0.26    | 1.44 | 4.88  | 4.16  | 7.68  |
| Minimum         | 1505.00   | 2       | 2.17    | 1.00 | 65.25 | 65.80 | 56.18 |
| Maximum         | 15983.00  | 6.20    | 5.00    | 9.00 | 98.36 | 99.00 | 99.95 |
| Coefficient of Variation | 0.23 | 0.13 | 0.08 | 0.36 | 0.06 | 0.05 | 0.09 |

TABLE 2 Distribution of SNPs after quality control.

| BTA | Length (Mb) | No. SNP (Chip data) | No. SNP (after QC) | Density length/SNP(kb) |
|-----|-------------|----------------------|--------------------|-----------------------|
| 1   | 158.53      | 5556                 | 5188               | 30.56                 |
| 2   | 136.23      | 4688                 | 4367               | 31.20                 |
| 3   | 121.01      | 4508                 | 4158               | 29.10                 |
| 4   | 120.00      | 4049                 | 3760               | 31.92                 |
| 5   | 120.09      | 4523                 | 4083               | 29.41                 |
| 6   | 117.81      | 4364                 | 3977               | 29.62                 |
| 7   | 110.68      | 3903                 | 3551               | 31.17                 |
| 8   | 113.32      | 3805                 | 3539               | 32.02                 |
| 9   | 105.45      | 3695                 | 3469               | 30.40                 |
| 10  | 103.31      | 3626                 | 3376               | 30.60                 |
| 11  | 106.98      | 3801                 | 3522               | 30.38                 |
| 12  | 87.22       | 3044                 | 2842               | 30.69                 |
| 13  | 83.47       | 3064                 | 2822               | 29.58                 |
| 14  | 82.40       | 3045                 | 2796               | 29.47                 |
| 15  | 85.01       | 3119                 | 2885               | 29.47                 |
| 16  | 81.01       | 2826                 | 2586               | 31.33                 |
| 17  | 73.17       | 2668                 | 2506               | 29.20                 |
| 18  | 65.82       | 2605                 | 2389               | 27.55                 |
| 19  | 63.45       | 2726                 | 2406               | 26.37                 |
| 20  | 71.97       | 2737                 | 2498               | 28.81                 |
| 21  | 69.86       | 2573                 | 2374               | 29.43                 |
| 22  | 60.77       | 2201                 | 2038               | 29.82                 |
| 23  | 52.50       | 2110                 | 1951               | 26.91                 |
| 24  | 62.32       | 2259                 | 2081               | 29.95                 |
| 25  | 42.35       | 1726                 | 1589               | 26.65                 |
| 26  | 51.99       | 1823                 | 1708               | 30.44                 |
| 27  | 45.61       | 1699                 | 1624               | 28.09                 |
| 28  | 45.94       | 1735                 | 1630               | 28.18                 |
| 29  | 51.10       | 1871                 | 1731               | 29.52                 |
| Total | 2489.37    | 90349                | 83446              | 29.58                 |

SCS, respectively. These significant SNPs are mainly distributed in BTA 1, BTA 2, BTA 5, BTA 6, BTA 11, BTA 14, and BTA 20, with as many as 9 SNPs on BTA 14.

In addition, this study reported an interesting phenomenon where four SNPs were found to be related to multi-traits, including BovineHD0500025853 (BTA 5:90.66 Mb), BovineHD1400000206 (BTA 14:0.49 Mb), BovineHD1400000287 (BTA 14:0.88 Mb), and BovineHD1400011649 (BTA 14:38.57 Mb) (see Table 6). The bovinehd1400000287 SNP located in the 58th intron of the PLEC gene was found to be associated with MY, FP, and PP. The fat yield and the somatic cell score trait shared one SNP bovinehd140000206 located 1.46 kb away from TONSL on BTA 14.

This study detected 10, 7, and 11 significant SNPs related to BFCI, DCCI, and FTLEG, respectively. There were 4 SNPs distributed on BTA 16. Three SNPs were found to be possibly as pleiotropism SNPs, including BovineHD4100004660 (BTA 6:38.22 Mb), BovineHD1300012605 (BTA 13:42.81 Mb) and BTA-50244-no-rs (BTA 20:34.30 Mb), respectively. Of these significant SNPs, the BTA-50244-no-rs SNP related to BFCI ($P = 5.84E-13$) was located downstream of the PTGER4 gene.

