Genome-wide association for milk production traits and somatic cell score in different lactation stages of Ayrshire, Holstein, and Jersey dairy cattle

H. R. Oliveira,^{1,2*} J. P. Cant,¹ L. F. Brito,^{1,3} F. L. B. Feitosa,¹ T. C. S. Chud,¹ P. A. S. Fonseca,¹ J. Jamrozik,^{1,4} F. F. Silva,² D. A. L. Lourenco,⁵ and F. S. Schenkel¹

¹Centre for Genetic Improvement of Livestock (CGIL), Department of Animal Biosciences, University of Guelph, Guelph, Ontario, N1G 2W1, Canada

²Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, 36570-000, Brazil

³Department of Animal Sciences, Purdue University, West Lafayette, IN 47907

⁴Canadian Dairy Network (CDN), Guelph, Ontario, N1K 1E5, Canada

⁵Department of Animal and Dairy Science, University of Georgia, Athens 30602

ABSTRACT

We performed genome-wide association analyses for milk, fat, and protein yields and somatic cell score based on lactation stages in the first 3 parities of Canadian Ayrshire, Holstein, and Jersey cattle. The genome-wide association analyses were performed considering 3 different lactation stages for each trait and parity: from 5 to 95, from 96 to 215, and from 216 to 305 d in milk. Effects of single nucleotide polymorphisms (SNP) for each lactation stage, trait, parity, and breed were estimated by back-solving the direct breeding values estimated using the genomic best linear unbiased predictor and single-trait random regression test-day models containing only the fixed population average curve and the random genomic curves. To identify important genomic regions related to the analyzed lactation stages, traits, parities and breeds, moving windows (SNP-by-SNP) of 20 adjacent SNP explaining more than 0.30% of total genetic variance were selected for further analyses of candidate genes. A lower number of genomic windows with a relatively higher proportion of the explained genetic variance was found in the Holstein breed compared with the Ayrshire and Jersey breeds. Genomic regions associated with the analyzed traits were located on 12, 8, and 15 chromosomes for the Ayrshire, Holstein, and Jersey breeds, respectively. Especially for the Holstein breed, many of the identified candidate genes supported previous reports in the literature. However, well-known genes with major effects on milk production traits (e.g., diacylglycerol O-acyltransferase 1) showed contrasting results among lactation stages, traits, and parities of different breeds. Therefore, our results suggest evidence of differential sets of candidate genes underlying the phenotypic expression of the ana-

lyzed traits across breeds, parities, and lactation stages. Further functional studies are needed to validate our findings in independent populations.

Key words: fat yield, longitudinal trait, milk components, protein yield

INTRODUCTION

Milk production traits, including milk (MY), fat (FY), and protein (PY) yields and SCS, are complex traits influenced by management practices, environmental conditions, animal physiological stage (e.g., age, stage of lactation), and genetic merit of the animals. For instance, adding n-3 fatty acids and CLA to a dairy cow's diet can change the milk fatty acid profile and FY (Bauman et al., 2006; Bionaz and Looor, 2008b). In addition, performing a proper dry-cow therapy and adjusting milking machines with adequate suction pressure can substantially decrease the average SCS in the herd (Nyman et al., 2007; Oliveira et al., 2015). However, it is well established that all of these traits are under hormonal control. In this sense, milk synthesis is directly influenced by the complementary action of several hormones, such as prolactin, somatotropin, IGF-1, corticosteroids, insulin, and thyroid hormone (Baumgard et al., 2017).

Genes related to milk production traits and SCS might be differentially expressed during lactation and across parities, as suggested by changes in genetic correlations (Strabel and Misztal, 1999; Oliveira et al., 2017). Strucken et al. (2012a) reported that the variance explained by certain loci changes from lactation to lactation in German Holstein-Friesian cows. The mammary glands of first-parity cows are still in development, and, therefore, gene expression in this parity might differ from subsequent lactations. Furthermore, some studies have shown that the genetic influence on milk traits varies throughout lactation (Strabel et al., 2005; Brito et al., 2017). Therefore, a better under-

Received February 7, 2019.

Accepted May 13, 2019.

*Corresponding author: holivier@uoguelph.ca

standing of the genetic and biological mechanisms associated with milk production traits and SCS in different lactation stages (e.g., beginning, middle, and end of lactation) and across parities is crucial to enable more efficient genetic selection for these traits, as well as better management practices.

The scientific knowledge concerning lactation biology of the bovine mammary gland in Holstein cattle has advanced substantially over the past decades (Akers, 2017; Almeida and Eckersall, 2018). For instance, studies have shown that the expression of genes related to milk traits, such as *ACSL1*, *AGPAT6*, *FABP3*, *LPIN1*, and *SLC27A6* is affected by lactation stage in Holstein cows of second or greater parity (Bionaz and Looor, 2008a). In addition, Strucken et al. (2012b) performed a genome-wide association study (**GWAS**) for milk production traits in Holstein-Friesian cows, defining the traits based on 10-d intervals over the first 60 lactation days, and reported significant changes in effect sizes over lactation. However, the underlying lactation biology and genes associated with milk traits might differ across breeds and populations, highlighting the need to investigate candidate genes affecting milk traits throughout the lactation stages and parities in different breeds.

In this context, GWAS is essential to identify genetic variants and genomic regions associated with complex traits, which may contribute to understanding their genetic architecture and underlying biology in different breeds. Therefore, the main objectives of this study were to (1) perform GWAS for MY, FY, PY, and SCS based on lactation stages in the first 3 parities of Canadian Ayrshire, Holstein, and Jersey animals; (2) investigate prospective candidate genes related to the mentioned traits in each lactation stage; and (3) identify candidate genes (or genomic regions) and biological processes that are shared among dairy cattle breeds, lactation stages, and parities.

MATERIALS AND METHODS

No animal care committee approval was necessary for the purposes of this study, because all information required was obtained from existing databases. Information from genotyped animals with reliable EBV in the official Canadian genetic evaluation for the Ayrshire, Holstein, and Jersey breeds were used. Details about the official genetic evaluation performed in Canada and the steps performed in this study are described below.

