Journal of Animal and Feed Sciences, 35, 4, 2026, XX, https://doi.org/10.22358/jafs/219999/2026
The Kielanowski Institute of Animal Physiology and Nutrition, Polish Academy of Sciences, Jabłonna

ORIGINAL PAPER

ARTICLE IN PRESS

(final version of manuscript before type-setted PDF)

Genome-wide association analysis of milk yield, composition and somatic cell count in Holstein dairy cattle

Yunus Arzik[1], Jale Metin Kıyıcı[2,*], Bilal Akyüz[3], Korhan Arslan[3], Mehmet Kizilaslan[4], Mahmut Kaliber[2] and Mehmet Ulaş Çınar[2,5]

[1] Aksaray University, Faculty of Veterinary Medicine, Department of Animal Science, 68100, Aksaray, Türkiye
[2] Erciyes University, Faculty of Agriculture, Department of Animal Science, 38280, Kayseri, Türkiye
[3] Erciyes University, Faculty of Veterinary Medicine, Department of Genetics, 38280, Kayseri, Türkiye
[4] University of Wisconsin–Madison, Department of Animal and Dairy Sciences, Madison, WI 53706, USA
[5] Washington State University, College of Veterinary Medicine, Department of Veterinary Microbiology and Pathology, Pullman, WA 99164, USA

*** Corresponding author:**
e-mail: jalemetin@erciyes.edu.tr

Received: 6 January 2026
Revised: 26 February 2026
Accepted: 25 March 2026

Short title: Genome-wide association of milk traits in Holstein

ABSTRACT. This study investigated the genetic basis of milk production traits in Holstein dairy cattle using a genome-wide association study approach with high-density single nucleotide polymorphism (SNP) genotyping. A total of 1 431 observations from 147 Holstein cows reared in Türkiye were analysed. The cows were phenotyped for milk somatic cell count (SCC), fat, protein, lactose, solid-not-fat, freezing point, density, electrical conductivity, pH, and milk yield. After quality control, 79 769 SNPs were retained for association analysis. Significant associations were identified, with 3 SNPs exceeding genome-wide significance thresholds and 12 exceeding chromosome-wide thresholds. Key findings included SNPs

(rs109536476, rs110213695, rs135761188, rs109679567) associated with SCC were located in or near the *CBY3* and *CNOT6* genes on BTA7, while variants associated with both fat content and density were located in the *RAB27B* gene on BTA24, indicating potential pleiotropic effects. The results emphasise the importance of functional validation to clarify the biological roles of the identified SNPs and genes. The findings contribute to understanding the genetic architecture of economically significant milk traits and supports the development of genomic selection strategies to improve dairy production. Future studies with larger populations and diverse environmental conditions are recommended to refine these associations.
KEY WORDS: *Bos taurus*, lactation performance, mastitis indicator traits, milk quality, quantitative traits

Introduction

Milk production and its compositional traits play a key role in the profitability and sustainability of dairy farming, especially in Holstein cattle, which are the most widely used breed in global milk production (Teissier et al., 2018; Liu et al., 2020). While conventional breeding programmes have substantially increased milk yield (MY), the improvement of milk quality traits, such as somatic cell count (SCC), fat content, solids-not-fat (SNF), density, protein, freezing point, time, lactose, electrical conductivity (EC), and pH, MY, remains difficult due to their moderate to low heritability and complex genetic architecture (Kirkpatrick and Berry, 2024; Bai et al., 2025; Bongers et al., 2025).

Among these traits, somatic cell count (SCC) is widely recognised as a reliable indicator for udder health and milk hygiene (Ruegg, 2017). SCC reflects the concentration somatic cells and leukocytes in the mammary gland, including macrophages, neutrophils, and lymphocytes, released in response to intramammary infection. Elevated SCC is strongly correlated with subclinical forms of mastitis and is routinely used for early detection of inflammatory episodes in dairy herds (Auldist et al., 1996; Ogola et al., 2007). Importantly, an increase in SCC is often associated with alterations in the milk synthesis, including reduced lactose and casein

production, increased proteolytic and lipolytic enzyme activity, and disturbances in ion balance, all of which adversely affect milk composition and its technological quality (Forsbäck et al., 2009; Bochniarz et al., 2023).

SCC is not only a diagnostic indicator but also a heritable trait of relevance in dairy cattle selection. High SCC is associated with reduced milk yield, impaired coagulation properties, and increased risk of culling. Therefore, integrating SCC into genomic evaluations provides two benefits by improving udder health and milk quality (Norberg, 2005; Liu et al., 2022). SCC is generally considered a low-heritability trait. Recent estimates indicate heritability (h^2) values of 0.06–0.17 for SCC and 0.10–0.13 for somatic cell score (SCS) across different dairy populations. These values confirm that SCC is characterised by low rather than moderate heritability. Similar patterns have been reported in various Holstein populations, including studies conducted in Türkiye and other countries, where milk production traits generally show higher heritability (e.g., $h^2 \approx 0.20$ for milk yield), whereas functional traits such as SCC show considerably lower genetic determinism (Sahin et al., 2012; Visentin et al., 2017).

Beside SCC, other phenotypic traits serve as complementary indicators of milk quality and udder health. Milk density is an integrative parameter that reflects total solids, including protein, fat, and lactose, and decreases significantly during mastitis due to cellular damage and altered milk secretion (Suranindyah et al., 2015; Lima et al., 2018). Similarly, freezing point (FP) depends on solute concentration and is frequently used in milk quality control to detect abnormalities such as dilution or inflammation (Davoodi et al., 2016). Damage to mammary epithelial cells leads to an increased leakage of sodium (Na^+) and chloride (Cl^-) ions into the milk, resulting in elevated electrical conductivity (EC). As such, EC is used as a rapid indicator of milk quality and early detection of mastitis. Studies have shown strong correlations (r = 0.65–

0.80) between EC and SCC, suggesting the utility of EC as a supporting phenotype in genomic evaluations for mastitis resistance (Nielen et al., 1992; Norberg, 2005; Boas et al., 2017).