QTL annotation analysis

The LD of cows decreases with the increase of distance, when the distance is extended to 200 Kb, the
Manhattan plots and Quantile-Quantile plots for the milk production and somatic cell score traits. MY (A,B), FY (C,D), FP (E,F), PP (G,H), PY (I,J), and SCS (K,L).

FIGURE 3
The decline rate of LD of cows tends to be gentle, and the average $r^2$ value of cows is 0.3 at this time (shown as Figure 2). The 100 Kb range of SNP upstream and downstream of significant trait association obtained from genome-wide association analysis is compared with the data that has been verified in the current cattle QTL database. Our significant SNPs associated with MY, FY, FP, PP and SCS overlapped with 1332, 1177, 3042, 1288, 24 QTLs, respectively. But there are also very few QTLs about body conformation traits overlapped with significant SNPs.

**Figure 4**
Manhattan plots and Quantile-Quantile plots for the body conformation traits: BFCI (A,B), DCCI (C,D) and FTLEG (E,F).

### Discussion

#### Comparison with the other GWAS studies

In this study, FarmCPU was applied for screening the QTLs related to the milk production traits, health traits, and body conformation traits. A total of 95 significant SNPs were detected, located on the 93 candidate genes. Of these genes, EHHADH, SLC37A1, PLEKHA5, TONS1, PLEC, and IL5RA were reportedly related to milk production traits in other studies (15, 20, 21, 24, 39). However, this study did not detect some important
TABLE 3 The SNPs and candidate genes with pleiotropic effect in this study.

| SNP name          | Traits | Gene       | Distance(kb)* | Gene full name | Gene function                                                                 |
|-------------------|--------|------------|---------------|----------------|--------------------------------------------------------------------------------|
| BovineHD1400000287 | MY, FP, PP | PLEC       | Intron         | Plectin        | Related to the MY, FP, and PP traits in Chinese Holsteins (20).                |
| BovineHD0500025853 | FP, FY  | PLEKHA5    | Intron         | Pleckstrin homology domain containing, family A member 5 | Significantly associated with FP (21).                                       |
| BovineHD1400011649 | MY, PY  | HNF4G      | Intron         | Hepatocyte nuclear factor 4 gamma | Associated with childhood obesity (22). Key regulators of beef cattle carcass IMF (23). |
| BovineHD1400000206 | FY, SCS | TONSL      | 1.65 (U)       | Tonsoku like, DNA repair protein   | Related to milk yield (24, 25) and affect the gamma–linolenic acid, long-chain saturated fatty acids and milk fat percent of the Canadian Holstein cows (26). |
| BovineHD4100004660 | DCCI, FTLEG | LCORL     | 665.01 (D)    | Ligand dependent nuclear receptor corepressor like | Affect human height (27), pig body length (28), chicken carcass weight (30), and the growth and development of cattle (31). Associated with the human skeletal frame size (32). |
| BovineHD1300002605 | BFCL, DCCI | PYGB      | 0.61 (D)       | Glycogen phosphorylase B | Inhibition of glycogen utilization (33).                                       |
| BTA−50244-no−rs  | BFCL, DCCI | PTGER4    | 541.63 (D)    | Prostaglandin E receptor 4 | Relaxation to the smooth muscle (34), leading to the phosphorylation of glycogen synthase kinase−3 (35), involving in osteoporosis (36), and regulating lipid droplet size and mitochondrial activity in the white adipose tissue (37, 38). |

*U, Upstream; D, Downstream.

candidate genes, such as DGAT1. Because in this study, the closest SNP on both flanks of DGAT1 are BovineHD1400000206 (109.2 kb) and ARS-BFGL-NGS-55227 (50.8 Mb), respectively. Of these, BovineHD1400000206 associated with fat yield (P value = 2.76E−17). But the nearest gene on this significantly SNP is the TONSL gene (1.65 kb), which is a neighboring gene to DGAT1. So, the DGAT1 gene was not detected in this study. The study by Ning et al. (40) used two models and a 70k SNP chip based on the Chinese Holsteins population and identified the DGAT1 gene to be related to milk (40). Kim et al. (41) also obtained DGAT1 affecting MY and FY in the Korean cattle population (41). Cole et al. (42) identified the PHKA2 gene to be highly significant for four body size traits (stature, strength, body depth, rump width) (42). The 770k BeadChip was used by An et al. (43) to identify five candidate genes (CSMD3, LAP3, SYN3, FAM19A5, and TIMP3) related to the body conformation traits. This study did not detect the above genes to be associated with body conformation traits.