Data Sets

The Canadian Dairy Network (**CDN**; Guelph, ON, Canada) provided the pedigree, genotypes, and the De-

cember 2012 official genetic evaluation data sets, containing the estimates of the random regression genetic coefficients and reliabilities for MY, FY, PY, and SCS from the first 3 parities of Ayrshire, Holstein, and Jersey breeds. Reliabilities made available by CDN were the average reliabilities in the first 3 lactations, calculated as described in Jamrozik et al. (2000). Therefore, reliabilities of EBV for all days in milk (from 5 to 305 d) of each trait were assumed to be the same in this study. In brief, the random regression genetic coefficients and reliabilities were calculated by CDN using 2 multiple-trait random regression test-day models (**RRM**): one considering MY, FY, and PY in the first 3 parities and the other considering SCS in the first 3 parities (Schaeffer et al., 2000). Both models included the fixed effects of herd–test-day–parity and DIM (from 5 to 305 d), fixed regression curves for year–season–region–parity–age of calving effect, and random regression curves for herd–year of calving and additive genetic and permanent environmental effects. Fixed and random regressions were modeled through fourth-order Legendre orthogonal polynomials (Kirkpatrick et al., 1990), and heterogeneity of residual variance was taken into account by fitting 10 different classes in each parity. More details about the RRM used by CDN, as well as the estimated variance components and heritabilities, are presented in Interbull (2018). In summary, average heritabilities estimated by CDN for MY, FY, PY, and SCS were 0.35, 0.27, 0.29, 0.14 for Ayrshire; 0.31, 0.27, 0.27, 0.14 for Holstein; and 0.37, 0.29, 0.31, 0.14 for Jersey in the first lactation; 0.32, 0.26, 0.30, 0.18 for Ayrshire; 0.29, 0.26, 0.27, 0.18 for Holstein; and 0.32, 0.27, 0.29, 0.18 for Jersey in the second lactation; and 0.30, 0.24, 0.28, 0.18 for Ayrshire; 0.26, 0.24, 0.24, 0.21 for Holstein; and 0.31, 0.26, 0.28, 0.21 for Jersey in the third lactation (Interbull, 2018). The total number of animals included in the December 2012 official genetic evaluations was 188,045, 4,888,808, and 134,631 for the Ayrshire, Holstein, and Jersey breeds, respectively.

Genotypes

Animals were initially genotyped with or imputed to the Illumina BovineSNP50K BeadChip (Illumina Inc., San Diego, CA) using the FImpute software (Sargolzaei et al., 2011). Details about the genotyping quality control performed before the imputation, and the imputation accuracy are presented in Larmer et al. (2014). After the initial genotype quality control and imputation process performed by CDN, SNP with Mendelian conflicts, call rate <0.95, minor allele frequency (**MAF**) <0.01, and with a difference between observed and expected heterozygosity frequency >0.15 (Wiggans et al., 2009) were removed using the

Table 1. Total number of genotyped animals (N, i.e., bulls and cows) and bulls (N bulls), years of birth, and mean (SD) of reliability of EBV (REL_{EBV}), for each breed and trait

Trait	Parameter	Ayrshire ¹	Holstein ²	Jersey ¹
Milk yield	N	1,748	10,798	973
	N bulls	526	7,322	546
	Year of birth	1961 to 2010	1957 to 2009	1960 to 2010
	REL_{EBV} ³ (SD)	0.74 (0.13)	0.85 (0.12)	0.74 (0.15)
Fat yield	N	1,715	12,034	917
	N bulls	537	7,325	529
	Year of birth	1961 to 2010	1957 to 2009	1960 to 2010
	REL_{EBV} (SD)	0.72 (0.14)	0.83 (0.12)	0.73 (0.15)
Protein yield	N	1,716	8,865	917
	N bulls	537	7,289	527
	Year of birth	1961 to 2010	1957 to 2008	1960 to 2010
	REL_{EBV} (SD)	0.72 (0.13)	0.88 (0.10)	0.73 (0.15)
SCS	N	1,141	7,329	524
	N bulls	526	7,098	430
	Year of birth	1961 to 2009	1957 to 2008	1960 to 2008
	REL_{EBV} (SD)	0.68 (0.17)	0.88 (0.06)	0.74 (0.16)

¹Minimum REL_{EBV} equal to 0.50.²Minimum REL_{EBV} equal to 0.65.³Average reliability in the first 3 lactations.

preGSf90 software (Aguilar et al., 2014). The number of SNP on the 29 bovine autosomes that remained after the quality control was 41,517, 43,556, and 38,194, for the Ayrshire, Holstein, and Jersey breeds, respectively. Only animals with minimum reliability of EBV equal to 0.65 for Holstein and 0.50 for Ayrshire and Jersey breeds were included in the genomic analyses. The number of animals that remained in the analyses, birth year ranges, and average reliability of EBV for each breed and trait, are shown in Table 1.

Prediction of Daily Direct Genomic Values

The steps to estimate daily direct genomic values (**DGV**) are fully presented in Oliveira et al. (2019). In brief, genomic random regression coefficients were predicted using RRM and the genomic BLUP (**GBLUP**) approach (VanRaden, 2008). For this purpose, random regression genetic coefficients provided by CDN were used to calculate EBV from 5 to 305 DIM. Thereafter, these daily EBV were deregressed according to VanRaden et al. (2009) and subsequently used as pseudo-phenotypes in the GBLUP analyses. These GBLUP analyses were performed using single-trait RRM (i.e., GBLUP analyses were performed individually for each trait and parity within each breed), as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{g} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of pseudo-phenotypes (i.e., deregressed daily EBV); $\mathbf{b}$ is the vector of fixed regression coefficients, which models the average population curve; $\mathbf{g}$ is the vector of random regression genomic coefficients; and $\mathbf{e}$ is the vector of residuals. Both $\mathbf{b}$ and

$\mathbf{g}$ were modeled over DIM by fourth-order Legendre orthogonal polynomials (Kirkpatrick et al., 1990); $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices for $\mathbf{b}$ and $\mathbf{g}$, respectively. The model assumptions are

$$E[\mathbf{y}] = \mathbf{X}\mathbf{b}$$

and

$$Var \begin{bmatrix} \mathbf{g} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \otimes \mathbf{G}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix},$$

where $E[\mathbf{y}]$ is the expectation of $\mathbf{y}$, $\mathbf{G}_0$ is the variance-covariance matrix of the genomic random regression coefficients; σ_e^2 is the residual variance (assumed homogeneous across DIM), $\mathbf{I}$ is an identity matrix, and $\mathbf{G}$ is the genomic relationship matrix created as proposed by VanRaden (2008):

$$\mathbf{G} = 0.95 \frac{\mathbf{W}\mathbf{W}'}{2 \sum_{k=1}^K p_k (1 - p_k)} + 0.05\mathbf{A},$$

where $\mathbf{W} = \mathbf{M} - \mathbf{P}$, in which $\mathbf{M}$ contains the centered genotypes (i.e., -1, 0, and 1 to represent AA, Aa, and aa, respectively); $\mathbf{P}$ contains the allele frequency for SNP k (p_k) in its k th column, expressed as $2(p_k - 0.5)$; $2 \sum_{k=1}^K p_k (1 - p_k)$ is a scaling parameter; and $\mathbf{A}$ is the traditional (pedigree-based) additive relationship matrix. In brief, adding $0.05 \mathbf{A}$ guarantees that $\mathbf{G}$ is positive definite (VanRaden, 2008). The AIREMLF90 software

(Misztal et al., 2002) was used to obtain the variance components and the mixed model solutions for all GB-LUP analyses. Individual reliabilities were not included as weights in the analyses because only animals with relatively accurate EBV were included in this study (Oliveira et al., 2018).