Milk pH, another often overlooked trait, increases during mastitis due to altered ion flux and the accumulation of basic metabolites. This milk parameter is closely associated with enzymatic activity, protein degradation (particularly casein), and microbial proliferation, all of which are influenced by SCC and udder health (Auldist et al., 1996; Hortet and Seegers, 1998).

Somatic cell count (SCC) has been extensively investigated in dairy cattle due to its strong association with mastitis, milk yield losses, and milk quality deterioration (Meredith et al., 2012; Buaban et al., 2022). Numerous studies have explored the genetic background of SCC and SCS, making SCC one of the most frequently analysed functional traits in dairy genomics (Hu et al., 2022). However, despite this extensive body of research, genome-wide association studies (GWAS) simultaneously addressing SCC together with multiple physicochemical milk quality traits, such as electrical conductivity, pH, density, and freezing point, are limited, particularly in Holstein populations raised under specific environmental and management conditions, such as those in Türkiye.

The advent of next-generation sequencing technologies has advanced commercial genotyping tools by enabling high-resolution mapping of complex traits. Single nucleotide polymorphisms (SNPs), due to their abundance and stability, have become central to GWAS and genomic evaluations (Pryce et al., 2014; Kizilaslan et al., 2024; Siberski-Cooper et al., 2024). Previous studies have identified key genes, including *DGAT1*, *SCD1*, *GH*, and *STAT5A* that influence milk composition traits (Grisart et al., 2002; Stoop et al., 2009). However, recent studies have provided additional evidence of genetic polymorphisms associated with milk production and resistance to mastitis. For instance, Bagnicka et al. (2023) demonstrated significant associations between candidate gene polymorphisms and both milk production and

mastitis-related phenotypes in dairy cattle. Similarly, recent GWAS have identified novel genomic regions associated with SCC and functional milk traits in different Holstein populations (Grisart et al., 2002; Stoop et al., 2009; Metin Kiyici et al., 2020; Kiyici et al., 2022; Erdoğan et al., 2024). These findings underline the importance of integrating current genomic evidence when investigating milk production and udder health traits.

Here, a GWAS was conducted using high-density SNP genotyping data to identify genomic regions and candidate genes associated with SCC, fat, SNF, density, protein, freezing point, time, lactose, electrical conductivity, pH, and milk yield in Black-and-White Holstein cows raised in Türkiye. The primary objective was to detect significant SNPs associated with both production and functional milk traits. These findings contribute to improved biological understanding and support the development of genomic selection strategies.

Material and methods

Study population and phenotypes

Study area: The study was conducted in the Develi district of Kayseri province, characterised by a continental climate with hot, dry summers and cold winters. During winter months, temperatures frequently fall to −10 °C or below, while summer temperatures may exceed 30–35 °C. The region has an average annual rainfall of approximately 389 mm, an average annual temperature of 11.7 °C, and an average altitude of 1 330 meters. These climatic conditions result in marked seasonal variations that may influence animal performance and milk production traits.

Animal selection and management:

The animal study protocol was approved by the Local Ethics Committee for Experimental Animals of Erciyes University, Türkiye, (file number 13/72). The study population consisted of 147 Black-and-White Holstein dairy cows reared under commercial

farm conditions in Türkiye. All animals were in their second lactation and raised under uniform management, including an intensive production system and a standardised feeding strategy. Cows were fed a total mixed ration (TMR) formulated to meet or exceed the nutrient requirements recommended by the National Research Council (NRC), providing approximately 2 750 kcal/kg of metabolisable energy and 16–25% crude protein. Water and mineral licking blocks were provided *ad libitum* throughout the study period. Animals were housed in semi-open, free-stall barns in group pens. Health status was monitored daily by the farm veterinarian, and individual health records were maintained for each cow. All cows were milked three times daily as part of the routine management protocol.

Phenotypes: Cows were monitored throughout their lactation period, and phenotypic performance records were collected, including ear tag numbers and calf birth dates. On average, 10 (range: 7–12) milk samples were collected per cow during lactation.

Each 50 ml milk sample was analysed for SCC, fat, SNF, density, protein, freezing point, time, lactose, EC, and pH. SCC was expressed as 1000×10^3 cells/ml, fat, SNF, protein, and lactose as percentages, density in g/ml, freezing point in °C, time in seconds, EC in mS/cm, and pH using a standard pH meter. MY was defined as the 305-day lactation yield expressed in kilograms.

A total of 1 431 observations were included in the dataset. MY was recorded once per cow (n = 147) as the 305-day lactation yield. This value was estimated using data from an automated milking system (DeLaval, Stockholm, Sweden), which continuously recorded daily milk yield and enabled calculation of standardised lactation production.

Genotyping and quality control: Blood samples were collected from the jugular veins of each cow into 10 ml EDTA-coated tubes by licensed veterinarians, following standard animal welfare and biosafety procedures. DNA extraction was performed using the QIAcube HT automated system and a commercial kit (Qiagen DNeasy Blood-Tissue Kit, Hilden, Germany).

The extracted DNA was subjected to quality control, and samples meeting the criteria (A260/280 > 1.8, A260/230 > 1.5, >70 ng/µl, and high DNA integrity) were genotyped using the GeneTitan Multi Channel (MC) Instrument and the 100K Axiom Bovine (Applied Biosytems, Waltham, MA, USA) SNP genotyping array, following the manufacturer's instructions.

Subsequently, data filtering was conducted using quality control criteria established by McCarthy et al. (2008), The Wellcome Trust Case Control Consortium, (2007), and Weale (2010) to reduce type I and type II error rates. SNPs with a minor allele frequency (MAF) below 0.05, call rates below 95%, or mapped to sex chromosomes were excluded. Additionally, samples with a call rate below 90% and those with identity by state (IBS) values exceeding 95% were removed. SNPs deviating from Hardy-Weinberg equilibrium (HWE) (P = 0.05/number of SNPs) were excluded using a Bonferroni correction.