These inconsistencies might be due to differences in the detection platforms or algorithms used in the corresponding analysis, changes in the genetic background of the analyzed cattle, differences in the size and structure of the study population, or random or technical errors in some analyses. This also indicated that there are many important genetic markers or candidate genes in the bovine genome that are yet to be discovered.

Genetic analysis of pleiotropic genes

Organisms have hundreds of thousands of genes and tens of thousands of phenotypes. The relationship between genes and epigenetic factors is complex. There are various associations such as pleiotropism, multigenic effect, polygene effect and so on. Pleiotropy is defined as the phenomenon where a single locus affects two or more distinct phenotypic traits (44, 45). It
Table 4: Genome-wide significant SNPs are associated with milk production traits.

| Traits  | SNP name | BTA | Position (Mb) | MAF  | Nearest gene | Distance (kb)* | P-value  | SNP effect |
|---------|----------|-----|---------------|------|--------------|---------------|----------|------------|
| MY      | BTB–00089434 | 2   | 33.86 | 0.0599 | KCNH7 | 39.78 (U) | 5.87E–08 | −295.1002 |
| MY      | BovinHD1700012968 | 17 | 45.45 | 0.2130 | SF51 | 90.47(D) | 3.53E–07 | −153.0319 |
| MY      | BovinHD3500014407 | 15 | 49.22 | 0.3053 | OR51L4 | 0.17(U) | 1.23E–07 | −155.6556 |
| MY      | BovinHD1700012968 | 17 | 45.45 | 0.2130 | SF51 | 90.47(D) | 3.53E–07 | −153.0319 |
| FY      | BovinHD2800000275 | 28 | 1.65 | 0.2076 | UBD2 | 45.14(D) | 2.75E–07 | −164.7354 |
| FY      | BovinHD2300011633 | 23 | 40.58 | 0.1126 | ZNF94F | 7.11E–07 | 3.37E–08 | 3.4575 |
| MY      | BovinHD1700012968 | 17 | 45.45 | 0.2130 | SF51 | 90.47(D) | 3.53E–07 | −153.0319 |
| FY      | BovinHD1400000206 | 14 | 4.9 | 0.3182 | TONSL | 0.17(U) | 1.23E–07 | −155.6556 |
| MY      | BovinHD1400002871 | 14 | 0.88 | 0.2240 | PLEC | 2.71E–07 | 3.4575 |
| FY      | BovinHD1700012968 | 17 | 45.45 | 0.2130 | SF51 | 90.47(D) | 3.53E–07 | −153.0319 |
| MY      | BovinHD1300016576 | 11 | 53.86 | 0.4229 | – | – | 1.28E–09 | −0.0479 |
| FY      | BovinHD1400000287 | 14 | 0.88 | 0.2240 | PLEC | 2.71E–07 | 3.4575 |
| MY      | BovinHD1300016576 | 11 | 53.86 | 0.4229 | – | – | 1.28E–09 | −0.0479 |
| FY      | BovinHD1700012968 | 17 | 45.45 | 0.2130 | SF51 | 90.47(D) | 3.53E–07 | −153.0319 |
| MY      | BovinHD1700012968 | 17 | 45.45 | 0.2130 | SF51 | 90.47(D) | 3.53E–07 | −153.0319 |
| FY      | BovinHD1400002871 | 14 | 0.88 | 0.2240 | PLEC | 2.71E–07 | 3.4575 |
| MY      | BovinHD1300016576 | 11 | 53.86 | 0.4229 | – | – | 1.28E–09 | −0.0479 |
| FY      | BovinHD1700012968 | 17 | 45.45 | 0.2130 | SF51 | 90.47(D) | 3.53E–07 | −153.0319 |
| MY      | BovinHD1300016576 | 11 | 53.86 | 0.4229 | – | – | 1.28E–09 | −0.0479 |
| FY      | BovinHD1700012968 | 17 | 45.45 | 0.2130 | SF51 | 90.47(D) | 3.53E–07 | −153.0319 |
| MY      | BovinHD1300016576 | 11 | 53.86 | 0.4229 | – | – | 1.28E–09 | −0.0479 |