The vector of DGV of each animal i ($\mathbf{DGV}_i$), which included daily DGV from all DIM (from 5 to 305 d) for each trait and parity within each breed, was obtained by multiplying the vector of predicted genomic random regression coefficients by the matrix of Legendre orthogonal polynomial covariates; that is,

$$\mathbf{DGV}_i = \mathbf{T}\hat{\mathbf{g}}_i,$$

where $\hat{\mathbf{g}}_i$ is the vector of predicted genomic coefficients for each animal i , and $\mathbf{T}$ is a matrix of orthogonal covariates associated with the Legendre orthogonal polynomial functions.

GWAS Analyses

The GWAS analyses were performed for all traits, parities, and breeds considering 3 stages of lactation: (1) from 5 to 95 DIM, representing the ascending production stage and lactation peak; (2) from 96 to 215 DIM, representing the lactation persistency stage; and (3) from 216 to 305 DIM, representing the production decline up to the end of the lactation. Therefore, DGV for each lactation stage of each animal i (for each trait and parity) were obtained by summing the daily DGV solutions of the specific DIM; that is,

$$\begin{aligned} DGV1_i &= D\hat{G}V_{i5} + D\hat{G}V_{i6} + \dots + D\hat{G}V_{i95}, \\ DGV2_i &= D\hat{G}V_{i96} + D\hat{G}V_{i97} + \dots + D\hat{G}V_{i215}, \text{ and} \\ DGV3_i &= D\hat{G}V_{i216} + D\hat{G}V_{i217} + \dots + D\hat{G}V_{i305}, \end{aligned}$$

where $DGV1_i$, $DGV2_i$, and $DGV3_i$ are the DGV for the first, second, and third lactation stages of animal i obtained by summing the DGV from 5 to 95, 96 to 215, and 216 to 305 DIM, respectively.

The SNP effects for each lactation stage were estimated individually for each trait, parity, and breed using the postGSf90 software (Aguilar et al., 2014). In summary, the SNP effects were estimated as follows:

$$\begin{aligned} \hat{\mathbf{u}}1 &= \mathbf{M}'[\mathbf{MM}']^{-1} \mathbf{D}\hat{\mathbf{G}}\mathbf{V}1, \\ \hat{\mathbf{u}}2 &= \mathbf{M}'[\mathbf{MM}']^{-1} \mathbf{D}\hat{\mathbf{G}}\mathbf{V}2, \text{ and} \\ \hat{\mathbf{u}}3 &= \mathbf{M}'[\mathbf{MM}']^{-1} \mathbf{D}\hat{\mathbf{G}}\mathbf{V}3, \end{aligned}$$

in which $\hat{\mathbf{u}}1$, $\hat{\mathbf{u}}2$, and $\hat{\mathbf{u}}3$ are the vectors of estimated SNP effects for the first, second, and third lactation stage, respectively; $\mathbf{M}$ is the previously mentioned matrix that contains the centered genotypes, and $\mathbf{D}\hat{\mathbf{G}}\mathbf{V}1$, $\mathbf{D}\hat{\mathbf{G}}\mathbf{V}2$, and $\mathbf{D}\hat{\mathbf{G}}\mathbf{V}3$ are the vectors of estimated DGV, which include the DGV from all animals of the same trait, parity, and breed for the first, second and third lactation stages, respectively.

To identify important genomic regions related to the analyzed traits, moving windows (SNP-by-SNP) of 20 adjacent SNP explaining more than 0.30% of the total genetic variance were selected for further analyses. Genetic variance explained by the SNP k ($\hat{\sigma}_{u,k}^2$) was estimated as follows (Zhang et al., 2010; Fragomeni et al., 2014):

$$\hat{\sigma}_{u,k}^2 = \hat{u}_k^2 2p_k(1 - p_k),$$

in which p_k is the observed allele frequency for the first allele of the SNP k , and $\hat{u}_k^2$ is the estimated effect of SNP k for the analyzed breed, trait, and lactation stage. Therefore, the total genetic variance explained by each genomic window was calculated by summing the variance explained by 20 consecutive SNP.

Because moving windows were used in this study, only the combination of SNP that explained the highest variance was retained for further analyses (i.e., only nonoverlapping windows were retained). This criterion was used to avoid double counting of SNP effects, as suggested by Fragomeni et al. (2014). In addition, a threshold of 0.30% was chosen based on other studies that have investigated complex traits in cattle (Silva et al., 2017; Soares et al., 2017) and to allow the retrieval of important genomic windows for all analyzed traits. Moreover, genomic windows were defined based on 20 adjacent SNP to avoid large noise and absence of peaks due to the use of small and large window sizes, respectively (Fragomeni et al., 2014).

Linkage Disequilibrium

To understand and compare the GWAS results from each breed, we estimated the extension of linkage disequilibrium (measured as r^2 ; Hill and Robertson, 1968) using the PLINK 2.0 software (Chang et al., 2015). For computational ease, LD was determined only at distances <0.50 Mb for all syntenic marker pairs.

Identification of Positional Candidate Genes

To identify possible candidate genes associated with milk production traits and SCS, genes that were located within the most important nonoverlapping genomic

windows (i.e., between the start and end of genomic coordinates) were further investigated. Positional candidate genes were annotated using the Biomart tool (Kinsella et al., 2011) embedded in the Ensembl Genes database version 93 (<http://useast.ensembl.org/index.html>). The bovine genome UMD 3.1 (https://oct2018.archive.ensembl.org/Bos_taurus/Info/Index) was used as the reference genome.

Functional Prioritization and Candidate Genes Identification

To identify the candidate genes and, consequently, to reduce the number of false-positive candidate genes, we performed a functional prioritization analysis using the MeSH ORA tool (Tsuyuzaki et al., 2015), available in the Bioconductor package of the R software (R Core Team, 2011). The MeSH ORA performs an enrichment analysis based on the Medical Subject Headings (MeSH) vocabulary to retrieve statistically overrepresented annotations (ORA); that is, annotations that appear more frequently in a selected gene group relative to their occurrence by chance in a set of reference genes (Nelson et al., 2004). Such MeSH analysis has proven to be useful for deducing meaning from sets of genes across several species, including cattle, swine, horses, and chickens (Morota et al., 2015, 2016). The statistical significance of enrichment analyses was assessed by hypergeometric test (Adams and Skopek, 1987), in which the P -value was given by

$$P = \sum_{i=k}^{\min(M,n)} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}},$$

where N is the total number of genes that were analyzed for each trait (reference genes), M is the total number of selected genes, n is the total number of genes in the MeSH term under study, k is the number of selected genes that belong to the MeSH term under study, and $\binom{y}{x} = \frac{y!}{x!(y-x)!}$ is the binomial coefficient. To increase the power of the hypergeometric test for the smaller breeds (i.e., Ayrshire and Jersey) and reduce the number of false-positive results, all lactation stages, parities, and breeds from a given trait were analyzed together in the MeSH ORA.