Statistical analyses

Phenotypic outliers exceeding ±3 SD from the mean were removed from the further analysis prior to GWAS. The repeated generalised least squares (rGLS) function from the RepeatABEL R package was employed to perform GWAS on phenotypic data with repeated measurements (R Core Team, 2020). This function is specifically designed for datasets involving related individuals and repeated observations. It fits a linear mixed model that accounts for random additive polygenic effects and random effects due to the repeated measurements of the same animal (i.e., permanent environment effect). For each genetic marker, the rGLS function calculates *P*-values based on Wald's F test to assess the null hypothesis that the additive marker effect equals zero. By incorporating a genomic relationship matrix (GRM), the model accounts for genetic relatedness between individuals, thereby improving the accuracy of the GWAS. This approach ensures that both between-individual relatedness and within-individual repeated measures are appropriately accounted for.

The statistical model applied was as follows:

$$y = X\beta + Zg + Zpe + e,$$

where:y is the vector of the trait under analysis; β represents fixed effects, including SNP effects; *g* denotes random additive genetic effects; *pe* denotes random permanent environmental effects; and *e* represents residuals with variance σe^2, where e $\sim N(0, \sigma_e^2)$. X and Z are design matrices.

Random effects were assumed to follow a multivariate normal distribution: g $\sim N(0, G\sigma_g^2)$ and pe $\sim N(0, I\sigma_{pe}^2)$, where G is the genomic relationship matrix.

The estimated variance matrix for the model was:

$$V = ZGZ^T\sigma_g^2 + ZZ^T\sigma_{pe}^2 + I\sigma_e^2.$$

Wald tests were used to compute *P*-values for SNP effects (β). Computations were optimised using the eigen decomposition of V and the built-in QR decomposition function in R for fitting linear models. Variance components for the specified model were extracted from the GWAS object.

Q-Q plots of observed test statistics for each SNP were compared with expected values under the null hypothesis to identify potential systematic biases. Genomic control was applied to *P*-values to correct for inflation in test statistics.

Visualisation and significance thresholds: Manhattan plots were generated to visualise $-\log_{10}$(*P*-value) for all SNPs based on their chromosomal positions. Genome-wide and chromosome-wide significance thresholds were applied to minimise type I error rates due to multiple testing. Genome-wide significance was set at 1.26×10^{-6} (0.05/79 769), and chromosome-wide significance at 1.81×10^{-5} ((0.05/79 769) × 29), based on Bonferroni correction for multiple testing.

Functional gene annotation: Positional information for significant SNPs and nearby genes was retrieved using the ARS-UCD1.3 genome assembly via the NCBI Genome Data Viewer.

Genes overlapping with SNPs were considered candidates. When a significant SNP was not located within a gene, a ±100 kb window was scanned to identify nearby candidate genes. Functional annotation of the identified candidate genes was performed using the DAVID Bioinformatics Resources 2021 (Huang et al., 2009). Where annotations in the bovine genome were limited, orthology-based annotations from sheep, goats, mice, and humans were utilised. Biological processes associated with the genes were described using Gene Ontology (GO) terms, and further examined in the QuickGO browser (Binns et al., 2009).

Results

Outliers in the phenotypic data were identified and removed prior to analysis, resulting in a dataset suitable for genome-wide association studies. The analysed traits included SCC, fat content, SNF, density, protein content, freezing point, time, lactose content, electrical conductivity, pH, and milk yield. Descriptive statistics of the phenotypic data are summarised in Table 1.

The initial genotypic dataset comprised 85 455 SNPs from 147 animals. After applying quality control criteria, 79 769 SNPs were retained for subsequent GWAS.

The analysis identified 15 significant SNPs, including 3 exceeding the genome-wide significance threshold and 12 exceeding chromosome-wide thresholds. The distribution of significant associations across chromosomes is illustrated in the Manhattan plots (Figures 2 and 3). Quantile-Quantile plots indicated appropriate control of population stratification (Figure 1).

Among the identified loci, the most relevant associations were detected on BTA7 and BTA24. SNP rs135761188 located on BTA7 was significantly associated with SCC and mapped to the *CBY3* (chibby family member 3) gene. Another genome-wide significant SNP, rs109679567 on BTA7, was located within the *CNOT6* (CCR4-NOT transcription complex subunit 6) gene and was also associated with SCC. In addition, significant SNPs identified on

BTA24 were mapped to the *RAB27B* (member RAS oncogene family) gene and were associated with fat and density traits, suggesting potential pleiotropic genetic effects. These findings indicate candidate genes potentially involved in immune response and milk production mechanisms. Detailed positional and functional information on significant SNPs and candidate genes is presented in Table 2.

Functional annotation of significant loci showed that several SNPs were located within or near biologically relevant genes (Table 2). Candidate genes identified in the vicinity of significant SNPs included *CBY3*, *CNOT6*, *RAB27B*, *PHLDB1*, and *LRRC4C*. These genes are involved in immune response, cellular secretion, metabolic regulation, and signalling pathways related to milk production and udder health.

Discussion

The present study identified significant SNPs associated with various milk traits in Holstein cattle. The results revealed the chromosomal locations of SNPs linked to traits such as SCC, fat, SNF, density, protein, freezing point, lactose, conductivity, pH, and milk yield, along with closely positioned genes.

Interestingly, some SNPs were associated with multiple milk traits. For example, the *RAB27B* gene on BTA24 was associated with both fat and density traits. This gene is involved in cellular secretion and membrane trafficking, which may influence milk composition and quality. The association of *RAB27B* with milk fat percentage and density suggests a role in metabolic and cellular processes related to milk composition. These findings are consistent with previous studies describing the involvement of RAB27B in exosome biogenesis and secretion (Izumi, 2021). RAB27B regulates the docking and fusion of multivesicular endosomes (MVEs) with the plasma membrane, enabling the release of exosomes that carry lipids, proteins, and

signalling molecules involved in cellular communication and metabolic regulation (Blanc and Vidal, 2018).