(Continued)
is common in nature. For example, the \textit{DGAT1} gene is related to milk yield (40) and fat yield (26, 41). The genes \textit{PIK3R6} and \textit{PIK3R1} showed direct functional associations with height and body size (10). Production and health constitute fundamental dairy functions while body conformation traits are related to the functionality of the cow’s body. So, the milk production traits and body conformation traits of dairy cows tend to complement each other. Certain identified regions related to conformation traits overlap with the performance traits such as reproduction (24), and milk production (47). Some genes in these regions were also involved in regulating the cell cycle or cell division, homeostasis, and lipid metabolism (10).

This study also reported this interesting phenomenon where the \textit{PLEC}, \textit{PLEKHA5}, and \textit{TONSL} genes were found to belong to the pleiotropism gene for milk traits, and the \textit{LCORL}, and \textit{PTGER4} were pleiotropic genes for the body conformation traits. The \textit{PLEC} gene (Plectin) can interlink different elements of the cytoskeleton. The \textit{PLEC} gene was found to be associated with multiple traits, like MY, FP, and PP. Dan Wang et al. (20) also detected \textit{PLEC} to have potential effects on the MY, FP, and PP traits, which could be useful for molecular breeding for milk production in Chinese Holsteins. The \textit{PLEKHA5} gene, located on BTA 5, was predicted to enable the activity of phosphatidylinositol phosphate (48), and milk production (47). Some genes in these regions were found to be significantly associated with FP using two different methods using 294,079 Holstein cows. The \textit{TONSL} protein was considered to be an NF-

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
\textbf{Traits} & \textbf{SNP name} & \textbf{BTA} & \textbf{Postion (Mb)} & \textbf{MAF} & \textbf{Nearest gene} & \textbf{Distance(kb)} & \textbf{P-value} & \textbf{SNP effect} \\
\hline
\textit{PP} & BovineHD20000009361 & 20 & 32.69 & 0.2844 & OCXCT1 & Intron & 5.35E-08 & -0.0294 \\
\textit{PP} & BovineHD2500004479 & 25 & 15.71 & 0.1275 & XYL1T & Intron & 1.78E-10 & 0.0365 \\
\textit{PY} & BovineHD1000027261 & 1 & 95.18 & 0.3279 & FNDCSB & Intron & 1.20E-11 & 3.3681 \\
\textit{PY} & BovineHD3000017107 & 3 & 56.64 & 0.3195 & LMO4 & Intron & 54.14(U) & 1.26E-07 & 2.5817 \\
\textit{PY} & Hapmap26317-BTC–059618 & 6 & 80.53 & 0.1557 & EPHA5 & Intron & 306.96(U) & 4.69E-08 & -3.2564 \\