Subsequently, selected genes retrieved from the significant ($P < 0.01$) MeSH terms were considered candidate genes for the lactation stage, parity, and breed in which they were identified. Thereafter, to understand

the biological processes related to the studied traits and define the most likely functional genes, complete gene functions were obtained from the National Center for Biotechnology Information database (NCBI; www.ncbi.nlm.nih.gov/gene/) and Animal QTL Database (QTLdb; www.animalgenome.org/cgi-bin/QTLdb/BT/index). The most likely candidate genes that were identified in each lactation stage, trait, and breed were selected for further discussion in this paper.

RESULTS AND DISCUSSION

Identification of Important Genomic Windows and LD

Descriptive statistics of the important genomic windows in each lactation stage, parity, and breed are summarized in Supplemental Tables S1, S2, S3, and S4 (<https://doi.org/10.3168/jds.2019-16451>) for MY, FY, PY, and SCS, respectively. The location of each important genomic window as well as the proportion of genetic variance explained by each one of them are shown in Supplemental Tables S5, S6, S7, and S8 (<https://doi.org/10.3168/jds.2019-16451>) for MY, FY, PY, and SCS, respectively. The number of important genomic windows was higher for Jersey (705), followed by Ayrshire (337) and Holstein (118), which may be related to the sample size for each breed. A smaller sample size usually decreases the power to detect SNP or genomic windows affecting the traits of interest, especially when assuming the same threshold to define genomic windows as important, as done in this study. Therefore, a smaller number of important genomic regions would be expected for Jersey and Ayrshire. However, the observed results might have been caused by an increase in the proportion of animals carrying variants with low MAF in these small-sized breeds, which consequently may affect the estimated SNP effects. This and other limitations of this study are further described in the Comparison Between Breeds, Statistical Challenges, and Opportunities section.

It is worth noting that using a larger genomic window size to analyze the Jersey breed would reduce the number of genomic regions identified in this breed, which will be useful for future GWAS. This is related to the higher average extension of LD in the Jersey breed (0.284) than in the Ayrshire (0.214) and Holstein (0.223) breeds, which is a consequence of the smaller effective population size in Jerseys (Stachowicz et al., 2011). Therefore, the probability of finding the same QTL in 2 consecutive nonoverlapping windows is higher for the Jersey breed, because the window size was limited to 20 consecutive SNP in this study. The levels of LD found in this study were consistent with those reported in the literature (Larmer et al., 2014).

A smaller number of windows with a relatively higher proportion of explained genetic variance was found in the Holstein breed compared with the Ayrshire and Jersey breeds, for MY, FY, and PY. This fact suggests that there might be some major genes being differentially expressed in the Holstein breed compared with the others. Pryce et al. (2010), validating SNP and haplotypes associated with milk production traits in Holstein-Friesian and Jersey sires, also found that the average variation explained by SNP was higher in Holsteins compared with Jerseys. However, those authors suggested that their findings were likely a consequence of the population used to select the SNPs, which was mostly composed of Holstein animals.

Similar numbers of important windows were found between different lactation stages of the same parity and trait for all breeds (Supplemental Tables S1 to S4; <https://doi.org/10.3168/jds.2019-16451>). However, for most traits and breeds, a smaller number of important windows was found for third parity compared with the first and second parities of the same trait. The smaller number of important genomic windows found in third parity suggests that the heritability reduction observed for milk production traits and SCS from the first to the third parity is also related to a reduction in their genetic control, not only to the greater influence of environmental effects. This may be related to the reduction in phenotypic information in later parities, because many cows are culled after the first or second lactation. For all breeds and parities, genomic windows accounting for a low proportion of the genetic variance were found, especially for SCS, which might be related to the more quantitative nature of this trait compared with MY, FY, and PY.

Identification of Candidate Genes

The total number of candidate genes identified in the important nonoverlapping genomic windows for each lactation stage, trait, parity, and breed before and after the prioritization analyses are shown in Supplemental Table S9 (<https://doi.org/10.3168/jds.2019-16451>). In general, the prioritization analyses performed using MeSH ORA reduced the number of candidate genes considerably, which facilitated interpretation of the results. After the prioritization analyses, the statistically significant MeSH ORA terms were associated with 18, 23, and 39 genes grouped in 4 categories (i.e., Anatomy, Chemical and Drugs, Diseases, and Phenomena and Process), for the Ayrshire, Holstein and Jersey breeds, respectively. The statistically significant ($P < 0.01$) MeSH ORA terms and the genes associated with them are shown in Tables 2, 3, 4 and 5 for MY, FY, PY and SCS, respectively.

The proportion of genes shared between the 3 breeds, 4 traits, 3 parities, and 3 lactation stages are shown in Supplemental Table S10 (<https://doi.org/10.3168/jds.2019-16451>). In general, different candidate genes were found among traits and breeds. However, relatively similar genes were identified across lactation stages and parities for the same trait. Genomic regions associated with the analyzed traits were located on 12, 8, and 15 chromosomes for the Ayrshire, Holstein, and Jerseys breeds, respectively (Tables 2 to 5). This result is in agreement with Buitenhuis et al. (2014), who reported candidate genes located in a larger number of chromosomes in Jersey compared with Holstein for milk fat composition in Danish cattle populations. No studies comparing candidate gene results for Ayrshire with Holstein and Jersey breeds were found in the literature.

Candidate Genes for MY. Overall, no similar candidate genes for MY were identified in all 3 breeds. Genomic regions associated with MY were located on BTA11 (*SDC1*), BTA16 (*C4BPA* and *C4BPB*), and BTA28 (*MTR*) for the Ayrshire; BTA6 (*CSN1S1*, *CSN1S2*, and *CSN3*), BTA13 (*ARFRP1*), and BTA14 (*BOP1*, *CPSF1*, *CYHR1*, *DGAT1*, *EEF1D*, *FBXL6*, *FOXH1*, *LRRC24*, *RECQL4*, *TONSL*, and *VPS28*) for the Holstein; and BTA2 (*ARPC2*, *GPBAR1* and *SLC11A1*), BTA5 (*HSP90B1*), BTA11 (*TTL*), BTA20 (*MYO10*), BTA26 (*TCF7L2*), and BTA27 (*SLC25A4*) for the Jersey breed.

Our results suggest that the *C4BPA* (complement component 4 binding protein α -chain) and *C4BPB* (*C4BP* β -chain) genes are related to MY in mid lactation for the Ayrshire breed. The *C4BPA* and *C4BPB* gene products circulate in blood as regulators of the complement system that enhances inflammation and bacterial lysis functions of the immune system (Ricklin et al., 2010). In addition, *SDC1* (syndecan 1) appears to be an important candidate gene related to MY for this breed, especially at the end of the lactation. The *SDC1* gene encodes a transmembrane heparan sulfate proteoglycan that interacts with adhesion molecules, cytokines, coagulation factors, and extracellular matrix proteins. As such, the *SDC1* product is involved in coordinating development and differentiation of epithelial structures during embryogenesis and into adulthood, including those of the mammary glands and uterus. Hambruch et al. (2017) showed that the placental expression of *SDC1* increases at the end of gestation (from 250 to 270 gestation days), which may contribute to the association of *SDC1* with MY, especially at the end of lactation.

