



Genetic analysis of milk citrate predicted by milk mid-infrared spectra of Holstein cows in early lactation

Yansen Chen,^{1*} Hongqing Hu,¹ Hadi Atashi,^{1,2} Clément Grelet,³ Katrien Wijnrocx,¹ Pauline Lemal,¹ and Nicolas Gengler¹

¹TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech (ULiège-GxABT), 5030 Gembloux, Belgium

²Department of Animal Science, Shiraz University, 71441-13131 Shiraz, Iran

³Walloon Agricultural Research Center (CRA-W), 5030 Gembloux, Belgium

ABSTRACT

Milk citrate is regarded as an early biomarker of negative energy balance in dairy cows during early lactation and serves as a suitable candidate phenotype for genomic selection due to its wide availability across a large number of cows through milk mid-infrared spectra prediction. However, its genetic background is not well known. Therefore, the objectives of this study were to (1) analyze the genetic parameters of milk citrate; (2) identify genomic regions associated with milk citrate; and (3) analyze the functional annotation of candidate genes and quantitative trait loci (QTL) related to milk citrate in Walloon Holstein cows. In total, 134,517 test-day milk-citrate phenotypes (mmol/L) collected within the first 50 d in milk on 52,198 Holstein cows were used. These milk-citrate phenotypes, predicted by milk mid-infrared spectra, were divided into 3 traits according to the first (citrate1), second (citrate2), and third to fifth parity (citrate3+). Genomic information for 566,170 SNPs was available for 4,479 animals. A multiple-trait repeatability model was used to estimate genetic parameters. A single-step GWAS was used to identify candidate genes for citrate and post-GWAS analysis was done to investigate the relationship and function of the identified candidate genes. The heritabilities estimated for citrate1, citrate2, and citrate3+ were 0.40, 0.37, and 0.35, respectively. The genetic correlations among the 3 traits ranged from 0.98 to 0.99. The genomic correlations among the 3 traits were also close to 1.00 across the genomic regions (1 Mb) in the whole genome, which means that citrate can be considered as a single trait in the first 5 parities. In total, 603 significant SNPs located on 3 genomic regions (chromosome 7, 68.569–68.575 Mb; chromosome 14, 0.15–1.90

Mb; and chromosome 20, 54.00–64.28 Mb), were identified to be associated with milk citrate. We identified 89 candidate genes including *GPT*, *ANKH*, *PPP1R16A*, and 32 QTL reported in the literature related to the identified significant SNPs. These identified QTL were mainly reported associated with milk fatty acids and metabolic diseases in dairy cows. This study suggests that milk citrate in Holstein cows is highly heritable and has the potential to be used as an early proxy for the negative energy balance of Holstein cows in a breeding objective.

Key words: milk citrate, negative energy balance, candidate genes, QTL

INTRODUCTION

High-yield dairy cows in early lactation often experience negative energy balance (NEB) because of insufficient energy intake to support high milk production demands (Churakov et al., 2021). The NEB can lead to multiple metabolic diseases (e.g., ketosis) and reproductive problems (e.g., decrease in fertility rates; Walsh et al., 2011; Zachut et al., 2020), reducing animal welfare and farmers' economic profitability. During NEB, dairy cows undergo a mobilization of body-fat reserves, resulting in the liberation of glycerol and fatty acids (FA). The excessive presence of nonesterified FA (NEFA) in dairy cows' blood can precipitate the development of hepatic lipidosis, commonly known as fatty-liver disease (Herd, 2000). In such circumstances, the bovine liver significantly amplifies the synthesis of acetyl-CoA from FA, subsequently leading to the conversion into ketone bodies, especially β -hydroxybutyric acid (BHBA), due to diminished glucose levels and an oversupply of FA. The accumulation of these ketone bodies ultimately causes ketosis in cattle (Herd, 2000; Ospina et al., 2010).

Direct NEB of dairy cows can be calculated by subtracting energy expended (here included milk, excreta, and maintenance) from energy intake. However,

Received June 26, 2023.

Accepted November 8, 2023.

*Corresponding author: yanzen.chen@uliege.be

monitoring the NEB of individual cows under current commercial conditions is challenging for dairy farmers. Monitoring NEB requires frequent measurement of DMI, milk components, excreta, and body weight to calculate energy requirements and energy intake from nutrients. This approach is therefore not suitable for large-scale commercial assessment of NEB for dairy cows (Coffey et al., 2001; Friggens et al., 2007). The blood NEFA and BHBA were proven as indicators for detecting NEB in dairy cows (Ospina et al., 2010; Zachut et al., 2020; Pires et al., 2022). However, both NEFA and BHBA respond more slowly to NEB than milk citrate (Bjerre-Harpøth et al., 2012). When cows enter the NEB (reduced glucose), FA synthesis stops because it is a highly energy-consuming process. Citrate plays a role in this process that inhibits de novo FA synthesis in dairy cows (Garnsworthy et al., 2006). Therefore, citrate can be considered as a potential early biomarker for identifying NEB in dairy cows (Bjerre-Harpøth et al., 2012; Xu et al., 2020).

The first requirement for conducting a successful genetic selection for a given trait is to establish a method to measure the trait on a large number of animals at a low cost. Furthermore, it is necessary to estimate the genetic variation of the trait in the considered population to prove whether the trait is heritable. Despite the complexity of conventional methods for measuring milk citrate, novel mid-infrared (MIR) predictions can provide milk citrate concentration at the individual level on a large scale (Grelet et al., 2016).

Although some studies have been conducted on the genetic background of milk citrate in Montbéliarde cows (Sanchez et al., 2018, 2019, 2021), the genetic background of citrate in Holstein cows has not yet been studied. However, citric acid, a citrate conjugate, has been shown to be heritable ($h^2 \pm \text{SE}$: 0.54 ± 0.19) in small populations ($n = 371$) in Holstein cows and has been shown to be associated with metabolic energy (Buitenhuis et al., 2013).

Therefore, the objectives of this study were to (1) analyze the genetic parameters of milk citrate; (2) identify genomic regions associated with milk citrate through a single-step GWAS (ssGWAS); and (3) analyze the functional annotation of candidate genes and QTL related to milk citrate in Walloon Holstein cows. The potential relationships between milk citrate and traits of interest were also investigated.

MATERIALS AND METHODS

No human or animal subjects were used, so this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board.

Data

Phenotypic Data. All milk samples were collected by Elevéo (Awé groupe, Ciney, Belgium) from 2012 to 2019 during the official milk recording in the Walloon Region of Belgium. The milk samples were analyzed by MIR spectrometry (commercial instruments from FOSS) to generate MIR spectra. The milk spectra were harmonized into the common European format as described by Grelet et al. (2015) and then were used to predict milk citrate phenotypes (mmol/L, hereafter called citrate) for the MIR spectra. The prediction equation developed by Grelet et al. (2016) was applied to the milk MIR spectra of the present study. The coefficient of determination and root mean square error of validation for the citrate equation were 0.86 and 0.07 mmol/L, respectively. A total of 506 milk citrate reference data from 3 countries (Germany, France, and Luxembourg) and 3 cattle breeds (Holstein, Abondance, and Montbéliarde) was used to develop (380 samples) and validate (126 samples) the predicted model of milk citrate (Grelet et al., 2016). The citrate phenotypes were divided into 3 traits according to parity: **citrate1** for the first parity, **citrate2** for the second parity, and **citrate3+** for the third to fifth parity.

To remove outliers, the filtering methods proposed by Chen et al. (2021) were used. Briefly, (1) the predicted MIR spectra, for which the standardized Mahalanobis distance between the MIR data and the calibration dataset is ≤ 3 , were retained; (2) the predicted value of citrate was restricted within the range of ± 3 standard deviations of the mean. Then, the citrate was restricted to the first 50 DIM, a period during which most high-yield Holstein cows are in NEB.

Genotypic Data. Genotypic data related to cows with citrate phenotypes were extracted for 4,479 animals from the routine genetic evaluation system of Holstein cattle in the Walloon Region of Belgium. The used chip versions were BovineSNP50 K v1 to v3 (Illumina, San Diego, CA). The SNPs common among all 3 chips were kept. Nonmapped SNPs, SNPs located on sex chromosomes (**Chr**), and nonbiallelic SNPs were excluded. A minimum GenCall Score of 0.15 and a minimum GenTrain Score of 0.55 were used to keep SNPs (Wilmot et al., 2022). Next, genotypes were imputed to high density with a reference population of 4,352 high-density individuals (1,046 bulls and 3,288 cows) using the FImpute V2.2 software (Sargolzaei et al., 2014). The SNPs with Mendelian conflicts and those with minor allele frequency less than 5% were excluded. The difference between observed heterozygosity and that expected under Hardy-Weinberg equilibrium was estimated, and if the difference was greater than 0.15, the SNP was excluded (Wiggans et al., 2009).

Table 1. Range, mean, and SD of milk citrate (mmol/L) of Holstein cows in the first 50 DIM

Trait ¹	Minimum	Maximum	Mean	SD	Number of citrates	Number of animals
Citrate1	4.07	14.14	8.93	1.48	41,035	33,376
Citrate2	4.05	14.18	8.93	1.63	36,584	29,835
Citrate3+	3.97	14.18	9.18	1.76	56,898	28,775

¹Citrate1 = milk citrate in the first parity; citrate2 = milk citrate in the second parity; citrate3+ = milk citrate from third to fifth parity.

Finally, 566,170 out of 730,539 SNPs, distributed on 29 Chr, were kept.

After filtering, 134,517 citrate phenotypes on 52,198 Holstein cows distributed across 774 farms remained. The number of citrate phenotypes (animals) in each parity is shown in Table 1. The pedigree related to the dataset comprised 122,218 animals, of which 4,479 (3,215 cows and 1,264 bulls) had SNP information.