\textit{PY} & ARS-BFGL-NGS–4974 & 11 & 106.56 & 0.1069 & ZMYND19 & 5.22(D) & 4.52E-09 & 2.4183 \\
\textit{PY} & BovineHD1400011649 & 14 & 38.57 & 0.1725 & CRISPLD1 & Intron & 226.89(D) & 4.93E-08 & 3.2889 \\
\textit{PY} & BovineHD1700016449 & 17 & 55.89 & 0.4740 & CCDC60 & Intron & 1.62E-07 & -2.2566 \\
\textit{PY} & ARS-USMARC-Parent-EF034086–no-rs & 26 & 37.90 & 0.4607 & EMX2 & Intron & 63.30(D) & 4.29E-09 & 2.6483 \\
\hline
\textit{SCS} & Hapmap59481–rs29019616 & 1 & 56.94 & 0.1893 & GCSAM & Intron & 1.59E-10 & -0.1547 \\
\textit{SCS} & BovineHD0200033155 & 2 & 113.94 & 0.2847 & NYAP2 & Intron & 90.45(D) & 1.98E-08 & 0.1198 \\
\textit{SCS} & BovineHD0600020300 & 6 & 71.35 & 0.3580 & CEP135 & Intron & 6.10(U) & 1.94E-10 & -0.1304 \\
\textit{SCS} & BovineHD1100011547 & 11 & 39.19 & 0.2080 & CCDC68 & Intron & 118.86(D) & 3.15E-08 & 0.1376 \\
\textit{SCS} & BovineHD1300019252 & 13 & 67.15 & 0.2282 & KIAA1755 & Intron & 21.69(D) & 1.73E-09 & -0.1423 \\
\textit{SCS} & BovineHD1400000206 & 14 & 0.49 & 0.2172 & TONS1 & Intron & 1.65(U) & 4.03E-07 & 0.1159 \\
\textit{SCS} & BovineHD1400011508 & 14 & 38.01 & 0.2939 & PI15 & Intron & 147.13(U) & 3.10E-09 & -0.1307 \\
\textit{SCS} & BovineHD1600013229 & 16 & 47.05 & 0.3996 & ACOT7 & Intron & 7.12E-08 & 0.1065 \\
\textit{SCS} & BovineHD1600015783 & 16 & 53.27 & 0.0657 & SERPINCI1 & Intron & 2.54E-08 & -0.2208 \\
\textit{SCS} & BTA–65815–no-rs & 16 & 59.73 & 0.2267 & RASAL2 & Intron & 2.31E-10 & -0.1528 \\
\textit{SCS} & UA-JFASA–5305 & 19 & 59.21 & 0.1271 & SOX9 & Intron & 289.73(D) & 3.61E-09 & -0.1751 \\
\textit{SCS} & BovineHD2000017315 & 20 & 61.61 & 0.4405 & CTNND2 & Intron & 11.32(U) & 1.18E-07 & 0.1036 \\
\textit{SCS} & BTA–52343–no-rs & 21 & 42.73 & 0.1042 & AKAP6 & Intron & 4.65E-09 & -0.1536 \\
\textit{SCS} & Hapmap46118-BTA–108252 & 22 & 19.45 & 0.4435 & GRM7 & Intron & 3.00E-08 & -0.1121 \\
\textit{SCS} & ARS-BFGL–NGS–24519 & 25 & 10.59 & 0.1378 & GSPT1 & Intron & 0.93(D) & 2.31E-07 & 0.1394 \\
\textit{SCS} & ARS-BFGL–NGS–37189 & 25 & 32.40 & 0.07786 & RCC1L & Intron & 267.69(U) & 3.28E-08 & 0.1948 \\
\textit{SCS} & Hapmap42542-BTA–40776 & 26 & 27.93 & 0.2504 & SORCS1 & Intron & 20.95(D) & 5.60E-10 & 0.1395 \\
\hline
\end{tabular}
\caption{Genome–wide significant SNPs are associated with somatic cell score.}
\end{table}
affect the gamma-linolenic acid, long-chain saturated fatty acids and milk fat percent of the Canadian Holstein cows (26).