The observed association between *RAB27B* and milk fat percentage may be explained by its involvement in lipid metabolism. RAB27B is involved in exosome secretion, and these vesicles participate in the transport and secretion of lipids, which are essential for milk fat synthesis. Previous research has demonstrated that RAB27B influences lipid droplet formation and trafficking, processes directly related to milk fat production (Li and Yu, 2016). Moreover, the pleiotropic effects of RAB27B, as suggested by its concurrent association with milk density, reflect its broader role in regulating the secretion and distribution of cellular components that affect physical milk traits. These findings are consistent with reports describing the role of RAB27B in membrane trafficking and its interaction with actin filaments in regulating exosome positioning and release (Ostrowski et al., 2010). Additionally, RAB27B-mediated exosome secretion in milk-producing cells may similarly influence the transport of essential lipids and proteins, affecting both milk fat content and its structural density. The clustering of significant SNPs within the *RAB27B* locus and their association with multiple milk traits further supports potential pleiotropic effects, consistent with findings in other systems (Ji et al., 2022).

The role of RAB27B in regulating both milk fat percentage and density may involve its interactions with specific effectors such as Slp4 and Slac2b, as reported in previous studies (Alzahofi et al., 2020). These effectors facilitate the positioning of MVEs, ensuring efficient exosome release. The association of *RAB27B* with milk fat percentage and density aligns with its role in lipid metabolism and exosome-mediated transport. These findings support the hypothesis that exosome biogenesis and secretion contribute to milk composition traits. Future research is needed to validate RAB27B function and clarify its molecular role in milk production.

SNPs rs135761188 and rs109679567 located on BTA7 were associated with SCC and mapped to the *CBY3* and *CNOT6* genes, respectively. These genes are considered to be involved in immune response and host defence processes. Additionally, SNP rs41615984 was associated with both SNF and protein traits, suggesting possible pleiotropic effects and genetic correlations between these traits.

SNPs associated with milk yield included rs43718731 and rs132639440, both located on BTA15. These SNPs were mapped to the *PHLDB1* and *LRRC4C* genes, respectively, which are related to cellular signalling and metabolic processes. The presence of multiple SNPs associated with milk yield on the same chromosome may indicate a potential genetic hotspot in this region.

Overall, this study contributes to the understanding of the genetic architecture of milk production traits in Holstein cattle. It underscores the importance of population-specific studies in assessing genetic and environmental interactions influencing these traits. The results pinpoint candidate genes and chromosomal regions that may be considered in genomic selection for dairy cattle. However, further studies are required to validate these findings in different populations and environments, which would improve their applicability in breeding programmes.

Conclusions

This study identified genomic regions associated with milk production and functional milk traits in Black-and-White Holstein cattle, contributing to a better understanding of the genetic basis underlying economically important dairy characteristics. The results support the hypothesis that both production and udder health-related traits are influenced by distinct but potentially interconnected genomic loci. The findings reflect the complexity of milk quality and mastitis-related traits and support the use of genomic information in dairy breeding

programmes. Although the identified loci represent potential targets for selection, further validation in larger and independent populations is necessary to confirm their stability and biological relevance. Future studies integrating functional genomics and genotype-environment interactions should be conducted to clarify the biological mechanisms underlying these traits and to improve the efficiency of genomic selection in dairy cattle.

Funding

This study was supported by the Scientific Research Projects Department of Erciyes University under Project ID: FBA-2022-11669.

Acknowledgments

The authors would like to thank Saray Farm Dairy Operations Co. (Kayseri, Türkiye) for their help during the experiments. Special thanks are extended to the farm staff for their invaluable assistance in managing the animals and collecting data. We also express our gratitude to the laboratory personnel of Genome and Stem Cell Center at Erciyes University (GEN-KOK) for providing laboratory facilities. Finally, we are grateful to the reviewers and collaborators for their valuable comments, which improved the quality of this work.

Conflict of interest

The Authors declare that there is no conflict of interest.

References

Alzahofi N., Welz T., Robinson C.L., Page E.L. et al., 2020. Rab27a co-ordinates actin-dependent transport by controlling organelle-associated motors and track assembly proteins. Nat. Commun. 11, 3495, https://doi.org/10.1038/s41467-020-17212-6

Auldist M.J., Coats S.J., Sutherland B.J., Hardham J.F., McDowell G.H., Rogers G.L., 1996. Effect of somatic cell count and stage of lactation on the quality and storage life of ultra high temperature milk. J. Dairy Res. 63, 377–386, https://doi.org/10.1017/s0022029900031903

Bai H., Lai Z., Zhang J., Zheng X., Zhang J., Jin W., Lin L., Mao S., 2025. Host genetic regulation of specific functional groups in the rumen microbiome of dairy cows: Implications for lactation trait. J. Adv. Res. 75, 111–122, https://doi.org/10.1016/j.jare.2024.11.012

Binns D., Dimmer E., Huntley R., Barrell D., O'Donovan C., Apweiler R., 2009. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 25, 3045–3046, https://doi.org/10.1093/bioinformatics/btp536

Blanc L., Vidal M., 2018. New insights into the function of Rab GTPases in the context of exosomal secretion. Small GTPases 9, 95–106, https://doi.org/10.1080/21541248.2016.1264352

Boas D.F.V., Vercesi Filho A.E., Pereira M.A., Junior L.C.R., El Faro L., 2017. Association between electrical conductivity and milk production traits in dairy Gyr cows. J. Appl. Anim. Res. 45, 227–233, https://doi.org/10.1080/09712119.2016.1150849

Bochniarz M., Błaszczyk P., Szczubiał M., Vasiu I., Adaszek Ł., Michalak K., Pietras-Ożga D., Wochnik M., Dąbrowski R., 2023. Comparative analysis of total protein, casein, lactose, and fat content in milk of cows suffering from subclinical and clinical mastitis caused by Streptococcus spp. J. Vet. Res. 67, 251, https://doi.org/10.2478/jvetres-2023-0028

Bongers R., Rochus C.M., Houlahan K. et al., 2025. Estimation of genetic parameters and genome-wide association study for enzootic bovine leukosis resistance in Canadian Holstein cattle. J. Dairy Sci. 108, 611–622, https://doi.org/10.3168/jds.2024-25196