(Co)variance Component Estimation. A 3-trait repeatability model was used to estimate the (co)variance components. The model was fitted as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{h} + \mathbf{X}\mathbf{b} + \mathbf{Q}\mathbf{q} + \mathbf{W}_1\mathbf{c} + \mathbf{W}_2\mathbf{p} + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

where \mathbf{y} was the vector of citrate1, citrate2, and citrate3+. For each trait, \mathbf{h} was the vector of fixed herd-year-season of calving classes; \mathbf{b} was the vector of fixed regression coefficients for DIM, after standardization, and its quadratic; \mathbf{q} was the vector of fixed regression coefficients of the age of calving, after standardization, defined as a constant (parity effect), linear, and quadratic regression, defined within parities (first to fifth parity); \mathbf{c} was a vector of the nongenetic cow effect (within-parity permanent environment) random effects; \mathbf{p} was a vector of nongenetic cow \times parity effect (across-parity permanent environment) random effects, modeled only for citrate3+, as they allowed us to distinguish citrates for the same cow occurring during different parities (third to fifth parity); \mathbf{a} was a vector of random additive genetic effects; and \mathbf{e} was a vector of random residual effects. Additionally, \mathbf{H} , \mathbf{X} , \mathbf{Q} , \mathbf{W}_1 , \mathbf{W}_2 , and \mathbf{Z} were incidence matrixes assigning observations to effects. More detailed information about the model can be found in Chen et al. (2021). To calculate the relationship among animals, the \mathbf{H} relationship matrix was used. The \mathbf{H} matrix combines pedigree (\mathbf{A})-and genomic (\mathbf{G})-based relationship matrixes. The \mathbf{A} is the numerator relationship matrix for all animals included in the pedigree; \mathbf{G} is the genomic relationship matrix of genotyped animals obtained using the first formula described by VanRaden (2008):

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum_{i=1}^N p_i (1 - p_i)},$$

where \mathbf{Z} is a matrix of gene content adjusted for allele frequencies (0, 1, or 2 for aa, Aa, and AA, respectively); N is the number of SNPs; and p_i is the minor allele frequency of the i th SNP.

(Co)variance components were estimated by using the BLUPF90+ (version 2.42) program through the AI-REML method (Misztal et al., 2014). The h^2 , repeatability, and genetic and phenotypic correlations were calculated based on the estimated (co)variance components as previously described by Chen et al. (2021). The approximate SE of all calculated parameters were obtained according to the algorithm of Meyer and Houle (2013).

Genomic Correlations Among 3 Classes of Citrate Across the Whole Genome

The deregressed proofs (DRP) of animals were used in this section because not all animals had citrate phenotypes. Genomic breeding values (GEBV) of citrate and its reliability for all animals were calculated based on the estimated (co)variance components through BLUPF90+ (version 2.42; <http://nce.ads.uga.edu/html/projects/programs/Linux/64bit/>). Only those genotyped animals ($n = 4,435$) that had a reliability of over 0.30 for the 3 included traits were used for the next step (Chen et al., 2021). The DRP of selected animals were calculated based on the GEBV through DEPROOFSF90 (version 1.4) from the method developed by Garrick et al. (2009). Each selected animal got 3 DRP (DRP1, DRP2, DRP3+) for the 3 traits respectively (citrate1, citrate2, and citrate3+). A 3-trait marker-based BayesCII-model was used to estimate SNP effects for the predicted DRP (Cheng et al., 2018a). The model was as follows:

$$\mathbf{y}_i = \boldsymbol{\mu} + \sum_{j=1}^p m_{ij} \mathbf{D}_j \boldsymbol{\beta}_j + \mathbf{e}_i,$$

where \mathbf{y}_i is a vector of 3 DRP for animal i ; $\boldsymbol{\mu}$ is a vector of overall means for 3 DRP; p is the number of SNPs; m_{ij} is the genotype at SNP j (coded as 0, 1, 2) for animal i , with allele substitution effect $\mathbf{D}_j \boldsymbol{\beta}_j$, where \mathbf{D}_j is a diagonal matrix whose k th diagonal entry is an indicator

variable indicating whether the marker effect of locus j for trait DRP is zero or nonzero ($2^3 = 8$ combinations), and β_j follows a multivariate normal distribution; e_i is a vector of random residuals of 3 DRP for animal i , which is a priori assumed to be an independently and identically multivariate normal distribution.

The genomic correlation analyses were performed with the JWAS software (version 1.1.2; Cheng et al., 2018b) using a Monte-Carlo Markov Chain length of 50,000, with the first 5,000 iterations discarded as burn-in. Genomic correlations among the studied traits across the whole genome (1 Mb as one moving genomic region) were estimated as the posterior mean of the correlation across animals between the sampled breeding values (1 Mb as one moving genomic region). Its standard deviation (SD) was calculated across the kept samples as the approximated SE (Cheng et al., 2022). The breeding values of each 1 Mb genomic region were calculated as the sum of its SNP genotypes multiplied by the sampled marker effects for each kept iteration. In addition, genomic covariances of each 1 Mb genomic region were calculated as the sum of its SNP covariances for each kept iteration, and its SE was estimated by the SD across saved iterations.

Single-Step GWAS

The GEBV of genotyped animals were used in this section. The SNP effects for each trait (citrate1, 2, and 3+) were estimated by back-solving the animals' GEBV (Wang et al., 2012). The P -values of SNP were calculated as follows (Aguilar et al., 2019):

$$P\text{-value}_i = 2 \left(1 - f \left(\left| \frac{\hat{a}_i}{sd(\hat{a}_i)} \right| \right) \right),$$

where $P\text{-value}_i$ and \hat{a}_i are the P -value and effect of SNP i , respectively, and f is the cumulative standard normal function. The 8.83E-8 (0.05/566,170) was used as the significance threshold, which was calculated based on the Bonferroni correction for multiple testing. Linkage disequilibrium (LD; squared correlation coefficient, r^2) was calculated for significant SNPs. The SNP effects and P -values were calculated using POSTGSF90 software (version 1.73; Misztal et al., 2014).

Functional Annotation Analysis

The protein-encoding genes within 50 kb of the significant SNPs were considered candidate genes for citrate. Based on the results of Cánovas et al. (2022), the position (coordinate) of significant SNPs on reference genome assembly UMD3.1 (the used chip version) was

converted to the new position (coordinate) on the new reference genome assembly ARS-UCD1.2 through the Lift Genome Annotations tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). The gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were carried out on the identified candidate gene sets through the g:Profiler website (Raudvere et al., 2019). Next, a protein-protein interaction (PPI) analysis was performed through STRING (Szklarczyk et al., 2021) to reveal the relationship among the identified candidate genes. The PPI relationship was based on text mining, experiments, database, co-expression, neighborhood, gene fusion, and co-occurrence, and the minimum required interaction score was set to 0.40 (Zhou et al., 2019).

To explore potential relationships between citrate and other traits, the same genomic regions identified for citrate were annotated with Cattle QTLdb (<https://www.animalgenome.org/cgi-bin/QTLdb/BT/index>, accessed on Oct. 25, 2022; Hu et al., 2019). At present, Cattle QTLdb has 170,536 QTL, which were divided into 6 classes including exterior, production, health, reproduction, milk, meat, and carcass (https://www.animalgenome.org/cgi-bin/QTLdb/BT/ontrait?class_ID=1). To avoid the deviation caused by the annotation richness of the different traits, the hypergeometric-test approach was adopted for the enrichment analysis (Fonseca et al., 2020). The candidate genes and QTL annotations were performed using the GALLO package in R (Fonseca et al., 2020). In all enrichment analyses (GO, KEGG, QTL), the Benjamini-Hochberg method was used for multiple testing corrections. It should be noted that the Cattle QTL dataset currently does not include NEB.

Furthermore, the genome region with the most significant effect on citrate was subjected to the GO, KEGG, and QTL analysis. The data preparation and processing were performed using R (version 4.1.2; <https://www.r-project.org/>).

RESULTS

Descriptive Statistics and Genetic Parameters

Descriptive statistics of the studied traits are presented in Table 1. Citrate ranged from 3.97 to 14.18 mmol/L of milk. The means (SD) of citrate1, citrate2, and citrate3+ were 8.93 (1.48), 8.93 (1.63), and 9.18 (1.76) mmol/L, respectively. The average daily citrate showed a decreasing trend with increasing DIM in early lactation (Figure 1) and remained relatively stable during the period from 20 to 32 DIM.

The estimated (co)variance components, h^2 , and repeatability of 3 classes of citrate are presented in

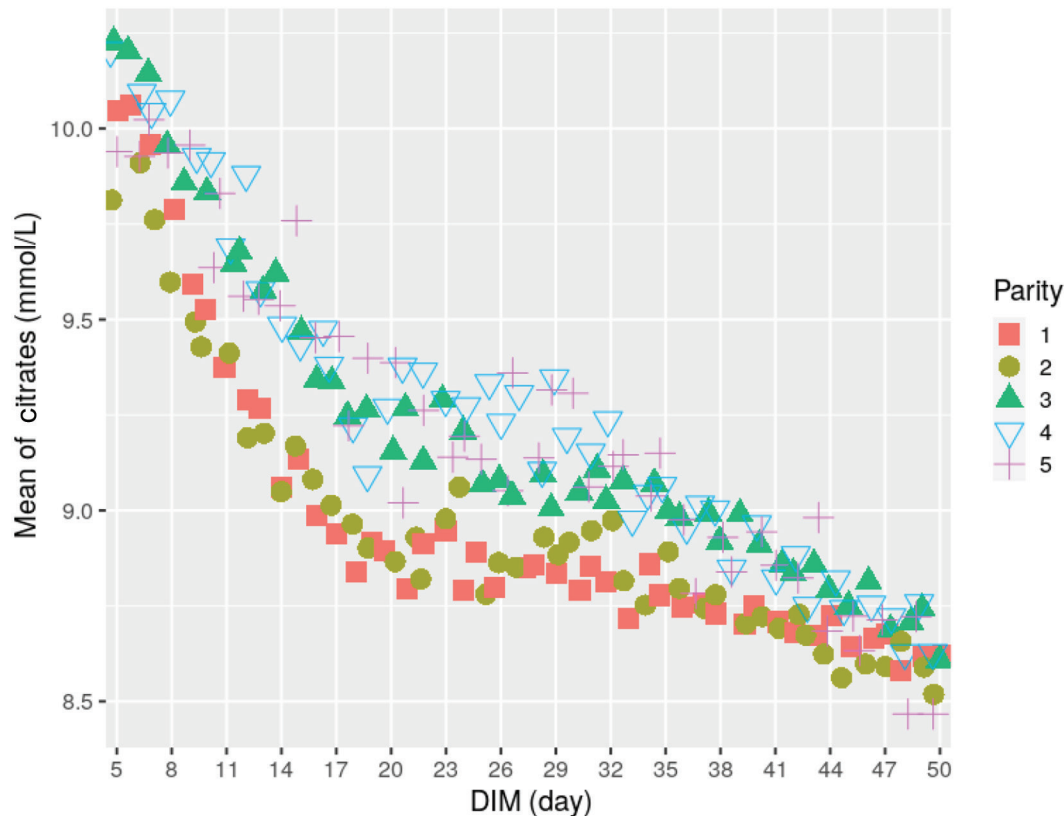


Figure 1. Average milk citrate phenotypes of Holstein cows, according to DIM.