Many of the candidate genes detected for MY in the Holstein breed are corroborated by previous literature (Nayeri et al., 2016). In general, several candidate genes appeared consistently in all lactation stages and pari-

Table 2. Statistically significant ($P < 0.01$) Medical Subject Headings (MeSH) terms associated with the selected genes in 3 categories (C1, Anatomy; C2, Chemical and Drugs; C3, Phenomena and Process) for milk yield in each lactation stage, parity, and breed

Parity	Breed ¹	Stage	(Category) MeSH term	Gene
First	AY	1	(C3) Lactation	<i>MTR</i>
		2	(C2) Integrin alphaXbeta2, Protein S	<i>C4BPA, C4BPB</i>
		3	(C2) Integrin alphaXbeta2, Protein S	<i>C4BPA, C4BPB</i>
	HO	1	(C1) Golgi Apparatus, (C2) Adenosine Diphosphate, Amyloid, Bicuculline, Glucagon, Lactalbumin, Luminescent Proteins, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen, ² Prekallikrein, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, (C3) Lactation, Hydrodynamics	<i>BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		2	(C1) Golgi Apparatus, (C2) Adenosine Diphosphate, Amyloid, Bicuculline, Glucagon, Lactalbumin, Luminescent Proteins, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen, ² Prekallikrein, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, (C3) Lactation, Hydrodynamics	<i>BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		3	(C2) Diacylglycerol O-Acyltransferase, (C3) Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
	JE	1	(C2) Adenosine Diphosphate	<i>TTL</i>
		2	(C2) Adenosine Diphosphate	<i>TTL</i>
		3	(C2) Adenosine Diphosphate	<i>TTL</i>
Second	AY	1	—	—
		2	(C2) Integrin alphaXbeta2, Protein S	<i>C4BPA, C4BPB</i>
		3	(C3) Hydrodynamics	<i>SDC1</i>
	HO	1	(C1) Golgi Apparatus, (C2) Adenosine Diphosphate, Amyloid, Bicuculline, Glucagon, Lactalbumin, Luminescent Proteins, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen, ² Prekallikrein, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, (C3) Lactation, Hydrodynamics	<i>ARFRP1, BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		2	(C2) Diacylglycerol O-Acyltransferase, (C3) Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		3	(C2) Diacylglycerol O-Acyltransferase, (C3) Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
	JE	1	(C3) Adenosine Diphosphate, Lactation, Luminescent Proteins	<i>ARPC2, GPBAR1, HSP90B1, MYO10, SLC11A1, SLC25A4, TCF7L2</i>
		2	(C2) Adenosine Diphosphate	<i>TTL</i>
		3	(C2) Adenosine Diphosphate	<i>TTL</i>
Third	AY	1	—	—
		2	(C2) Integrin alphaXbeta2, Protein S	<i>C4BPA, C4BPB</i>
		3	(C3) Hydrodynamics	<i>SDC1</i>
	HO	1	(C1) Golgi Apparatus, (C2) Amyloid, Bicuculline, Glucagon, Lactalbumin, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen ¹ , Prekallikrein, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, (C3) Lactation, Hydrodynamics	<i>BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		2	(C2) Diacylglycerol O-Acyltransferase, (C3) Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		3	(C2) Diacylglycerol O-Acyltransferase, (C3) Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
	JE	1	—	—
		2	(C2) Adenosine Diphosphate	<i>TTL</i>
		3	(C2) Adenosine Diphosphate	<i>TTL</i>

¹AY = Ayrshire; HO = Holstein; JE = Jersey.²Plasminogen: D010958, D010959, D014568, and D055293.

ties, including *CYHR1* (cysteine and histidine rich 1), *DGAT1* (diacylglycerol o-acyltransferase 1), *EEF1D* (eukaryotic translation elongation factor 1 delta), *FBXL6* (F-box and leucine rich repeat protein 6), and

FOXH1 (forkhead box H1). This indicates that those genes are likely affecting physiological processes of milk synthesis along the entire lactation period. However, interestingly, the *CSN1S1* (α_{S1} -casein), *CSN1S2* (α_{S2} -

Table 3. Statistically significant ($P < 0.01$) Medical Subject Headings (MeSH) terms associated with the selected genes in 2 categories (C1, Anatomy; C2, Chemical and Drugs) for fat yield in each lactation stage, parity, and breed

Parity	Breed ¹	Stage	(Category) MeSH term	Gene
First	AY	1	(C1) Kidney Glomerulus, (C2) Follicle Stimulating Hormone, Nucleoplasmins	<i>BMP2, ICAM1, PIN1</i>
		2	(C1) Kidney Glomerulus, (C2) Follicle Stimulating Hormone, Nucleoplasmins	<i>BMP2, ICAM1, PIN1</i>
		3	(C1) Kidney Glomerulus, (C2) Follicle Stimulating Hormone, Myosin Light Chains	<i>ICAM1, PIN1, ROCK2</i>
	HO	1	(C2) Diacylglycerol O-Acyltransferase	<i>CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, RECQL4, TONSL, VPS28</i>
		2	(C2) Diacylglycerol O-Acyltransferase	<i>BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, MGP, RECQL4, TONSL, VPS28</i>
		3	(C2) Diacylglycerol O-Acyltransferase	<i>BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, RECQL4, TONSL, VPS28</i>
	JE	1	(C2) Calcitriol	<i>ANXA2</i>
		2	(C2) Calcitriol	<i>ANXA2</i>
		3	(C2) Calcitriol, Follicle Stimulating Hormone	<i>ANXA2, CYP19A1</i>
Second	AY	1	(C1) Kidney Glomerulus, (C2) Follicle Stimulating Hormone, Nucleoplasmins	<i>BMP2, ICAM1, PIN1</i>
		2	(C1) Kidney Glomerulus, (C2) Follicle Stimulating Hormone, Myosin Light Chains, Nucleoplasmins	<i>BMP2, HIF1A, ICAM1, PIN1, ROCK2</i>
		3	—	—
	HO	1	(C2) Diacylglycerol O-Acyltransferase	<i>BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		2	(C2) Calcitriol, Diacylglycerol O-Acyltransferase	<i>BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MGP, RECQL4, TONSL, VPS28</i>
		3	(C2) Diacylglycerol O-Acyltransferase	<i>BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
	JE	1	(C2) Calcitriol, Follicle Stimulating Hormone, Glucuronosyltransferase, Myosin Light Chains, Interferon Regulatory Factor 3 and 7, Diacylglycerol O-Acyltransferase	<i>ANXA2, BOP1, DGAT1, FAS, HSP90B1, IFIT3, IFIT5, STAB2</i>
		2	(C1) Cumulus Cells, (C2) Calcitriol, Metalloendopeptidases, Glucuronosyltransferase	<i>ANXA2, TMSB4</i>
		3	(C1) Cumulus Cells, (C2) Calcitriol, Metalloendopeptidases, Glucuronosyltransferase	<i>ANXA2, TMSB4</i>
Third	AY	1	(C1) Kidney Glomerulus, (C2) Follicle Stimulating Hormone, Metalloendopeptidases, Protease Inhibitors, Tissue Inhibitor of Metalloproteinase-1, Matrix Metalloproteinase, Nucleoplasmins	<i>BMP2, HIF1A, ICAM1, PIN1, TIMP3</i>
		2	(C1) Kidney Glomerulus, (C2) Follicle Stimulating Hormone, Nucleoplasmins	<i>BMP2, ICAM1, KPNA7, PIN1</i>
		3	Nucleoplasmins	<i>KPNA7</i>
	HO	1	(C2) Diacylglycerol O-Acyltransferase	<i>BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MGP, RECQL4, TONSL, VPS28</i>
		2	(C2) Calcitriol, Diacylglycerol O-Acyltransferase	<i>BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MGP, RECQL4, TONSL, VPS28</i>
		3	(C2) Diacylglycerol O-Acyltransferase	<i>BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
	JE	1	(C2) Calcitriol	<i>ANXA2</i>
		2	(C2) Calcitriol	<i>ANXA2</i>
		3	(C1) Cumulus Cells, (C2) Follicle Stimulating Hormone, Glucuronosyltransferase, Myosin Light Chains, Diacylglycerol O-Acyltransferase	<i>BOP1, CYP26A1, DGAT1, FASLG, HSP90B1, STAB2</i>