Buaban S., Lengnudum K., Boonkum W., Phakdeedindan P., 2022. Genome-wide association study on milk production and somatic cell score for Thai dairy cattle using weighted single-step approach with random regression test-day model. J. Dairy Sci. 105, 468–494, https://doi.org/10.3168/jds.2020-19826

Davoodi S.H., Shahbazi R., Esmaeili S., Sohrabvandi S., Mortazavian A., Jazayeri S., Taslimi A., 2016. Health-related aspects of milk proteins. Iran. J. Pharm. Res. 15, 573–591

Erdoğan M., Çinkaya S., Brenig B., Çelikeloğlu K., Demirtaş M., Sarıibrahimoğlu S., Tekerli

M., 2024. Genome-wide association studies for milk production traits and persistency of first calving Holstein cattle in Türkiye. Front. Vet. Sci. 11, 1461075, https://doi.org/10.3389/fvets.2024.1461075

Forsbäck L., Lindmark-Månsson H., Andrén A., Åkerstedt M., Svennersten-Sjaunja K., 2009. Udder quarter milk composition at different levels of somatic cell count in cow composite milk. Animal 3, 710–717, https://doi.org/10.1017/S1751731109004042

Grisart B., Coppieters W., Farnir F. et al., 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. Genome Res. 12, 222–231, https://doi.org/10.1101/gr.224202

Hortet P., Seegers H., 1998. Loss in milk yield and related composition changes resulting from clinical mastitis in dairy cows. Prev. Vet. Med. 37, 1–20, https://doi.org/10.1016/s0167-5877(98)00104-4

Hu Z.L., Park C.A., Reecy J.M., 2022. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res. 50, D956–D961, https://doi.org/10.1093/nar/gkab1116

Huang D.W., Sherman B.T., Lempicki R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13, https://doi.org/10.1093/nar/gkn923

Izumi T., 2021. In vivo roles of Rab27 and its effectors in exocytosis. Cell Struct. Funct. 46, 79–94, https://doi.org/10.1247/csf.21043

Ji Z.-H., Ren W.-Z., Wu H.-Y., Zhang J.-B., Yuan B., 2022. Exosomes in mastitis—research status, opportunities, and challenges. Animals 12, 2881, https://doi.org/10.3390/ani12202881

Kirkpatrick B.W., Berry D.P., 2024. A genomic analysis of twinning rate and its relationship with other reproductive traits in Holstein-Friesian cattle. J. Dairy Sci. 108, 1686–1698, https://doi.org/10.3168/jds.2024-25236

Kiyici J.M., Akyüz B., Kaliber M., Arslan K., Aksel E.G., Cinar M.U., 2022. Association of *GH*, *STAT5A*, *MYF5* gene polymorphisms with milk somatic cell count, EC and pH levels of Holstein dairy cattle. Anim. Biotechnol. 33, 401–407, https://doi.org/10.1080/10495398.2020.1800483

Kizilaslan M., Arzik Y., Behrem S., Yavuz E., White S.N., Cinar M.U., 2024. Unravelling the genetic architecture of serum biochemical indicators in sheep. Genes 15, 990, https://doi.org/10.3390/genes15080990

Li C., Yu S.S.B., 2016. Rab proteins as regulators of lipid droplet formation and lipolysis. Cell

Biol. Int. 40, 1026–1032, https://doi.org/10.1002/cbin.10650

Lima R.S., Danielski G.C., Pires A.C.S., 2018. Mastitis detection and prediction of milk composition using gas sensor and electrical conductivity. Food Bioprocess Technol. 11, 551–560, https://doi.org/10.1007/s11947-017-2029-6

Liu D., Xu Z., Zhao W. et al., 2022. Genetic parameters and genome-wide association for milk production traits and somatic cell score in different lactation stages of Shanghai Holstein population. Front. Genet. 13, 940650, https://doi.org/10.3389/fgene.2022.940650

Liu L., Zhou J., Chen C.J., Zhang J., Wen W., Tian J., Zhang Z., Gu Y., 2020. GWAS-based identification of new loci for milk yield, fat, and protein in Holstein cattle. Animals 10, 2048, https://doi.org/10.3390/ani10112048

McCarthy M.I., Abecasis G.R., Cardon L.R., Goldstein D.B., Little J., Ioannidis J.P.A., Hirschhorn J.N., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. 9, 356–369, tttchttps://doi.org/10.1038/nrg2344

Meredith B.K., Kearney F.J., Finlay E.K., Bradley D.G., Fahey A.G., Berry D.P., Lynn D.J., 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet. 13, 21, https://doi.org/10.1186/1471-2156-13-21

Metin Kiyici J., Akyüz B., Kaliber M., Arslan K., Aksel E.G., Çinar M.U., 2020. *LEP* and *SCD* polymorphisms are associated with milk somatic cell count, electrical conductivity and pH values in Holstein cows. Anim. Biotechnol. 31, 498–503, https://doi.org/10.1080/10495398.2019.1628767

Nielen M., Deluyker H., Schukken Y.H., Brand A., 1992. Electrical conductivity of milk: measurement, modifiers, and meta analysis of mastitis detection performance. J. Dairy Sci. 75, 606–614, https://doi.org/10.3168/jds.S0022-0302(92)77798-4

Norberg E., 2005. Electrical conductivity of milk as a phenotypic and genetic indicator of bovine mastitis: A review. Livest. Prod. Sci. 96, 129–139, https://doi.org/10.1016/j.livprodsci.2004.12.014

Ogola H., Shitandi A., Nanua J., 2007. Effect of mastitis on raw milk compositional quality. J. Vet. Sci. 8, 237–242, https://doi.org/10.4142/jvs.2007.8.3.237

Ostrowski M., Carmo N.B., Krumeich S. et al., 2010. Rab27a and Rab27b control different steps of the exosome secretion pathway. Nat. Cell Biol. 12, 19–30, https://doi.org/10.1038/ncb2000

Pryce J.E., Johnston J., Hayes B.J. et al., 2014. 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. J. Dairy Sci. 97,