Table 2. The h^2 and repeatability of citrate ranged from 0.35 to 0.40 and from 0.43 to 0.44, respectively. The h^2 of citrate decreased with advancing parity; however, its repeatability was stable.

Correlations Among the Different Parities

The genetic and phenotypic correlations of 3 classes of citrate are reported in Table 3. The genetic and phenotypic correlations were around 0.98 and 0.41, respectively.

To demonstrate the consistency of the estimated genetic correlations, genomic covariances and correlations of citrate1, citrate2, and citrate3+ were calculated across genomic regions (1 Mb) for the whole genome (Figure 2). The highest genomic covariances and correlations among the 3 classes of citrate were found in the same genomic region (Chr20 58.00–59.00 Mb, UMD3.1; Chr20 57.94–58.93 Mb, ARS-UCD1.2). The highest genomic covariance of the 3 classes of citrate was much higher than that of all other genomic regions (around 19 times). However, the highest genomic correlations

Table 2. Heritability, repeatability and (co)variance components¹ of milk citrate (mmol/L) of Holstein cows in the first 50 DIM

Trait ²	Heritability	Repeatability	σ_a^2	σ_c^2	σ_p^2	σ_e^2
Citrate1	0.40 ± 0.01	0.43 ± 0.01	0.71 ± 0.02	0.06 ± 0.01	NA ³	1.03 ± 0.01
Citrate2	0.37 ± 0.01	0.43 ± 0.01	0.82 ± 0.03	0.13 ± 0.02	NA	1.25 ± 0.02
Citrate3+	0.35 ± 0.01	0.44 ± 0.01	0.92 ± 0.04	0.17 ± 0.03	0.05 ± 0.02	1.45 ± 0.02

¹ σ_a^2 = additive genetic variance; σ_c^2 = within-parity permanent environment (nongenetic cow) variance; σ_p^2 = across-parity permanent environment (nongenetic cow × parity) variance, only for Citrate3+ traits; σ_e^2 = residual variance.

²Citrate1 = milk citrate in the first parity; citrate2 = milk citrate in the second parity; citrate3+ = milk citrate from third to fifth parity.

³NA = not applicable.

Table 3. Genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) among 3 classes of milk citrate of Holstein cows in the first 50 DIM

Trait ¹	Citrate1	Citrate2	Citrate3+
Citrate1		0.979 ± 0.005	0.967 ± 0.007
Citrate2	0.418 ± 0.006		0.991 ± 0.003
Citrate3+	0.406 ± 0.006	0.419 ± 0.006	

¹Citrate1 = milk citrate in the first parity; citrate2 = milk citrate in the second parity; citrate3+ = milk citrate from third to fifth parity.

among 3 classes of citrate were similar to that of all other genomic regions (around 0.99).

Single-Step GWAS

The Manhattan plots for the results of ssGWAS are presented in Figure 3A, and they were similar among the 3 classes of citrate. The number of significant SNPs associated with citrate1, citrate2, and citrate3+ were 589, 578, and 600, respectively, and 95% of these SNPs

were common among the 3 classes of citrate (Figure 3B). These significant SNPs are distributed on Chr7, 14, and 20, respectively, and their positions are close when they are on the same Chr (Table 4). The $-\log_{10}$ *P*-values and position (in the UMD3.1 and ARS-UCD1.2) for each significant SNP associated with 3 classes of citrate are shown in Supplemental File S1 (<https://github.com/Yansen0515/citrate>; Chen, 2023).

The LDs were calculated for all SNPs located in the identified genomic regions (Figure 4). Significant SNPs located in the genomic region identified in Chr7 were highly linked, whereas those located in the regions identified in Chr14 and 20 were not.

Functional Annotation Analysis

A respective total of 89, 88, and 88 candidate genes were identified for citrate1, citrate2, and citrate3+, respectively, and 99% of the identified genes were common among the 3 classes of citrate (Figure 3C). The

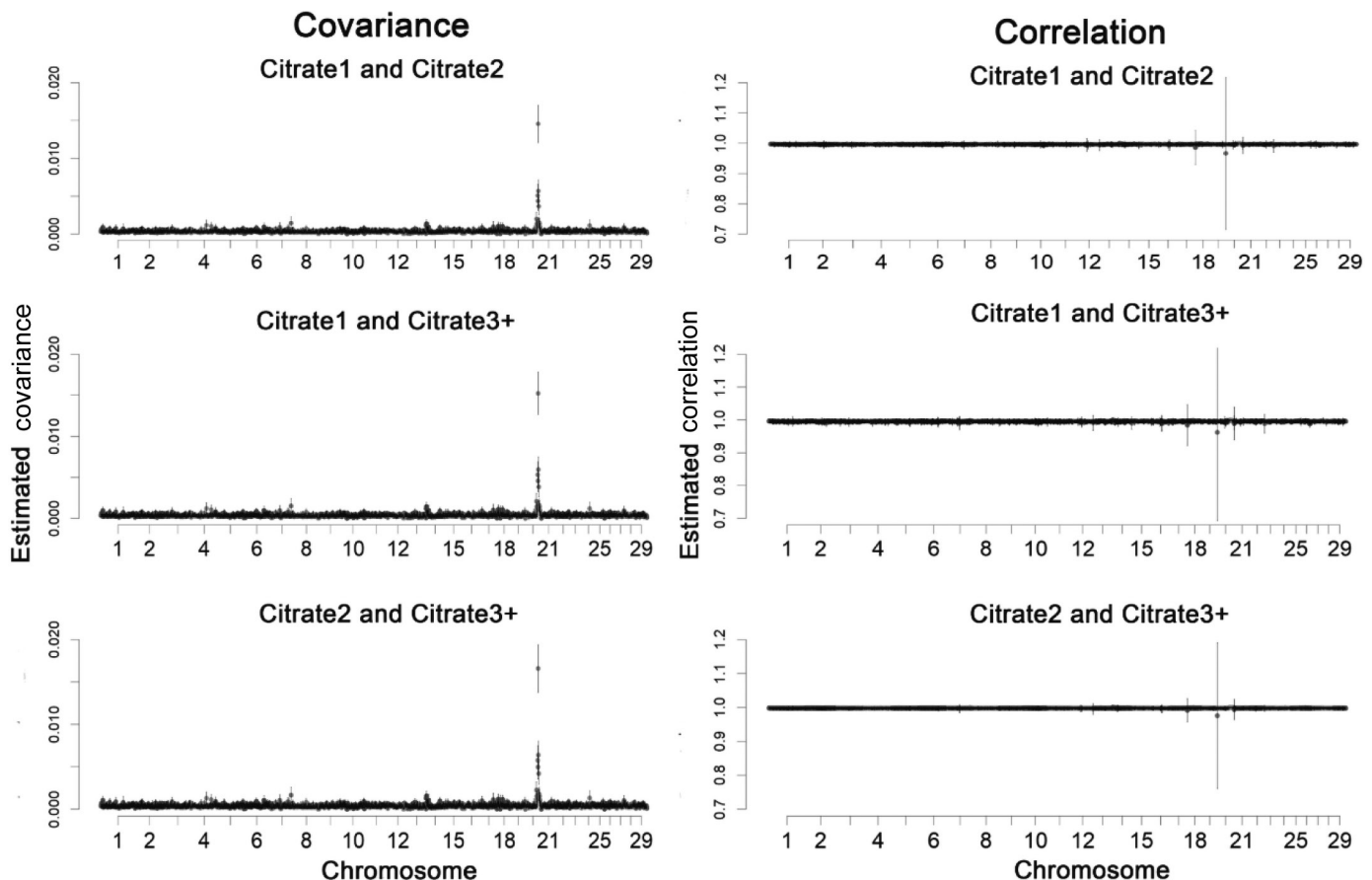


Figure 2. Genomic covariances (left) and correlations (right) among 3 classes of milk citrate in the first 5 parities. Citrate1 = milk citrate in the first parity; citrate2 = milk citrate in the second parity; citrate3+ = milk citrate from third to fifth parity. The black line represents the SE value of the effect in the genetic window (1 Mb).