¹AY = Ayrshire; HO = Holstein; JE = Jersey.

Table 4. Statistically significant ($P < 0.01$) Medical Subject Headings (MeSH) terms associated with the selected genes in 3 categories (C1, Anatomy; C2, Chemical and Drugs; C3, Phenomena and Process) for protein yield in each lactation stage, parity, and breed

Parity	Breed ¹	Stage	(Category) MeSH term	Gene
First	AY	1	(C1) Colostrum, Kidney Glomerulus, (C3) Lactation	<i>BAX, FCGRT, ICAM1</i>
		2	(C1) Kidney Glomerulus, (C3) Lactation	<i>ICAM1, MTR</i>
		3	(C1) Kidney Glomerulus	<i>ICAM1</i>
	HO	1	(C2) Growth Hormone, Diacylglycerol O-Acyltransferase, (C3) Energy Metabolism, Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		2	(C1) Colostrum, (C2) Aldehyde Reductase, Amyloid, Bicuculline, Cortisone Reductase, Glucagon, Lactalbumin, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen, Prekallikrein, Prolactin, Growth Hormone, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, (C3) Energy Metabolism, Lactation	<i>AKR1C4, BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		3	(C2) Aldehyde Reductase, Cortisone Reductase, Growth Hormone, Diacylglycerol O-Acyltransferase, (C3) Energy Metabolism, Lactation	<i>AKR1C4, BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		1	—	—
		2	—	—
		3	—	—
	JE	1	—	—
		2	—	—
		3	—	—
Second	AY	1	(C1) Kidney Glomerulus, (C2) Prolactin, Growth Hormone, (C3) Lactation	<i>ADRB2, CSH2, ICAM1</i>
		2	—	—
		3	Lactation	<i>PIK3R3</i>
	HO	1	(C2) Growth Hormone, Diacylglycerol O-Acyltransferase, (C3) Energy Metabolism, Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		2	(C2) Growth Hormone, Diacylglycerol O-Acyltransferase, (C3) Energy Metabolism, Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		3	(C2) Aldehyde Reductase, Cortisone Reductase	<i>AKR1C4</i>
		1	—	—
		2	(C3) Energy Metabolism, Lactation	<i>COX4I1, HSP90B1</i>
		3	—	—
Third	AY	1	(C1) Kidney Glomerulus, (C2) Prolactin, Growth Hormone, (C3) Energy Metabolism, Lactation	<i>ADRB2, CSH2, COX4I1, ICAM1</i>
		2	—	—
		3	—	—
	HO	1	(C2) Growth Hormone, Diacylglycerol O-Acyltransferase, (C3) Energy Metabolism, Lactation	<i>BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		2	(C2) Aldehyde Reductase, Cortisone Reductase, Growth Hormone, Diacylglycerol O-Acyltransferase, (C3) Energy Metabolism, Lactation	<i>AKR1C4, BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28</i>
		3	(C2) Aldehyde Reductase, Cortisone Reductase	<i>AKR1C4</i>
		1	—	—
		2	—	—
		3	(C1) Colostrum, Kidney Glomerulus, (C3) Energy Metabolism, Lactation	<i>COL4A3, COX4I1, CYP26A1, HSP90B1</i>

¹AY = Ayrshire; HO = Holstein; JE = Jersey.

casein), and *CSN3* (κ -casein) genes appeared as candidate genes only in the beginning of lactation (i.e., in the first lactation stage) for all 3 parities, suggesting that these genes are related mainly to the beginning of MY, possibly associated with lactogenesis at the onset of lactation. In contrast, *DGAT1* was identified as significant in all lactation stages of the Holstein breed.

The involvement of *DGAT1* variants in increasing fat and protein percentages in milk and decreasing MY is well known (Spelman et al., 2002). Association of the *DGAT1* genotype with parameters of fitted lactation curve equations revealed that the known effects of the *DGAT1* gene only appeared after 40 DIM and that effects before d 40 were opposite to those expected for

Table 5. Statistically significant ($P < 0.01$) Medical Subject Headings (MeSH) terms associated with the selected genes in 3 categories (C1, Anatomy; C2, Chemical and Drugs; C3, Phenomena and Process) for SCS in each lactation stage, parity, and breed