1799–1811, https://doi.org/10.3168/jds.2013-7368

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna (Austria), https://www.r-project.org/

Ruegg P.L., 2017. A 100-Year Review: Mastitis detection, management, and prevention. J. Dairy Sci. 100, 10381–10397, https://doi.org/10.3168/jds.2017-13023

Siberski-Cooper C.J., Mayes M.S., Gorden P.J., Kramer L., Bhatia V., Koltes J.E., 2024. The genetic architecture of complete blood counts in lactating Holstein dairy cows. Front. Genet. 15, 1360295, https://doi.org/10.3389/fgene.2024.1360295

Stoop W.M., Schennink A., Visker M.H.P.W., Mullaart E., van Arendonk J.A.M., Bovenhuis H., 2009. Genome-wide scan for bovine milk-fat composition. I. Quantitative trait loci for short- and medium-chain fatty acids. J. Dairy Sci. 92, 4664–4675, https://doi.org/10.3168/jds.2008-1966

Suranindyah Y., Wahyuni E., Bintara S., Purbaya G., 2015. The effect of improving sanitation prior to milking on milk quality of dairy cow in farmer group. Procedia Food Sci. 3, 150–155, https://doi.org/10.1016/j.profoo.2015.01.016

Teissier M., Sanchez M.P., Boussaha M., Barbat A., Hoze C., Robert-Granie C., Croiseau P., 2018. 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. J. Dairy Sci. 101, 3126–3139, https://doi.org/10.3168/jds.2017-13587

The Wellcome Trust Case Control Consortium, 2007. The Wellcome Trust Case Control Consortium.

Visentin G., McParland S., De Marchi M., McDermott A., Fenelon M.A., Penasa M., Berry D.P., 2017. Processing characteristics of dairy cow milk are moderately heritable. J. Dairy Sci. 100, 6343–6355, https://doi.org/10.3168/jds.2017-12642

Weale M.E., 2010. Quality control for genome-wide association studies. In: J.M. Walker. Methods in Molecular Biology. Humana Press. Totowa, NJ (USA), pp. 341–372, https://doi.org/10.1007/978-1-60327-367-1_19

Table 1. Descriptive statistics for milk composition and yield traits

Traits	Observations	Mean	Standard error	Min.	Max.	SD	CV, %
SCC, 1000 × 10^3	1431	369	17	2	507	677	32
Fat, %	1431	3.17	0.030	0.32	9.54	1.22	38
Solid-not-fat, %	1431	9.08	0.010	6.07	11.60	0.45	04
Density, g/ml	1431	1.031	0.000	1.017	1.040	2.49	08
Protein, %	1431	3.42	0.004	2.34	4.30	0.16	04
Freezing point, °C	1431	−0.597	0.000	−0.417	−0.748	0.031	05
Time, s	1431	21.07	0.070	13.90	57.50	2.66	12
Lactose, %	1431	4.98	0.006	3.10	6.12	0.25	05
Conductivity, mS/cm	1431	5.01	0.012	3.62	9.68	0.48	09
pH	1431	6.89	0.003	4.05	8.88	0.13	02
Milk yield, kg	147	7.99	61.80	2.23	13.41	2.33	29

SCC – somatic cell count, Min. – minimum, Max. – maximum, SD – standard deviation, CV – coefficient of variation

Table 2. Significant SNPs associated with milk composition and quality traits

Trait	rs IDs	Chr	Position, bp	*P*-value	SNP effect size	Significance level	Associated genes	
							Associated genes	Distance
SCC	rs135761188	7	1662092	5.72×10^{-07}	304.26	GW	*CBY3*	Within
SCC	rs109679567	7	633624	7.56×10^{-06}	230.28	CW	*CNOT6*	Within
Fat	rs29011587	1	17672335	5.79×10^{-06}	0.44	CW	–	–
Fat	rs135638793	24	50744312	6.51×10^{-06}	0.29	CW	*LOC132343789*	Within
Fat	rs110213695	24	54100787	7.01×10^{-06}	−0.28	CW	*RAB27B*	Within
SNF	rs41615984	10	42037030	3.20×10^{-06}	−0.13	CW	–	–
Density	rs109536476	24	54118399	7.26×10^{-07}	0.65	GW	*RAB27B*	Within
Density	rs110213695	24	54100787	1.51×10^{-05}	0.58	CW	*RAB27B*	Within
Protein	rs41615984	10	42037030	1.87×10^{-05}	−0.04	CW	–	–
Freezing point	rs42154364	28	42643759	1.12×10^{-05}	−0.78	CW	*FRMPD2*	Within
Freezing point	rs135018781	28	43501640	1.30×10^{-05}	−0.86	CW	*C28H10orf71, TMEM273*	Within, ~96 kb
Lactose	rs41615984	10	42037030	1.03×10^{-05}	−0.07	CW	–	–
Lactose	rs135018781	28	43501640	1.21×10^{-05}	−0.07	CW	*C28H10orf71, TMEM273*	Within, ~96 kb
Conductivity	rs43422798	5	4262056	1.45×10^{-05}	0.23	CW	–	–
Conductivity	rs110028024	20	41853874	1.92×10^{-05}	0.15	CW	*PDZD2*	Within
pH	rs110813247	27	19417940	4.62×10^{-06}	0.03	CW	*FGL1*	~65 kb
Milk yield	rs43718731	15	29295061	4.20×10^{-07}	141.41	GW	*PHLDB1*	Within
Milk yield	rs29016808	16	12818152	3.70×10^{-06}	−136.57	CW	*LOC107133222, RGS21*	Within ~71 kb
Milk yield	rs132639440	15	71590615	1.14×10^{-05}	136.60	CW	*LRRC4C*	Within