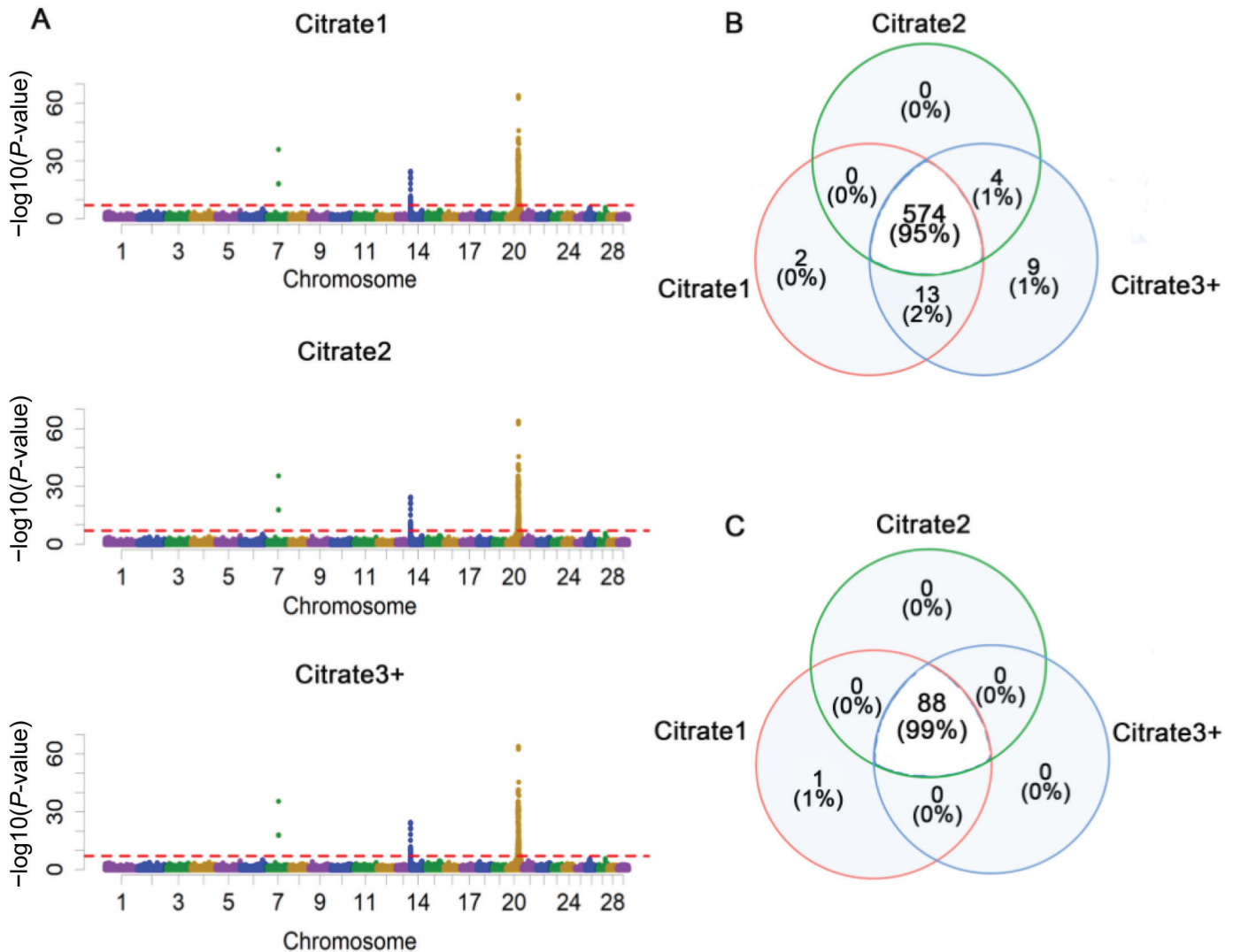


Figure 3. (A) Manhattan plot (P -value of each SNP), (B) number of significant SNPs, and (C) number of candidate genes of milk citrate. Citrate1 = milk citrate in the first parity; citrate2 = milk citrate in the second parity; citrate3+ = milk citrate from third to fifth parity. The red line in A is the threshold: $7.05 = -\log_{10}(0.05/566,170)$.

number of candidate genes in the 3 identified genomic regions is presented in Table 4. Citrate1 lacked only the *BASP1* gene on Chr20, as compared with citrate2 and citrate3+.

Figure 5 shows the PPI networks of the identified candidate genes. The candidate gene connection pairs *ARHGAP39-MYO10*, *DGAT1-ANKH*, *SHARPIN-OTULIN*, and *BOP1-ZNF622* represent interactions between candidate genes in Chr14 and Chr20. The GO analyses results are presented in Table 5. The results identified 2 GO terms based on 89 candidate genes located on Chr7, 14, and 20 (response to salt and linear ubiquitin chain assembly complex [LUBAC]), and one GO term based on 16 candidate genes located on Chr20

(inorganic diphosphate transmembrane transporter activity). For KEGG analyses, none of the pathways were found to be associated with the candidate genes for citrate.

The 3 selected genomic regions of citrate1, citrate2, and citrate3+ were significantly associated with the same 32 QTL, respectively (Supplemental File S2; <https://github.com/Yansen0515/citrate>; Chen, 2023). The classes and top 10 of 32 enrichment QTL are shown in Figures 6A and 6C. Milk and fat yields have the highest QTL numbers. In addition, 9 QTL related to milk compositions and cow health traits were obtained when only analyzing the Chr20 53–64 Mb region (Figures 6B and 6D).

Table 4. Significant SNPs and candidate genes of milk citrate (mmol/L) of Holstein cows in the first 50 DIM

Trait	Chr ¹	Position (bp; ARS-UCD1.2)	Significant SNPs	Number of candidate genes	Candidate gene names
Citrate1	7	68,569,696–68,575,260	5	1	<i>ENSBTAG00000052849</i>
	14	146,715–1,896,438	58	72	<i>VPS28, ENSBTAG00000053637, SLC39A4, CPSF1, ADCK5, SLC52A2, FBXL6, TMEM249, SCRT1, DGAT1, HSF1, BOP1, SCX, MROH1, TSSK5, HGH1, WDR97, MAF1, ENSBTAG00000051469, SHARPIN, CYC1, GPAA1, EXOSC4, OPLAH, SMPD5, SPATC1, GRINA, PARP10, PLEC, EPPK1, NRBP2, PUF60, SCRIB, IQANK1, FAM83H, MAPK15, CCDC166, ZNF623, ENSBTAG00000052472, GFUS, PYCR3, TIGD5, TOP1MT, ZNF696, GLL4, GPIHBP1, LY6H, ENSBTAG00000054483, ENSBTAG00000037824, LY6K, LY6D, ADGRB1, TSNARE1, OR10AG83, ZNF250, ZNF16, C14H8orf33, ZNF34, RPL8, ZNF7, COMMD5, ARHGAP39, C14H8orf82, LRRC24, LRRC14, MFSD3, GPT, PPP1R16A, FOXH1, KIFC2, CYHR1, TONSL</i>
	20	53,998,037–64,284,738	526	16	<i>ENSBTAG00000052828, ENSBTAG00000012971, ENSBTAG00000053528, BASP1, MYO10, RETREG1, ZNF622, MARCHF11, FBXL7, ANKH, OTULIN, OTULINL, TRIO, ENSBTAG00000054860, DNAH5, ENSBTAG00000049263</i>
Citrate2	7	a ²	a	a	a
	14	a	46	a	a
	20	a	527	15	b ³
Citrate3+	7	a	a	a	a
	14	a	61	a	a
	20	a	534	15	b

¹Chr = chromosome.²The letter a represents the same data as in citrate1.³The letter b represents the same data as in citrate1, except that *BASP1* is missing.

DISCUSSION

Citrate Potential for Genetic Selection

Citrate was considered as a proxy for NEB because it can be predicted accurately and cheaply on a large scale (Grelet et al., 2016), which is one of the requirements for the possibility of implementing genetic selection. The citrate was predicted by milk MIR and its prediction equation was suggested to be used for quantitative screening at the individual level (Grelet et al., 2021). Milk MIR is very inexpensive to obtain because it has been used worldwide to predict milk protein and fat percentages (Gengler et al., 2016). The range for predicted milk citrate obtained in this study was within the range of the reference data (directly measured values) from Grelet et al. (2016). Average citrate was relatively stable between 20 and 32 DIM (Figure 1), which is possibly due to a relatively stable period between cattle intake and metabolic needs. Our predicted mean values (8.93–9.18 mmol/L) are higher than the predicted mean value (8.27 mmol/L for first parity) reported for Montbéliarde by Sanchez et al. (2018). The difference could be because we only used data from early lactation (first 50 DIM), and they used data from the entire lactation (from 7 to 350 DIM).

The heritabilities of citrates in Holstein cows are high (0.35–0.40), which is largely high enough for genetic selection. However, the heritabilities in this study were lower than those previously reported for Montbéliarde cows (0.48; Sanchez et al., 2018, 2021), which may be due to breed differences. The heritabilities of citrates in this study decreased with increasing parity, which may be caused by the gradual increase in nongenetic effects, especially the residual variance (Table 2). The SD of citrate also increased with increasing parity, which also helps explain the increase in phenotypic variation of citrate (Table 1).

Genetic correlations among citrate in different parities in early lactation were close to 1.00, which suggests that the 3 classes of citrates can be considered as a single trait. Also, the results of ssGWAS and functional annotation confirm that citrate in the different parities is similar.