Interesting, the TONSL gene is a neighboring gene to DGAT1 (flanking < 200 kb), associated with the fat percentage of milk (26).

Some studies on the LCORL gene showed it to affect human height (27), pig body length (28), horse height (29), chicken carcass weight (30), and the growth and development of cattle (31). This gene might have been a novel loci associated with the human skeletal frame size (32). PTGER4 encodes a protein that is one of the members of the G-protein coupled receptor family, which imparts relaxation to the smooth muscle (34), leading to the phosphorylation of glycogen synthase kinase-3 (35), involved in osteoporosis (36), and regulating lipid droplet size and mitochondrial activity in the white adipose tissue (37, 38).

Important candidate genes related to the fat metabolism or mammary gland development

Fatty acids are essential components of milk with known positive associations with human cardiovascular diseases and so on. This study identified genes such as GPAT3, ARNTL2, EHHADH, CEBPB, DNAJB9, ZNF496, AOG2, GALNT18, and NEGR1 as critical for obesity traits or adipose metabolism (see Table 7).

GPAT3 is highly expressed in the adipose tissue with an important role in adipogenesis (50). This gene can be regulated by folic acid for controlling lactation and metabolic function of the dairy cows (51) and is also involved in fat and lipid metabolism in the Yunling cattle (52). EHHADH involved in fatty acid oxidation is essential for producing medium-chain fatty acids.
TABLE 7 Important candidate genes related to the fat metabolism or mammary gland development.

| Gene name | Location (BTA:Start–End, Mb) | Full name | Gene function |
|-----------|-----------------------------|-----------|---------------|
| GPAT3     | 6:98.29–98.36               | Glycerol-3-phosphate acyltransferase 3 | Highly expressed in the adipose tissue with an important role in adipogenesis (56). Can be regulated by folic acid for controlling lactation and metabolic function of the dairy cows (51). Involved in fat and lipid metabolism in the Yunling cattle (52). |
| ARNTL2    | 5:82.47–82.55               | Aryl hydrocarbon receptor nuclear translocator like 2 | Influencing Mexican–Mestizo childhood obesity (53). |
| EHHADH    | 1:81.88–81.93               | Enoyl-CoA hyd ratease and 3-hydroxyacyl CoA dehydrogenase | Involved in fatty acid oxidation is essential for producing medium-chain dicarboxylic acids (54). Impact on the characteristics of milk fatty acid traits in Chinese Holstein (55). A pivotal gene in the fat-related pathway (56). |
| CEBPB     | 13:78.20–78.21              | CCAAT enhancer binding protein beta | Involved in regulating the expression of fatty acid synthase in dairy cow mammary epithelial cells and milk fat synthesis (57). |
| DNAJB9    | 4:59.58–59.59               | DNAJ heat shock protein family (Hsp40) member B9 | The prognostic biomarkers of breast cancer (58). Correlated with the abdominal fat weight (59). |
| ZNF496    | 7:40.57–40.61               | Zinc finger protein 496 | Associated with milk fat and fertility (60). |
| AGO2      | 14:3.06–3.14                | Argonaute RISC catalytic component 2 | Related to mitochondrial oxidation and obesity–associated pathophysiology (61). |
| GALNT18   | 15:41.06–41.42              | Polypeptide N-acetylgalactosaminyltransferase 18 | Associated with milk protein and fat traits (62). |
| NEGR1     | 3:72.81–73.84               | Neuronal growth regulator 1 | Associated with obesity and BMI (body mass index) (63–65). |
| SLC37A1   | 1:142.81–142.87             | Solute carrier family 37 member 1 | Over-expressed in the bovine mammary tissue (66). Increases milk yield, decreases phosphorus concentration (66). |
| FNDC3B    | 1:95.12–95.41               | Fibronectin type III domain containing 3B | Biomarker for the bovine mammary stem/progenitor cells, and Essential for the growth and maintenance of the mammary epithelium (67). |

U, Upstream; D, Downstream.

dicarboxylic acids (54). Hence, this gene has a key impact on the characteristics of milk fatty acid traits in Chinese Holstein (55). In porcine adipogenesis, EHHADH has been proposed to be a pivotal gene in the fat-related pathway (56). The DNAJB9 gene is reportedly one of the prognostic biomarkers of breast cancer (58). Interestingly, DNAJB9 and DNAJB6 are members of the DNAJ gene family, with sequence similarity. The expression level of DNAJB6 in the chicken abdominal adipose tissue was significantly negatively correlated with the abdominal fat weight (59). ZNF496 is reportedly associated with milk concentration (milk fat) and fertility (60). According to Gao et al. (62), the GALNT18 gene was associated with milk protein and fat traits.

According to the known gene functions, some candidate genes were expressed in the mammary gland, such as the SLC37A1 and FNDC3B genes (see Table 7). SLC37A1, over-expressed in the bovine mammary tissue relative to the 17 other tissue types (66) transports glucose-6-phosphate in one direction and phosphorus in the other (68). Glucose is known to be essential for lactose synthesis in mammary cells. Kemper et al. (66) identified the causative mutation increasing the expression of SLC37A1 leading to an increase in milk yield and decreasing the phosphorus concentration.

**QTLs result overlapped with GWAS**

Although many quantitative trait loci (QTLs) related to economically important traits in dairy cows have been identified, due to insufficient sample size and insufficient marker density used in QTL mapping research in history, not all genetic variations of these traits have been captured (69), in the study, we used GWAS to analyze the milk production traits, body conformation traits and somatic cells of dairy cows, and most of the results were also verified in the QTL analysis of dairy cows. Interestingly, our study found many SNP related to pleiotropy, but no repeated QTL regions were found in the QTL analysis (70). Also found the same phenomenon in the study of multiple traits of beef cattle. With these results, we can get some
inspiration in verifying QTLs of some characteristics of interest shared among varieties (71).