GW – genome-wide significant, CW – chromosome-wide significant, SCC – somatic cell count, SNF – solid-not-fat, *CBY3* – chibby family member 3, *CNOT6* – CCR4-NOT transcription complex subunit 6, *RAB27B* – member RAS oncogene family, *FRMPD2* – FERM and PDZ domain containing 2, *C28H10orf71* – chromosome 10 open reading frame 71, *TMEM273* – transmembrane protein 273, *PDZD2* – PDZ domain containing 2, *FGL1* – fibrinogen like 1, *PHLDB1* – pleckstrin homology like domain family B member 1, *RGS21* – regulator of G protein signaling 21, *LRRC4C* – leucine rich repeat containing 4C

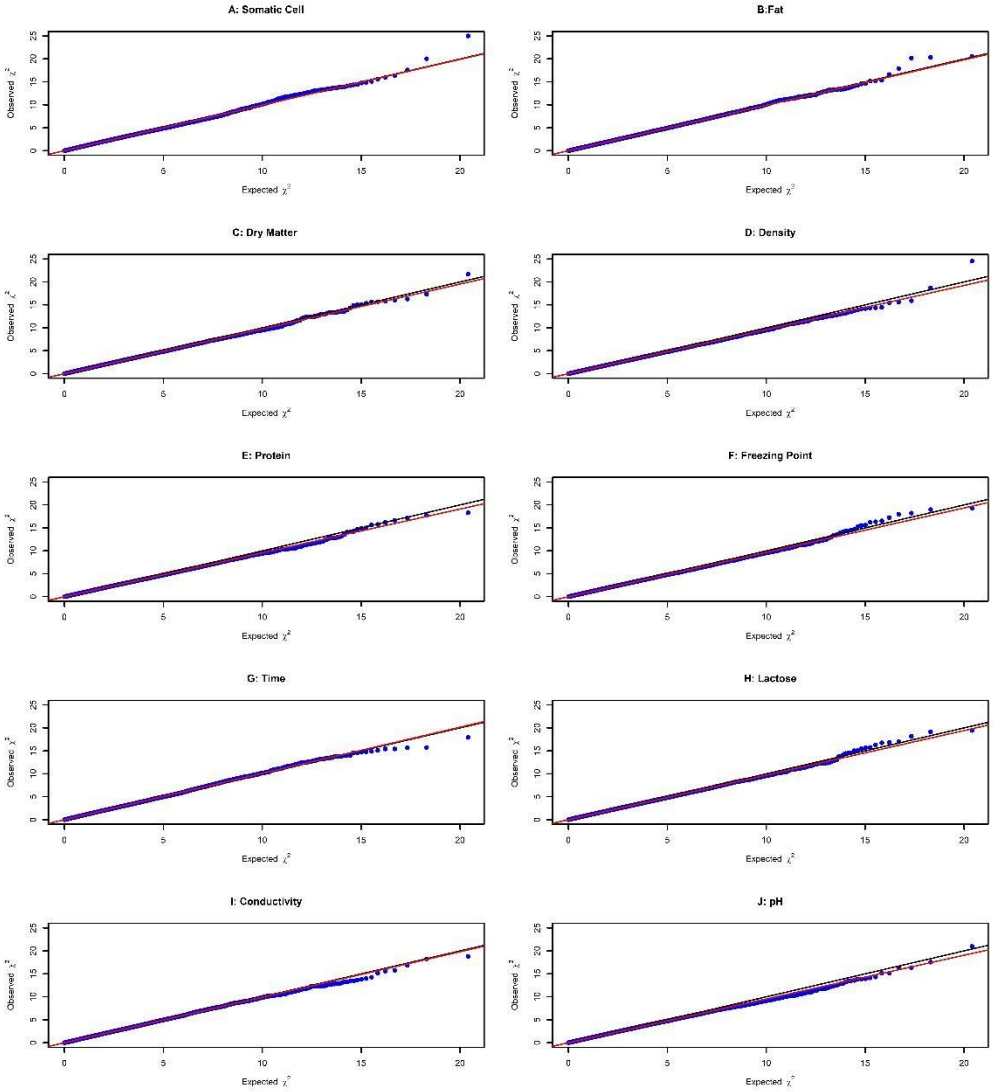

Figure 1. Quantile–Quantile (Q-Q) plots for milk composition traits. Q-Q plot observed and expected *P*-value (expressed as $-\log_{10}(P)$) of the GWAS for investigated traits