Genomic Background of Citrate

A total of 603 significant SNPs associated with citrate were identified, distributed in Chr7 (68.57–68.58 Mb), 14 (0.15–1.90 Mb), and 20 (54.00–64.28 Mb). The Chr7 region contains 5 consecutive SNPs in high LD (>0.80), suggesting that this region might be caused

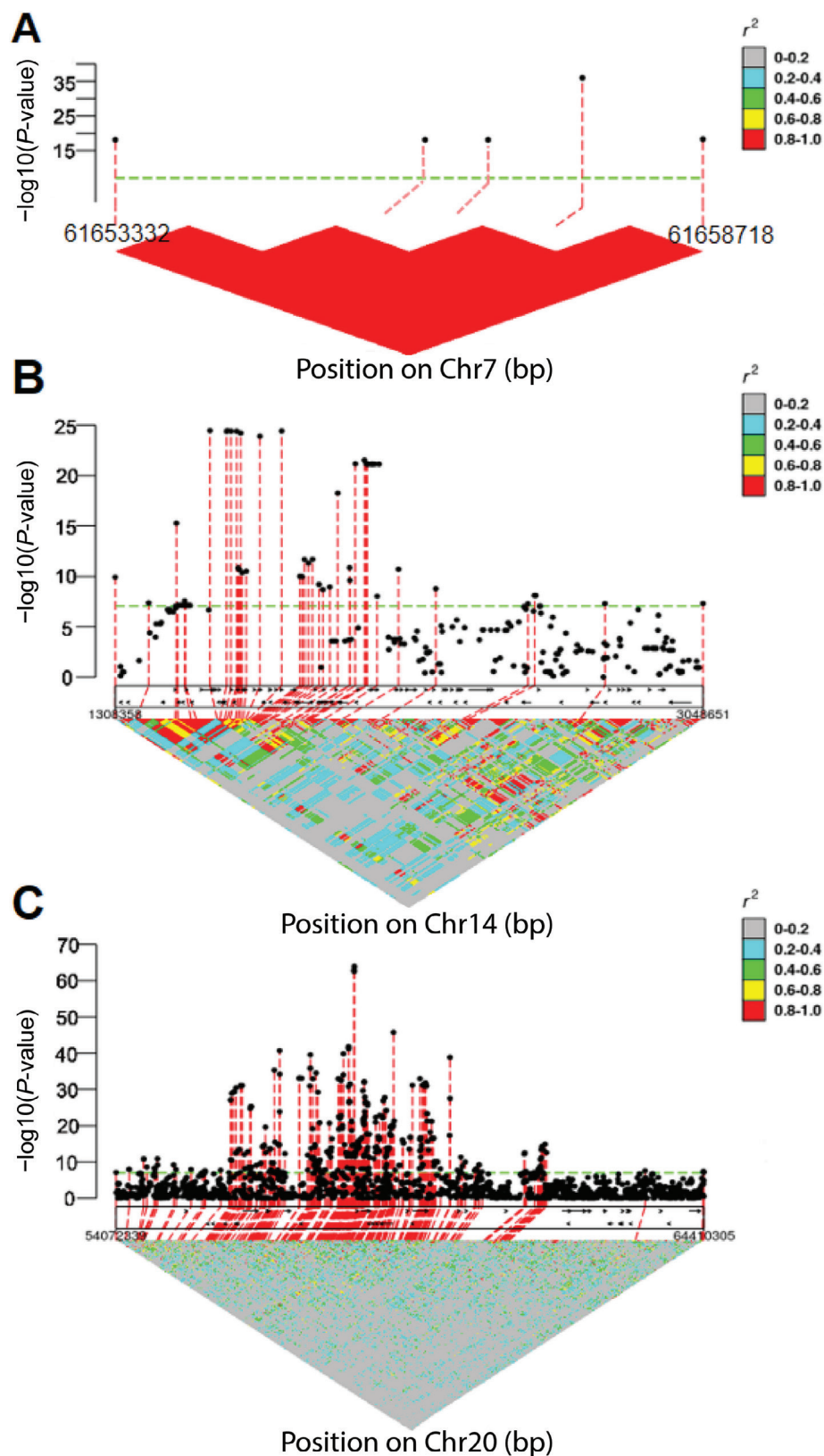


Figure 4. Linkage disequilibrium between 61.656 Mb and 61.658 Mb, 1.31 Mb and 3.05 Mb, 54.00 Mb and 64.28 Mb on (A) chromosome (Chr) 7, (B) Chr14, and (C) Chr20 in genome assembly UMD3.1 associated with milk citrate in the first parity. The green line in A is the threshold: $7.05 = -\log_{10}(0.05/566,170)$.

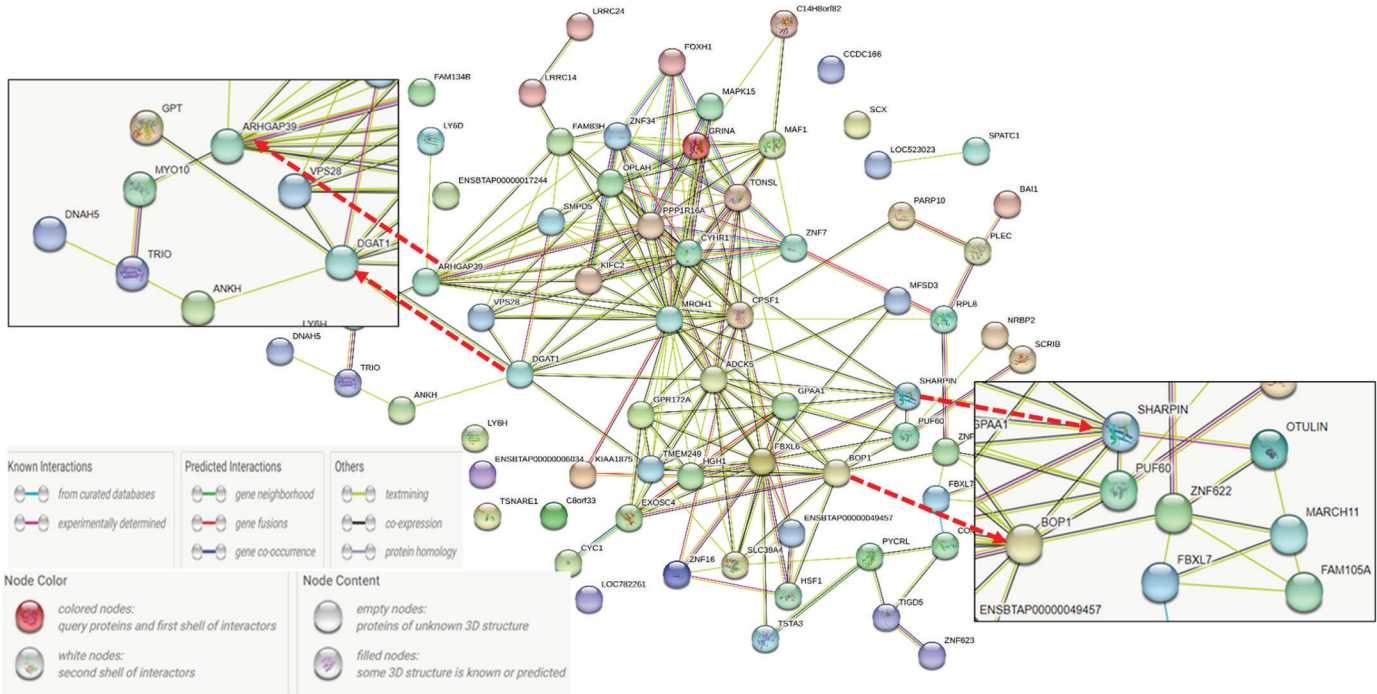


Figure 5. Protein–protein interaction network of candidate genes of milk citrate. The insets link candidate genes in chromosome 14 to candidate genes in chromosome 20.

by one SNP (Chr7 68,574,155 bp) with the highest significance. The Chr14 region has a relatively small effect on citrate compared with the Chr20 region, which showed the most significant effect on citrate. Sanchez et al. (2021) found similar results in Montbéliarde cows. The region on Chr20 (55–63 Mb in UMD3.1) was also found to be associated with subclinical ketosis (Nayeri et al., 2019). In contrast to our findings, Sanchez et al. (2021) identified a genomic locus on Chr11 associated with citrate. This difference found in Chr regions may be explained by the different citrate content in breeds of cattle (Sundekilde et al., 2011).

Among the 89 candidate genes identified for citrate in this study, *GPT*, *ANKH*, and *PPP1R16A* have previously been reported to be related to milk citrate (Sanchez et al., 2019, 2021). Indeed, *GPT* encodes the glutamic pyruvic transaminase that converts cytosolic

pyruvate and glutamate into α -ketoglutarate and alanine (Abla et al., 2020). The resulting α -ketoglutarate can then be converted into cytosolic citrate (Yoshimi et al., 2016). The *ANKH* encodes a membrane protein known to export pyrophosphate, but it seems also able to export citrate extracellularly (Szeri et al., 2020). Concerning *PPP1R16A*, it encodes a protein that interacts with the catalytic subunit of Ser/Thr protein phosphatase 1 (**PP1**; Wang et al., 2019). Numerous potential substrates of PP1 were studied including the ATP-citrate lyase that converts citrate into oxaloacetate and acetyl-CoA (Ingebritsen and Cohen, 1983).

The *DGAT1* was also identified as a candidate gene and is known to affect the blood NEFA (Oikonomou et al., 2009), which provides acetyl-CoA for citrate synthesis (Van et al., 2020); *OTULIN* and *SHARPIN* are associated with individual immunity and inflammation

Table 5. Gene ontology (GO) terms enrichment analysis of candidate genes of milk citrate of Holstein cows in the first 50 DIM

Number of candidate genes	Ontology	ID	Description	Adjusted P-value	Genes involved
89 from chromosome 7, 14, 20	Biological process	GO:1902074	Response to salt	0.0047	<i>ANKH</i> , <i>HSF1</i> , <i>ENSBTAG00000054022</i>
	Cellular component	GO:0071797	LUBAC complex	0.0147	<i>OTULIN</i> , <i>SHARPIN</i>
	Molecular function	GO:0030504	Inorganic diphosphate transmembrane transporter activity	0.0497	<i>ANKH</i>
16 from chromosome 20					