Parity	Breed ¹	Stage	(Category) MeSH term	Gene
First	AY	1	(C3) Lysosomal Storage C3	<i>FUCA1</i>
		2	(C3) Lysosomal Storage C3	<i>FUCA1</i>
		3	—	—
	HO	1	(C2) Receptor Epidermal Growth Factor	<i>SYNJ1</i>
		2	—	—
		3	(C1) Golgi Apparatus, (C2) Large-Conductance Calcium-Activated Potassium Channels, Phosphothreonine, Quercetin, Receptor Epidermal Growth Factor, Calcium Channels, Adiponectin, Amphiregulin, (C4) Energy Metabolism, Hydrodynamics	<i>NOS3, SYNJ1</i>
	JE	1	(C2) Caseins, Receptor Epidermal Growth Factor, Calcium Channels, Serine Proteinase Inhibitors, Amyloid β -Peptides, Diacylglycerol O-Acyltransferase, (C3) Lysosomal Storage C3	<i>CACNA1A, CTSA, CTSD, MMP9, PTK2, SERPINB6, SERPINB9, TUBB2B</i>
		2	(C2) Amyloid β -Peptides, Calcium Channels, Caseins, Diacylglycerol O-Acyltransferase, Large-Conductance Calcium-Activated Potassium Channels, Receptor Epidermal Growth Factor, Receptor Somatotropin, Serine Proteinase Inhibitors, (C3) Lysosomal Storage C3	<i>CACNA1A, CTSA, CTSD, IGF2, MMP9, PTI, PTK2</i>
		3	(C2) Amyloid β -Peptides, Calcium Channels, Caseins, Diacylglycerol O-Acyltransferase, Large-Conductance Calcium-Activated Potassium Channels, Receptor Epidermal Growth Factor, Receptors Somatotropin, Serine Proteinase Inhibitors, (C3) Lysosomal Storage C3	<i>CACNA1A, CTSA, CTSD, IGF2, MMP9, PTI, PTK2</i>
Second	AY	1	(C3) Lysosomal Storage C3	<i>FUCA1</i>
		2	—	—
		3	(C2) Amphiregulin, Receptor Epidermal Growth Factor	<i>NRG1</i>
	HO	1	(C2) Adiponectin, Calcium Channels, Diacylglycerol O-Acyltransferase, Receptor Epidermal Growth Factor, Receptor Somatotropin, (C4) Energy Metabolism	<i>ANGPTL4, BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MCOLN1, RECQL4, RETN, SYNJ1, TONSL, VPS28</i>
		2	(C1) Golgi Apparatus, (C2) Adiponectin, Amyloid, Amyloid β -Peptides, Bicuculline, Calcium Channels, Caseins, Clusterin, Diacylglycerol O-Acyltransferase, Glucagon, Lactalbumin, Pancreatic Hormones, Pancreatin, Pancreatic Elastase, Pituitary Hormones, Plasminogen, Prekallikrein, Quercetin, Receptor Somatotropin, Receptor Dopamine, Receptor Serotonin, Thyroid Hormones, (C4) Energy Metabolism, Hydrodynamics	<i>ANGPTL4, BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MCOLN1, RECQL4, RETN, STATH, TONSL, VPS28</i>
		3	(C1) Golgi Apparatus, (C2) Adiponectin, Amyloid, Amyloid β -Peptides, Bicuculline, Caseins, Calcium Channels, Clusterin, Diacylglycerol O-Acyltransferase, Glucagon, Lactalbumin, Pancreatic Hormones, Pancreatin, Pancreatic Elastase, Pituitary Hormones, Plasminogen, Prekallikrein, Quercetin Receptor Somatotropin, Receptor Dopamine, Receptor Serotonin, Thyroid Hormones, (C4) Energy Metabolism, Hydrodynamics	<i>ANGPTL4, BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MCOLN1, RECQL4, RETN, STATH, TONSL, VPS28</i>
	JE	1	(C1) Golgi Apparatus, (C2) Amyloid β -Peptides, Calcium Channels, Caseins, Diacylglycerol O-Acyltransferase, Large-Conductance Calcium-Activated Potassium Channels, Receptor Epidermal Growth Factor, Receptor Somatotropin, Serine Proteinase Inhibitors, Receptor Antigen T-Cell α - β , (C4) Genes MHC class II	<i>ARCNI, BOLA-DMA, BRD2, CD3E, CD3G, CTSD, IGF2, ITPR3, PLCB1, PTK2, PTI, PSMB9</i>
		2	(C2) Amyloid β -Peptides, Caseins, Calcium Channels, Diacylglycerol O-Acyltransferase, Large-Conductance Calcium-Activated Potassium Channels, Receptor Epidermal Growth Factor, Receptor Somatotropin, Serine Proteinase Inhibitors	<i>CACNA1A, CTSD, IGF2, PLCB1, PTK2, PTI</i>
		3	(C2) Amyloid β -Peptides, Calcium Channels, Caseins, Diacylglycerol O-Acyltransferase, Receptor Epidermal Growth Factor, Receptor Somatotropin	<i>CACNA1A, CTSD, IGF2, PTK2</i>
Third	AY	1	—	—
		2	—	—
		3	(C2) Amphiregulin, Receptor, Epidermal Growth Factor	<i>NRG1</i>
	HO	1	(C2) Large-Conductance Calcium-Activated Potassium Channels	<i>KCNU1</i>
		2	(C2) Large-Conductance Calcium-Activated Potassium Channels	<i>KCNU1</i>
		3	(C2) Large-Conductance Calcium-Activated Potassium Channels	<i>KCNU1</i>
	JE	1	(C1) Golgi Apparatus, (C2) Amyloid β -Peptides, Caseins, Calcium Channels, Diacylglycerol O-Acyltransferase, Receptor Epidermal Growth Factor, Receptor Somatotropin, Receptors Antigen T-Cell α - β , (C4) Genes MHC class II	<i>ARCNI, CACNA1A, CD3E, CD3G, CTSD, BOLA-DMA, BRD2, IGF2, ITPR3, PSMB9, PTK2</i>
		2	(C1) Golgi Apparatus, (C2) Amyloid β -Peptides, Caseins, Calcium Channels, Diacylglycerol O-Acyltransferase, Receptor Epidermal Growth Factor Receptor Somatotropin, Receptor Antigen T-Cell α - β	<i>ARCNI, CACNA1A, CD3E, CD3G, CTSD, IGF2, PTK2</i>
		3	(C2) Amyloid β -Peptides, Caseins, Calcium Channels, Diacylglycerol O-Acyltransferase, Phosphothreonine, Receptor Epidermal Growth Factor, Receptor Somatotropin	<i>CACNA1A, CTSD, IGF2, MAPT, PTK2</i>

¹AY = Ayrshire; HO = Holstein; JE = Jersey.

full 305-d traits (Strucken et al., 2011). However, in our study, no changes in the sign of SNP effects associated with candidate genes were observed between lactation stages. This may be due to the longer period defined as the first lactation stage in this study (from 5 and 95 DIM), compared with that in Strucken et al. (2011).

For the Jersey breed, the only candidate gene found consistently in more than one lactation stage and parity was *TTL* (tubulin tyrosine ligase). Similar to *SDC1* identified in Holsteins, *TTL* encodes an enzyme involved in cytoskeleton organization and function (Magiera et al., 2018), and it may be associated with MY through effects on mammary gland development and secretory capacity. In addition, many neurodevelopment diseases in humans are related to mutations in tubulin-related genes, including *TTL* (Magiera et al., 2018).

Candidate Genes for FY. A smaller number of candidate genes was found for FY in Ayrshire than in Holstein and Jersey breeds. Genomic regions associated with FY were located on BTA7 (*ICAM1* and *PIN1*), BTA10 (*HIF1A*), BTA11 (*ROCK2*), BTA13 (*BMP2*), and BTA25 (*KPNA7*) for the Ayrshire; on BTA5 (*MGP*) and BTA14 (*BOP1*, *CPSF1*, *CYHR1*, *DGAT1*, *FBXL6*, *FOXH1*, *RECQL4*, *TONSL*, and *VPS28*) for the Holstein; and BTA1 (*TMSB4*), BTA5 (*HSP90B1* and *STAB2*), BTA10 (*ANXA2* and *CYP19A1*), BTA14 (*BOP1* and *DGAT1*), BTA16 (*FASLG*), and BTA26 (*CYP26A1*, *FAS*, *IFIT3* and *IFIT5*) for the Jersey breed.