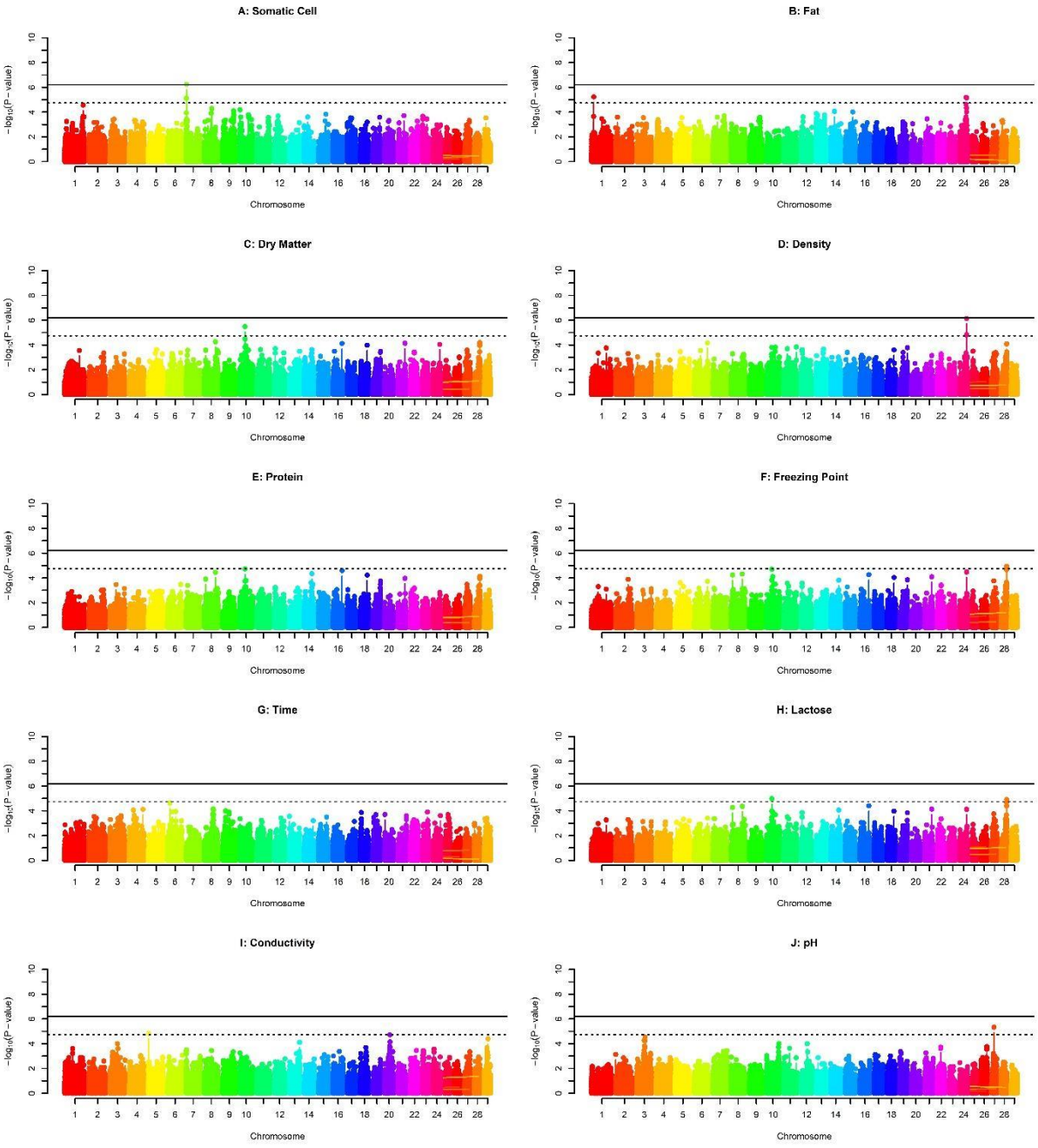

Figure 2. Manhattan plots for investigated traits

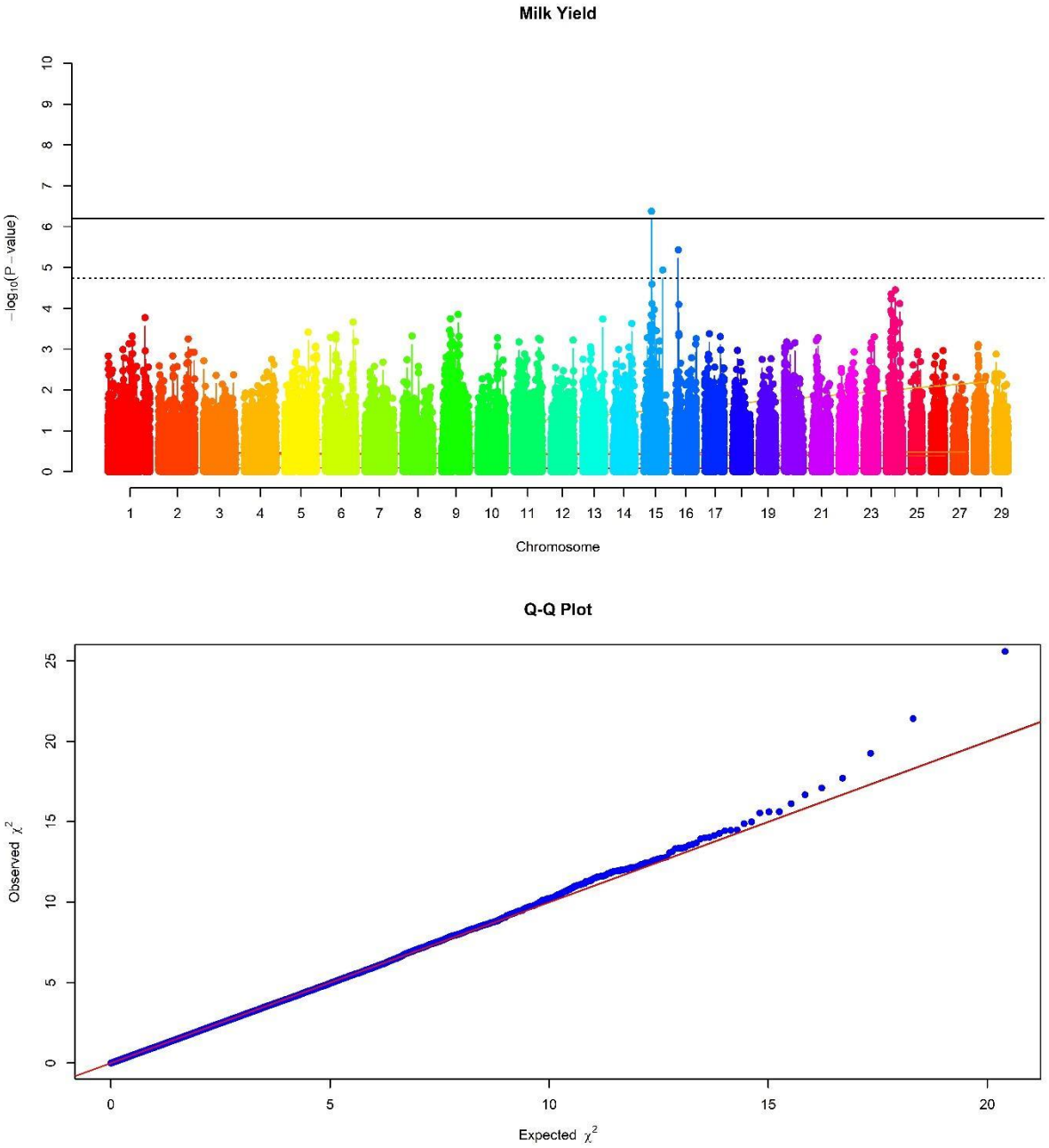

Figure 3. Manhattan and Quantile-Quantile (Q-Q) plots for milk yield. Q-Q plot observed and expected