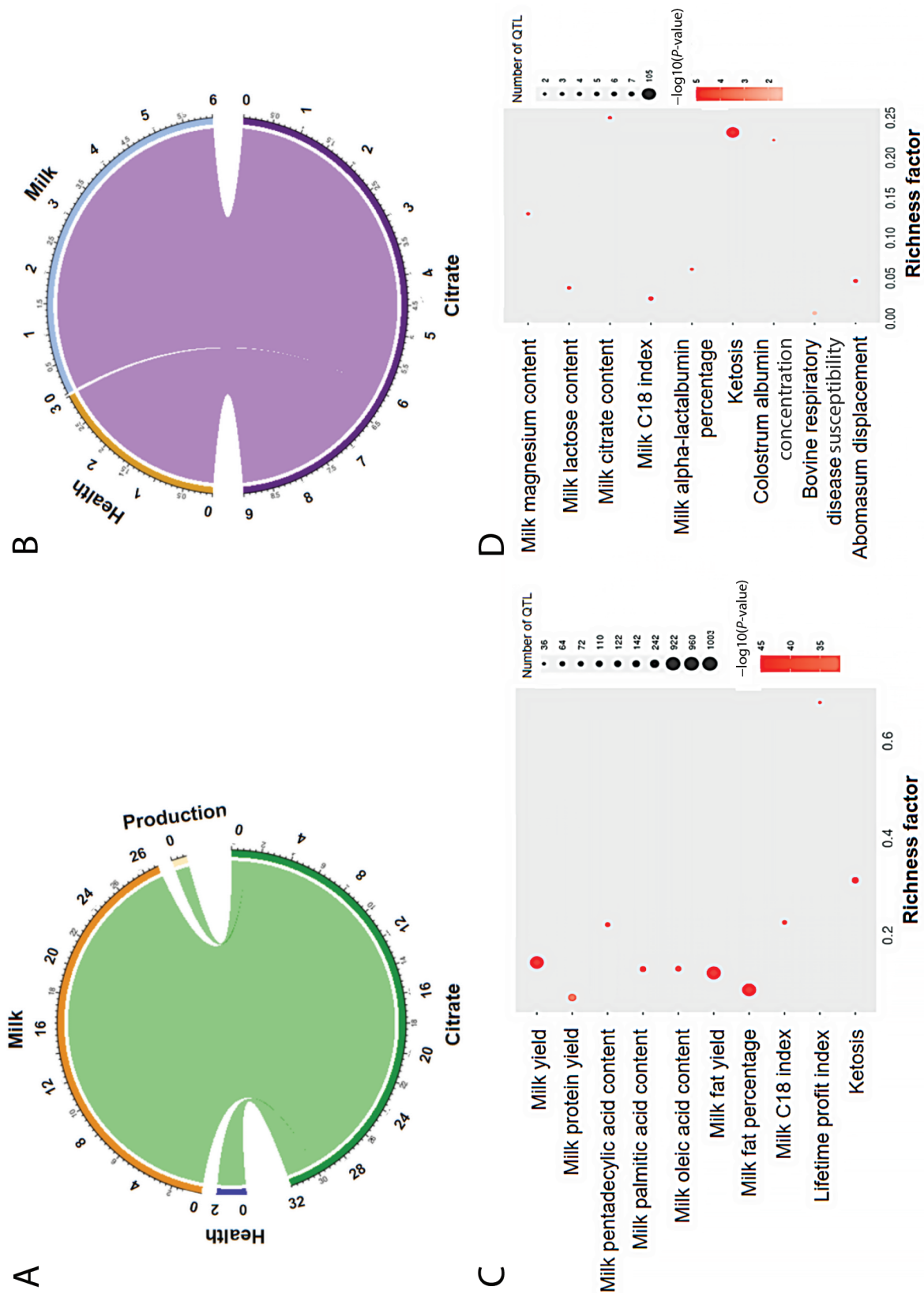


Figure 6. Number of enrichment QTL classes and top 10 (C) or 9 (D) QTL associated with significant SNPs in 3 chromosomes (7, 14, and 20; A and C) or chromosome 20 only (B and D) of milk citrate.

(Damgaard et al., 2016; Zinngrebe et al., 2016) and citrate is considered a key immunometabolite (Zotta et al., 2020). The PPI network of the identified 89 candidate genes shows an interaction between Chr14 and Chr20 regions (Figure 5), which helps in understanding the relationship between genomic regions that regulate citrate.

As for the GO analysis, 2 GO terms (response to salt and inorganic bisphosphate transmembrane transporter activity) were enriched by the genes associated with citrate, probably because citrate regulates the balance between Ca^{2+} and H^{+} ions (Sundekilde et al., 2011; Cánovas et al., 2013). The third enriched GO term, the LUBAC, stimulates the formation of linear ubiquitin chains, a signal that prevents inflammation and modulates immunity (Gerlach et al., 2011). This may be related to metabolic disease in cows that are in NEB for a long time.

Potential Relationship Between Citrate and Other Traits

The top 10 of 32 significant QTL that were identified via QTL enrichment analysis are mainly related to milk and protein yield, and FA content, which may be the effect of genes located on Chr14 (0.15–1.90 Mb; e.g., *DGAT1*). This region has been reported to be associated with milk and protein yield, as well as FA in Holstein cows (Bouwman et al., 2011; Atashi et al., 2020). The top 10 significant QTL were all directly related to the NEB of dairy cows (Xu et al., 2018; Gross and Bruckmaier, 2019; Pires et al., 2022), except for lifetime profit index. However, NEB is related to the production and reproduction performance of dairy cows, both of which are directly related to the lifetime profit index (VanRaden, 2004). Six out of the top 10 significant QTL are related to FA; these results were expected because citrate participates in the NEB of dairy cows by regulating FA. McCabe et al. (2012) reported differential gene expression in the liver of dairy cows with severe NEB primarily associated with FA metabolism pathways.

When focusing on Chr20, 9 significant QTL were associated with milk composition and cow health (Figure 6B and 6D). As expected, milk citrate-related QTL from other reports (Sanchez et al., 2019, 2021) were enriched. The milk C18 index and lactose content change because of the NEB of dairy cows (Xu et al., 2018; Churakov et al., 2021). The enrichment of milk magnesium may be due to its significant association with Chr20 (54.00–64.28 Mb; Sanchez et al., 2021). Amino acid content in milk changes significantly when cows are in NEB (Xu et al., 2018, 2020). This may be the reason for the enrichment of milk α -lactalbumin per-

centage and colostrum albumin concentration. Numerous diseases including ketosis, abomasum displacement, and subacute respiratory acidosis are associated with the state of NEB (Esposito et al., 2014). These can explain why we found 3 of 9 significant QTL linked to citrate and diseases.

Direction of Selection for Citrate Levels

Through the above Discussion section, we can conclude that citrate has a potential correlation with the traits related to the NEB of dairy cows. This indirectly helps to demonstrate that citrate can be a potential proxy of NEB of dairy cows. This suggests an interest in selecting cows with lower citrate levels to indirectly select cows with better energy balance at the beginning of lactation. However, citrate levels are also associated with other issues of dairy cows. It has been reported that a decrease of milk citrate is observed during mastitis (Hyvönen et al., 2010). Choosing animals with lower citrate levels may increase the likelihood of selecting individuals with a higher incidence of mastitis. An alternative is to select animals with lower citrate levels only during early lactation, as this period is often linked to NEB, and the decrease in citrate levels is more clearly associated with mastitis during late lactation (Hyvönen et al., 2010). From a different point of view, different milk citrate levels are observed during heat stress events and were higher or lower depending on the study (Tian et al., 2016; Yue et al., 2020). Selecting lower milk citrate levels could also affect cow thermotolerance. This is supported by *HSF1*, a candidate gene in our study, which has been implied in the thermotolerance (Li et al., 2011; Lemal et al., 2023).

Future research perspectives could involve extending the genetic analysis to cover the entire lactation period of citrate, as the h^2 of energy balance varied significantly throughout the lactation period (Liinamo et al., 2012). Furthermore, the genetic correlations of citrate with other traits of interest, especially milk production and reference energy balance or its other proxies, could be explored. This study shows the potential relationship between citrate and other traits, which requires further confirmation of real genetic correlations. In addition, citrate is a proxy of NEB, so we need to investigate the genetic correlation between citrate and NEB or its other proxies (e.g., BHBA). It is important to know how citrate relates to other traits of interest before including it in a breeding program.

CONCLUSIONS

This study showed that milk citrate predicted by milk MIR spectra has the potential to be used as a proxy of

NEB in dairy cows from an animal-breeding perspective. If citrate is proved as a suitable proxy for NEB, this could be included as a selection index trait, helping to reduce the disease incidence in Holstein cows.

ACKNOWLEDGMENTS

The China Scholarship Council (Beijing, China) is acknowledged for funding the PhD project of Yansen Chen and Hongqing Hu. The genotyping of animals (CDR “PREDICT-2 grant number J.0174.18) and the computation resources of the University of Liège–Gembloux Agro-Bio Tech (ULiège-GxABT, Gembloux, Belgium) were partly supported by the Fonds de la Recherche Scientifique (FRS-FNRS, Brussels, Belgium) which provided also support through the PDR projects “HTwoTHI” (grant number T.W005.23) and “DEEPSELECT” (grant number T.0095.19). The authors acknowledge the support of the Walloon Government (Service Public de Wallonie—Direction Générale Opérationnelle Agriculture, Ressources Naturelles et Environnement, SPW-DGARNE; Namur, Belgium) and the use of the computation resources of the ULiège-GxABT provided by the technical platform Calcul et Modélisation Informatique (CAMI) of the TERRA Teaching and Research Centre (Gembloux, Belgium). The authors are grateful to Jian Cheng (Bayer Crop-Science, Chesterfield, MO), Jiayi Qu (University of California–Davis, Davis, CA), and Hao Cheng (University of California–Davis, Davis, CA) for their help when we used the JWAS program. The high-resolution Figures 1 to 6 may be found in the Mendeley Dataset (<https://data.mendeley.com/datasets/2gbtt9xnx6/1>). Author contributions: YC was responsible for the study design, data analysis and writing of the manuscript; HH estimated the genetic parameters of citrate and revised the manuscript; HA provided help in genotype data imputation and manuscript revision; CG provided the milk citrate data and revised the manuscript; KW and PL revised the manuscript; and NG supervised the project and provided help in the study design and manuscript revision. The authors have not stated any conflicts of interest.