Conclusions

A total of 95 significant SNPs were identified to be related to the milk production, somatic cell score, and body conformation traits in Holstein cattle. Among them, 7 significant SNPs located on the PLEC, PLEKHA5, TONSL, PTGER4, and LCORL genes showed pleiotropic effects on milk production or body conformation traits. In addition, some important candidate genes, including GPAT3, CEBPB, AGO2, SLC37A1, and FNDC3B, were also found to be related to the fat metabolism or involved in mammary gland development. The above genes however need to be consolidated as new potential genes through future validation.

Data availability statement

The original contributions presented in the study are included in the article or supplementary material, the variation data reported in this article have been deposited in the Genome Variation Map (GVM) in Big Data Center, Beijing Institute of Genomics (BIG), and Chinese Academy of Sciences, under accession numbers GVM000388 that are publicly accessible at https://bigd.big.ac.cn/gvm/getProjectDetail?project=GVM000388. The Bioproject accession number is PRJCA011726. Further inquiries can be directed to the corresponding author.

Ethics statement

The animal study was reviewed and approved by Animal Care Advisory Committee, Northeast Agricultural University (Harbin, China).

Author contributions

ZW, PW, and XL conceived the study and participated in its design. YZ, JW, QK, and XN were involved in the acquisition of data. XL, JW and QZ performed all data analysis. XL and ZW drafted the manuscript. ZW, PW, XL, YZ, CZ, QK, and XN contributed to the writing and editing. All authors read and approved the final manuscript.

Funding

This work was financially supported by the Natural Science Foundation of China (No. 32070571), the Academic Backbone Project of Northeast Agricultural University (No. 15XG14), and the NEAU Research Founding for Excellent Young Teachers (2010RCB29).

Acknowledgments

The authors are grateful to all the members of the Dairy Breeding Group of the Heilongjiang Provincial Husbandry Department for collecting the data. We appreciate the support by the Project of Science and Education Office in the Department of Agriculture and Rural Affairs of Heilongjiang Province (Application of the genomic breeding technology for modernized large-scale dairy farm in Heilongjiang Province).

Conflict of interest

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

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Chapinal N, Koeck A, Sewalem A, Kelton DF, Mason S, Cramer G, et al. Genetic parameters for hoof lesions and their relationship with feet and leg traits in Canadian Holstein cows. J Dairy Sci. (2013) 96:2596–604. doi: 10.3168/jds.2012-6071
2. Lund T, Miglior F, Dekkers JCM, Burnside EB. Genetic relationships between clinical mastitis, somatic cell count, and udder conformation in Danish Holsteins. Livest Prod Sci. (1994) 39:243–51. doi: 10.1016/0301-6226(94)90203-8
3. Vollema AR, Van Der Beek S, Harbers AG, De Jong G. Genetic evaluation for longevity of Dutch dairy bulls. J Dairy Sci. (2000):2629–39. doi: 10.3168/jds.2002-0302(00)75156-3
4. Kadarmideen HN, Wegmann S. Genetic parameters for body condition score and its relationship with type and production traits in Swiss Holsteins. J Dairy Sci. (2003) 86:3685–93. doi: 10.3168/jds.S0022-0302(03)77974-5
5. Short TH, Lawlor TJ. Genetic parameters of conformation traits, milk yield, and herd life in Holsteins. J Dairy Sci. (1992) 75:1987–8. doi: 10.3168/jds.S0022-0302(92)77958-2
6. Schennink A, Stoop WM, Visser MH, van der Poel JJ, Bovenhuis H, van Arendonk JA. Short communication: genome-wide scan for bovine milk-fat composition. II Quantitative trait loci for long-chain fatty acids. J Dairy Sci. (2009) 92:4676–82. doi: 10.3168/jds.2008-1965
Gene and characterization of splicing variants of PLEKHA5 (Plekha5) during brain development. *Gene.* (2012) 492:270. doi: 10.1016/j.gene.2011.10.018

50. Khan MZ, Liu L, Zhang Z, Khan A, Wang D, Mi S, et al. Folic acid supplementation regulates milk production variables, metabolic associated genes and pathways in perinatal Holsteins. *J Anim Physiol Anim Nutr (Berl).* (2020) 104:483–92. doi: 10.1111/jpn.13313