REFERENCES

- Abla, H., M. Sollazzo, G. Gasparre, L. Iommarini, and A. M. Porcelli. 2020. The multifaceted contribution of α -ketoglutarate to tumor progression: An opportunity to exploit? *Semin. Cell Dev. Biol.* 98:26–33. <https://doi.org/10.1016/j.semcdb.2019.05.031>.
- Aguilar, I., A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco, and I. Misztal. 2019. Frequentist p-values for large-scale single step genome-wide association, with an application to birth weight in American Angus cattle. *Genet. Sel. Evol.* 51:28. <https://doi.org/10.1186/s12711-019-0469-3>.
- Atashi, H., M. Salavati, J. De Koster, J. Ehrlich, M. Crowe, G. Op-somer, and E. Gplus. Consortium, and M. Hostens. 2020. Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. *J. Anim. Breed. Genet.* 137:292–304. <https://doi.org/10.1111/jbg.12442>.
- Bjerre-Harpøth, V., N. C. Friggens, V. M. Thorup, T. Larsen, B. M. Damgaard, K. L. Ingvarsen, and K. M. Moyes. 2012. Metabolic and production profiles of dairy cows in response to decreased nutrient density to increase physiological imbalance at different stages of lactation. *J. Dairy Sci.* 95:2362–2380. <https://doi.org/10.3168/jds.2011-4419>.
- Bouwman, A. C., H. Bovenhuis, M. H. P. W. Visker, and J. A. M. van Arendonk. 2011. Genome-wide association of milk fatty acids in Dutch dairy cattle. *BMC Genet.* 12:43. <https://doi.org/10.1186/1471-2156-12-43>.
- Buitenhuis, A. J., U. K. Sundekilde, N. A. Poulsen, H. C. Bertram, L. B. Larsen, and P. Sørensen. 2013. Estimation of genetic parameters and detection of quantitative trait loci for metabolites in Danish Holstein milk. *J. Dairy Sci.* 96:3285–3295. <https://doi.org/10.3168/jds.2012-5914>.
- Cánovas, A., P. A. S. Fonseca, S. McKay, and J. F. Medrano. 2022. Impact of the bovine reference genome (ARS-UCD1.2) on functional annotation at genome, transcriptome and methylome level. *Proc. of 12th World Congr. Genet. Appl. to Livest. Prod.* https://doi.org/10.3920/978-90-8686-940-4_508.
- Cánovas, A., G. Rincón, A. Islas-Trejo, R. Jimenez-Flores, A. Laub-scher, and J. F. Medrano. 2013. RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. *J. Dairy Sci.* 96:2637–2648. <https://doi.org/10.3168/jds.2012-6213>.
- Chen, Y. 2023. Supplemental Files S1 and S2. Git Hub. <https://github.com/Yansen0515/citrate>.
- Chen, Y., S. Vanderick, R. R. Mota, C. Grelet, GplusE Consortium, and N. Gengler. 2021. Estimation of genetic parameters for predicted nitrogen use efficiency and losses in early lactation of Holstein cows. *J. Dairy Sci.* 104:4413–4423. <https://doi.org/10.3168/jds.2020-18849>.
- Cheng, H., R. Fernando, and D. Garrick. 2018b. JWAS: Julia implementation of whole-genome analysis software. *Proc. of 11th World Congr. Genet. Appl. to Livest. Prod.* 11:859. <https://interbull.org/static/web/1515DrHaoChengWCGALP2018.pdf>.
- Cheng, H., K. Kizilkaya, J. Zeng, D. Garrick, and R. Fernando. 2018a. Genomic prediction from multiple-trait Bayesian regression methods using mixture priors. *Genetics* 209:89–103. <https://doi.org/10.1534/genetics.118.300650>.
- Cheng, J., R. Fernando, H. Cheng, S. D. Kachman, K. Lim, J. C. S. Harding, M. K. Dyck, F. Fortin, G. S. PlastowPigGen Canada, and J. C. M. Dekkers. 2022. Genome-wide association study of disease resilience traits from a natural polymicrobial disease challenge model in pigs identifies the importance of the major histocompatibility complex region. *G3 Genes Gen. Genet.* 12:jkab441. <https://doi.org/10.1093/g3journal/jkab441>.
- Churakov, M., J. Karlsson, A. Edvardsson Rasmussen, and K. Holtenius. 2021. Milk fatty acids as indicators of negative energy balance of dairy cows in early lactation. *Animal* 15:100253. <https://doi.org/10.1016/j.animal.2021.100253>.
- Coffey, M. P., G. C. Emmans, and S. Brotherstone. 2001. Genetic evaluation of dairy bulls for energy balance traits using random regression. *Anim. Sci.* 73:29–40. <https://doi.org/10.1017/S1357729800058021>.
- Damgaard, R. B., J. A. Walker, P. Marco-Casanova, N. V. Morgan, H. L. Titheradge, P. R. Elliott, D. McHale, E. R. Maher, A. N. J. McKenzie, and D. Komander. 2016. The deubiquitinase OTULIN is an essential negative regulator of inflammation and autoimmunity. *Cell* 166:1215–1230.e20. <https://doi.org/10.1016/j.cell.2016.07.019>.
- Esposito, G., P. C. Irons, E. C. Webb, and A. Chapwanya. 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. *Anim. Reprod. Sci.* 144:60–71. <https://doi.org/10.1016/j.anireprosci.2013.11.007>.

- Fonseca, P. A. S., A. Suárez-Vega, G. Marras, and Á. Cánovas. 2020. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. *Gigascience* 9:giaa149. <https://doi.org/10.1093/gigascience/giaa149>.
- Friggensen, N. C., C. Ridder, and P. Løvendahl. 2007. On the use of milk composition measures to predict the energy balance of dairy cows. *J. Dairy Sci.* 90:5453–5467. <https://doi.org/10.3168/jds.2006-821>.
- Garnsworthy, P. C., L. L. Masson, A. L. Lock, and T. T. Mottram. 2006. Variation of milk citrate with stage of lactation and de novo fatty acid synthesis in dairy cows. *J. Dairy Sci.* 89:1604–1612. [https://doi.org/10.3168/jds.S0022-0302\(06\)72227-5](https://doi.org/10.3168/jds.S0022-0302(06)72227-5).
- Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41:55. <https://doi.org/10.1186/1297-9686-41-55>.
- Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M. L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. *J. Dairy Sci.* 99:4071–4079. <https://doi.org/10.3168/jds.2015-10140>.
- Gerlach, B., S. M. Cordier, A. C. Schmukle, C. H. Emmerich, E. Riser, T. L. Haas, A. I. Webb, J. A. Rickard, H. Anderton, W. W. L. Wong, U. Nachbur, L. Gangoda, U. Warnken, A. W. Purcell, J. Silke, and H. Walczak. 2011. Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature* 471:591–596. <https://doi.org/10.1038/nature09816>.
- Grelet, C., C. Bastin, M. Gelé, J. B. Davière, M. Johan, A. Werner, R. Reding, J. A. Fernandez Pierna, F. G. Colinet, P. Dardenne, N. Gengler, H. Soyeurt, and F. Dehareng. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β -hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. *J. Dairy Sci.* 99:4816–4825. <https://doi.org/10.3168/jds.2015-10477>.
- Grelet, C., P. Dardenne, H. Soyeurt, J. A. Fernandez, A. Vanlierde, F. Stevens, N. Gengler, and F. Dehareng. 2021. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. *Methods* 186:97–111. <https://doi.org/10.1016/j.ymeth.2020.07.012>.
- Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. *J. Dairy Sci.* 98:2150–2160. <https://doi.org/10.3168/jds.2014-8764>.
- Gross, J. J., and R. M. Bruckmaier. 2019. Review: Metabolic challenges in lactating dairy cows and their assessment via established and novel indicators in milk. *Animal* 13(Suppl. 1):s75–s81. <https://doi.org/10.1017/S175173111800349X>.
- Herd, T. H. 2000. Ruminant adaptation to negative energy balance: Influences on the etiology of ketosis and fatty liver. *Vet. Clin. North Am. Food Anim. Pract.* 16:215–230. [https://doi.org/10.1016/S0749-0720\(15\)30102-X](https://doi.org/10.1016/S0749-0720(15)30102-X).
- Hu, Z. L., C. A. Park, and J. M. Reecy. 2019. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 47:D701–D710. <https://doi.org/10.1093/nar/gky1084>.
- Hyvönen, P., T. Haarahiltunen, T. Lehtolainen, J. Heikkinen, R. Isomäki, and S. Pyörälä. 2010. Concentrations of bovine lactoferrin and citrate in milk during experimental endotoxin mastitis in early- versus late-lactating dairy cows. *J. Dairy Res.* 77:474–480. <https://doi.org/10.1017/S0022029910000579>.
- Ingebritsen, T. S., and P. Cohen. 1983. The protein phosphatases involved in cellular regulation. 1. Classification and substrate specificities. *Eur. J. Biochem.* 132:255–261. <https://doi.org/10.1111/j.1432-1033.1983.tb07357.x>.
- Lemal, P., K. May, S. König, M. Schroyen, and N. Gengler. 2023. Invited review: From heat stress to disease—Immune response and candidate genes involved in cattle thermotolerance. *J. Dairy Sci.* 106:4471–4488. <https://doi.org/10.3168/jds.2022-22727>.
- Li, Q., Z. Ju, J. Huang, J. Li, R. Li, M. Hou, C. Wang, and J. Zhong. 2011. Two novel SNPs in HSF1 gene are associated with thermal tolerance traits in Chinese Holstein cattle. *DNA Cell Biol.* 30:247–254. <https://doi.org/10.1089/dna.2010.1133>.
- Liinamo, A. E., P. Mäntysaari, and E. A. Mäntysaari. 2012. Short communication: Genetic parameters for feed intake, production, and extent of negative energy balance in Nordic Red dairy cattle. *J. Dairy Sci.* 95:6788–6794. <https://doi.org/10.3168/jds.2012-5342>.
- McCabe, M., S. Waters, D. Morris, D. Kenny, D. Lynn, and C. Creevey. 2012. RNA-seq analysis of differential gene expression in liver from lactating dairy cows divergent in negative energy balance. *BMC Genomics* 13:193. <https://doi.org/10.1186/1471-2164-13-193>.
- Meyer, K., and D. Houle. 2013. Sampling based approximation of confidence intervals for functions of genetic covariance matrices. Pages 523–526 in *Proc. of the Association for the Advancement of Animal Breeding and Genetics*. Vol. 20. Armidale, Australia.
- Misztal, I., S. Tsuruta, D. Lourenco, Y. Masuda, I. Aguilar, A. Le-garra, and Z. Vitezica. 2014. Manual for BLUPF90 family of programs. Accessed Jun. 26, 2023. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all8.pdf.
- Nayeri, S., F. Schenkel, A. Fleming, V. Kroezen, M. Sargolzaei, C. Baes, A. Cánovas, J. Squires, and F. Miglior. 2019. Genome-wide association analysis for β -hydroxybutyrate concentration in milk in Holstein dairy cattle. *BMC Genet.* 20:58. <https://doi.org/10.1186/s12863-019-0761-9>.
- Oikonomou, G., K. Angelopoulou, G. Arsenos, D. Zygogiannis, and G. Banos. 2009. The effects of polymorphisms in the DGAT1, leptin and growth hormone receptor gene loci on body energy, blood metabolic and reproductive traits of Holstein cows. *Anim. Genet.* 40:10–17. <https://doi.org/10.1111/j.1365-2052.2008.01789.x>.
- Ospina, P. A., D. V. Nydam, T. Stokol, and T. R. Overton. 2010. Evaluation of nonesterified fatty acids and β -hydroxybutyrate in transition dairy cattle in the northeastern United States: Critical thresholds for prediction of clinical diseases. *J. Dairy Sci.* 93:546–554. <https://doi.org/10.3168/jds.2009-2277>.
- Pires, J. A. A., T. Larsen, and C. Leroux. 2022. Milk metabolites and fatty acids as noninvasive biomarkers of metabolic status and energy balance in early-lactation cows. *J. Dairy Sci.* 105:201–220. <https://doi.org/10.3168/jds.2021-20465>.
- Raudvere, U., L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson, and J. Vilo. 2019. G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47:W191–W198. <https://doi.org/10.1093/nar/gkz369>.
- Sanchez, M. P., M. El Jabri, S. Minéry, V. Wolf, E. Beuvier, C. Laithier, A. Delacroix-Buchet, M. Brochard, and D. Boichard. 2018. Genetic parameters for cheese-making properties and milk composition predicted from mid-infrared spectra in a large data set of Montbéliarde cows. *J. Dairy Sci.* 101:10048–10061. <https://doi.org/10.3168/jds.2018-14878>.
- Sanchez, M. P., Y. Ramayo-Caldas, V. Wolf, C. Laithier, M. El Jabri, A. Michenet, M. Boussaha, S. Taussat, S. Fritz, A. Delacroix-Buchet, M. Brochard, and D. Boichard. 2019. Sequence-based GWAS, network and pathway analyses reveal genes co-associated with milk cheese-making properties and milk composition in Montbéliarde cows. *Genet. Sel. Evol.* 51:34. <https://doi.org/10.1186/s12711-019-0473-7>.
- Sanchez, M. P., D. Rocha, M. Charles, M. Boussaha, C. Hozé, M. Brochard, A. Delacroix-Buchet, P. Grosperin, and D. Boichard. 2021. Sequence-based GWAS and post-GWAS analyses reveal a key role of *SLC37A1*, *ANKH*, and regulatory regions on bovine milk mineral content. *Sci. Rep.* 11:7537. <https://doi.org/10.1038/s41598-021-87078-1>.
- Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478. <https://doi.org/10.1186/1471-2164-15-478>.
- Sundekilde, U. K., P. D. Frederiksen, M. R. Clausen, L. B. Larsen, and H. C. Bertram. 2011. Relationship between the metabolite profile and technological properties of bovine milk from two dairy breeds elucidated by NMR-based metabolomics. *J. Agric. Food Chem.* 59:7360–7367. <https://doi.org/10.1021/jf202057x>.
- Szeri, F., S. Lundkvist, S. Donnelly, U. F. H. Engelke, K. Rhee, C. J. Williams, J. P. Sundberg, R. A. Wevers, R. E. Tomlinson, R. S. Jansen, and K. Van De Wetering. 2020. The membrane protein ANKH is crucial for bone mechanical performance by medi-

- ing cellular export of citrate and ATP. *PLoS Genet.* 16:e1008884. <https://doi.org/10.1371/journal.pgen.1008884>.
- Szklarczyk, D., A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen, and C. von Mering. 2021. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49:D605–D612. <https://doi.org/10.1093/nar/gkaa1074>.
- Tian, H., N. Zheng, W. Wang, J. Cheng, S. Li, Y. Zhang, and J. Wang. 2016. Integrated metabolomics study of the milk of heat-stressed lactating dairy cows. *Sci. Rep.* 6:24208. <https://doi.org/10.1038/srep24208>.
- Van, Q. C. D., E. Knapp, J. L. Hornick, and I. Dufrasne. 2020. Influence of days in milk and parity on milk and blood fatty acid concentrations, blood metabolites and hormones in early lactation Holstein cows. *Animals (Basel)* 10:2081. <https://doi.org/10.3390/ani10112081>.
- VanRaden, P. M. 2004. Invited review: Selection on net merit to improve lifetime profit. *J. Dairy Sci.* 87:3125–3131. [https://doi.org/10.3168/jds.S0022-0302\(04\)73447-5](https://doi.org/10.3168/jds.S0022-0302(04)73447-5).
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- Walsh, S. W., E. J. Williams, and A. C. O. Evans. 2011. A review of the causes of poor fertility in high milk producing dairy cows. *Anim. Reprod. Sci.* 123:127–138. <https://doi.org/10.1016/j.anireprosci.2010.12.001>.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb.)* 94:73–83. <https://doi.org/10.1017/S0016672312000274>.
- Wang, X., M. Obeidat, L. Li, P. Pasarj, S. Aburahess, C. F. B. Holmes, and B. J. Ballermann. 2019. TIMAP inhibits endothelial myosin light chain phosphatase by competing with MYPT1 for the catalytic protein phosphatase 1 subunit PP1c β . *J. Biol. Chem.* 294:13280–13291. <https://doi.org/10.1074/jbc.RA118.006075>.
- Wiggans, G. R., T. S. Sonstegard, P. M. VanRaden, L. K. Matukumalli, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, and C. P. van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J. Dairy Sci.* 92:3431–3436. <https://doi.org/10.3168/jds.2008-1758>.
- Wilmot, H., J. Bormann, H. Soyeurt, X. Hubin, G. Glorieux, P. Mayeres, C. Bertozzi, and N. Gengler. 2022. Development of a genomic tool for breed assignment by comparison of different classification models: Application to three local cattle breeds. *J. Anim. Breed. Genet.* 139:40–61. <https://doi.org/10.1111/jbg.12643>.
- Xu, W., J. Vervoort, E. Saccenti, B. Kemp, R. J. van Hoeij, and A. T. M. van Knegsel. 2020. Relationship between energy balance and metabolic profiles in plasma and milk of dairy cows in early lactation. *J. Dairy Sci.* 103:4795–4805. <https://doi.org/10.3168/jds.2019-17777>.
- Xu, W., J. Vervoort, E. Saccenti, R. van Hoeij, B. Kemp, and A. van Knegsel. 2018. Milk metabolomics data reveal the energy balance of individual dairy cows in early lactation. *Sci. Rep.* 8:15828. <https://doi.org/10.1038/s41598-018-34190-4>.
- Yoshimi, N., T. Futamura, S. E. Bergen, Y. Iwayama, T. Ishima, C. Sellgren, C. J. Ekman, J. Jakobsson, E. Pålsson, K. Kakumoto, Y. Ohgi, T. Yoshikawa, M. Landén, and K. Hashimoto. 2016. Cerebrospinal fluid metabolomics identifies a key role of isocitrate dehydrogenase in bipolar disorder: Evidence in support of mitochondrial dysfunction hypothesis. *Mol. Psychiatry* 21:1504–1510. <https://doi.org/10.1038/mp.2015.217>.
- Yue, S., S. Ding, J. Zhou, C. Yang, X. Hu, X. Zhao, Z. Wang, L. Wang, Q. Peng, and B. Xue. 2020. Metabolomics approach explore diagnostic biomarkers and metabolic changes in heat-stressed dairy cows. *Animals (Basel)* 10:1741. <https://doi.org/10.3390/ani10101741>.
- Zachut, M., M. Šperanda, A. M. De Almeida, G. Gabai, A. Mobasheri, and L. E. Hernández-Castellano. 2020. Biomarkers of fitness and welfare in dairy cattle: Healthy productivity. *J. Dairy Res.* 87:4–13. <https://doi.org/10.1017/S0022029920000084>.
- Zhou, C., C. Li, W. Cai, S. Liu, H. Yin, S. Shi, Q. Zhang, and S. Zhang. 2019. Genome-wide association study for milk protein composition traits in a Chinese Holstein population using a single-step approach. *Front. Genet.* 10:72. <https://doi.org/10.3389/fgene.2019.00072>.
- Zinngrebe, J., E. Rieser, L. Taraborrelli, N. Peltzer, T. Hartwig, H. Ren, I. Kovács, C. Endres, P. Draber, M. Darding, S. Von Karstedt, J. Lemke, B. Dome, M. Bergmann, B. J. Ferguson, and H. Walczak. 2016. LUBAC deficiency perturbs TLR3 signaling to cause immunodeficiency and autoinflammation. *J. Exp. Med.* 213:2671–2689. <https://doi.org/10.1084/jem.20160041>.
- Zotta, A., Z. Zaslon, and L. A. O'Neill. 2020. Is citrate a critical signal in immunity and inflammation? *J. Cell. Signal.* 1:87–96. <https://doi.org/10.33696/Signaling.1.017>.

ORCIDS

- Yansen Chen  <https://orcid.org/0000-0002-8593-4384>
 Hongqing Hu  <https://orcid.org/0000-0002-6408-1845>
 Hadi Atashi  <https://orcid.org/0000-0002-6853-6608>
 Clément Grelet  <https://orcid.org/0000-0003-3313-485X>
 Katrien Wijnrocx  <https://orcid.org/0000-0001-5518-4933>
 Pauline Lemal  <https://orcid.org/0000-0003-4937-0064>
 Nicolas Gengler  <https://orcid.org/0000-0002-5981-5509>