51. Zhang F, Hanif Q, Luo X, Jin X, Zhang I, He Z, et al. Muscle transcriptome analysis reveal candidate genes and pathways related to fat and lipid metabolism in Yunling cattle. *Anim Biotechnol.* (2021) 7:1–8. doi: 10.1080/10495398.2021.2009846

52. Costa-Urrutia P, Colistro V, Jiménez-Osorio AS, Cárdenas-Hernández H, Solares-Tlapechco J, Ramirez-Alcántara M, et al. Genome-wide association study of body mass index and body fat in Mexican-Mestizo Children. *Genes (Basel).* (2019) 10:945. doi: 10.3390/genes10110945

53. Houten SM, Denis S, Argmann CA, Jia Y, Ferdinandusse S. Reddy JK, et al. Peroxisomal L-bifunctional enzyme (Ehhadh) is essential for the production of medium-chain dicarboxylic acids. *J Lipid Res.* (2012) 53:1296–303. doi: 10.1194/jlr.M024463

54. Li C, Sun D, Zhang S, Wang S, Wu X. Zhang Q, et al. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. *PLoS ONE.* (2014) 9:e96186. doi: 10.1371/journal.pone.0096186

55. Jin P, Wu X, Xu S, Zhang H, Li Y, Cao Z, et al. Differential expression of six genes and correlation with fitness traits in a unique broiler population. *Saudi J Biol Sci.* (2017) 24:945–9. doi: 10.1016/j.sjbs.2015.04.014

56. Golik M, Glick G, Rescher S, Shirak A, Ezra E, Zeron Y, et al. Differential expression of ruminant ZNF496 variants: association with quantitative trait affecting bovine milk concentration and fertility. *J Dairy Sci.* (2011) 94:2092–102. doi: 10.3168/jds.2010-3655

57. Bhattacharjee J, Borra VJ, Salem ES, Zhang C, Murakami K, Gill RK, et al. Hepatic aga2 regulates PPARα for oxidative metabolism linked to glycemic control in obesity and post bariatric surgery. *Endocrinology.* (2021) 162:bqab007. doi: 10.1210/endocr/bqab007

58. Gao Y, Jiang I, Yang S, Hou Y, Liu GE, Zhang S, et al. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. *BMC Genomics.* (2017) 18:265. doi: 10.1186/s12864-017-3636-3

59. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthoradottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet.* (2009) 41:18–24. doi: 10.1038/ng.274

60. Spelotes EK, Willer CJ, Berndt SI, Mondal KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* (2010) 42:937–48. doi: 10.1038/ng.686

61. Maji R, Manning S, Yousef S, Pucci A, Santini F, Karra E, et al. Contribution of 32 GWAS-identified common variants to severe obesity in European adults referred for bariatric surgery. *PLoS ONE.* (2013) 8:e70735. doi: 10.1371/journal.pone.0070735

62. Kemper KE, Reich CM, Bowman PJ, Vander Jagt CJ, Chamberlain AJ, Mason BA, et al. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breeding population leads to greater accuracy of across-breed genomic predictions. *Genet Sel Evol.* (2015) 47:29. doi: 10.1186/s12711-014-0074-4

63. Choudhary RK, Capuco AV. Expression of NRS5A2, NUP153, HNF4A, USP15 and FNDC3B is consistent with their use as novel biomarkers for bovine mammary stem/progenitor cells. *J Mol Histol.* (2021) 52:289–300. doi: 10.1007/s10735-020-09948-8

64. Pan CJ, Chen SY, Jun HS, Lin SR, Mansfield BC, Chou JY. SLC37A1 and SLC37A2 are phosphate-linked, glucose-6-phosphate antiporters. *PLoS ONE.* (2013) 6:e23157.

65. Saatchi M, Schnabel RD, Taylor JR, Garrick DJ. 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

66. Saatchi M, Beever JE, Decker JE, Faulkner DR, Freathy RC, Hansen SL, et al. QTLs associated with dry matter intake, metabolic mid-test weight, growth and feed efficiency have little overlap across 4 beef cattle studies. *BMC Genomics.* (2014) 15:1004. doi: 10.1186/1471-2164-15-1004

67. Tribout T, Crouseau P, Lefebvre R, Barbat A, Bousaha M, Fritz S, et al. Confirmed effects of candidate variants for milk production, udder health, and udder morphology in dairy cattle. *Genet Sel Evol.* (2020) 52:55. doi: 10.1186/s12711-020-00575-1