In general, *ICAM1* (intercellular adhesion molecule 1) and *PIN1* (peptidylprolyl *cis/trans* isomerase, NIMA-interacting 1) were important genes in the expression of FY for Ayrshire during all lactation periods. The *ICAM1* gene is expressed in leukocytes and used as a serum marker of inflammation (Rezamand and McGuire, 2011; Zhou et al., 2018). The *PIN1* gene is expressed in many cell types and is involved in cell signaling, especially in regard to cell proliferation (Matena et al., 2018). Different gene expression was suggested for *BMP2* (bone morphogenetic protein 2), *ROCK2* (Rho associated coiled-coil containing protein kinase 2), and *KPNA7* (karyopherin subunit α 7). In general, *BMP2*, *ROCK2*, and *KPNA7* were candidate genes for the beginning to middle, middle to end, and end of lactation, respectively. The *BMP2* gene product is an inducer of differentiation of many cell types, including those of the bone, gut, myocardium, white adipose tissue, and mammary glands (Wang et al., 2014b).

For the Holstein breed, several candidate genes seem to influence the expression of FY during all lactation periods, including *BOP1* (block of proliferation 1), *CPSF1* (cleavage and polyadenylation specific factor 1), *CYHR1* (cysteine and histidine rich 1), *DGAT1*, *FBXL6* (F-box and leucine rich repeat protein 6),

FOXH1, *LRRC24* (leucine rich repeat containing 24), *RECQL4* (RecQ like helicase 4), *TONSL* (tonsoku like, DNA repair protein), and *VPS28* (ESCRT-I subunit) genes. These genes have been reported in previous studies (Nayeri et al., 2016). The *MGP* (matrix Gla protein) gene was associated with FY during the early lactation stage and was previously reported to be associated with milk traits (Ron et al., 2007; Nayeri and Stothard, 2016).

Similarly to Holsteins, *DGAT1* and *BOP1* were identified as important genes throughout lactation for FY in Jerseys. The presence of *DGAT1* in Jerseys only for the FY trait may be related to the greater effect of the K232A *DGAT1* polymorphism on this trait compared with MY and PY, as shown by Spelman et al. (2002) in New Zealand Jersey cattle. However, specifically for Jersey animals, *ANXA2* (annexin A2) was also associated with FY during all lactation stages. The *CYP19A1* (cytochrome P450, family 19, subfamily A, polypeptide 1) and *CYP26A1* (cytochrome P450, family 26, subfamily A, polypeptide 1) genes were related to FY at the end of lactation in Jersey cattle. Similarly, Nayeri and Stothard (2016) considered *CYP19A1* to affect milk production, albeit during early lactation.

Candidate Genes for PY. Compared with the Holstein breed, a reduced number of candidate genes was found for PY in the Ayrshire and Jersey breeds. Genomic regions associated with PY were located on BTA3 (*PIK3R3*), BTA7 (*ADRB2* and *ICAM1*), BTA18 (*BAX*, *COX4I1*, and *FCGRT*), BTA23 (*CSH2*), and BTA28 (*MTR*) for the Ayrshire; BTA6 (*CSN1S1*, *CSN1S2*, and *CSN3*), BTA13 (*AKR1C4*), and BTA14 (*BOP1*, *CPSF1*, *CYHR1*, *DGAT1*, *EEF1D*, *FBXL6*, *FOXH1*, *LRRC24*, *RECQL4*, *TONSL*, and *VPS28*) for the Holstein; and BTA2 (*COL4A3*), BTA5 (*HSP90B1*), BTA18 (*COX4I1*), and BTA26 (*CYP26A1*) for the Jersey breed.

Among the genes identified, *ICAM1* appeared to play a major role in the expression of PY throughout lactation in the Ayrshire breed, whereas *ADRB2* (adrenoceptor β 2) and *CSH2* (chorionic somatomammotropin hormone 2) were especially related to the beginning of lactation. Although β -adrenergic receptors encoded by *ADRB2* are primarily recognized for their involvement in stimulating fatty acid mobilization from adipose tissue (Sumner-Thomson et al., 2011), an overlap with protein metabolism is also apparent from an association of *ADRB2* with MUN content in the Brown Swiss breed (Cecchinato et al., 2014).

Similar to MY and FY, the number of important candidate genes in PY along the entire lactation curve in the Holstein breed was large. The well-studied genes *BOP1*, *CPSF1* (cleavage and polyadenylation specific factor 1), *CYHR1* (cysteine and histidine rich 1),

DGAT1, *EEF1D*, *FBXL6* (F-box and leucine rich repeat protein 6), *FOXH1*, *LRRC24* (leucine rich repeat containing 24), *RECQL4*, *TONSL*, and *VPS28* (ESCRT-I subunit) influenced milk traits throughout the entire lactation. However, *AKR1C4* (3- α -hydroxysteroid dehydrogenase type 1) exerted effects specifically at the end of lactation. The product of *AKR1C4* is involved in bile acid synthesis and steroid hormone metabolism (Jin and Penning, 2007).

In Jersey cattle, the genes identified were *COX4I1* (cytochrome c oxidase subunit 4I1) and *HSP90B1* (heat shock protein 90 β family member 1), the latter of which has been associated with heat tolerance in cattle (Sengar et al., 2018).

Candidate Genes for SCS. Genomic regions associated with SCS were located on BTA2 (*FUCA1*) and BTA27 (*NRG1*) in Ayrshire; BTA1 (*SYNJ1*), BTA4 (*NOS3*), BTA7 (*ANGPTL4*, *MCOLN1*, and *RETN*), BTA14 (*BOP1*, *CPSF1*, *CYHR1*, *DGAT1*, *FBXL6*, *FOXH1*, *LRRC24*, *RECQL4*, *TONSL*, and *VPS28*) and BTA27 (*KCNU1*) in Holstein; and BTA7 (*CACNA1A*), BTA13 (*CTSA*, *MMP9*, *PLCB1*, and *PTI*), BTA14 (*PTK2*), BTA15 (*ARCNI*, *CD3E*, *CD3G*), BTA23 (*BOLA-DMA*, *BRD2*, *ITPR3*, *PSMB9*, *SERPINB6*, *SERPINB9*, and *TUBB2B*) and BTA29 (*CTSD* and *IGF2*) in the Jersey breed.

For the Ayrshire breed, *FUCA1* (α -L-fucosidase 1) and *NRG1* (neuregulin 1) were identified as candidate genes. The fucosidase encoded by *FUCA1* is a lysosomal protein-deglycosylating enzyme that also appears in blood plasma and on cell surfaces. Its expression level in monocytes and endothelial cells affects the ability of circulating leukocytes to penetrate into tissue at sites of inflammation (Ali et al., 2008). Expression of *FUCA1* in the bovine oviduct also influences the transport of sperm to the site of fertilization (Lefebvre et al., 1997), but a connection to SCS is not immediately apparent in this function. Neuregulin 1 is a neurotrophic factor of the epidermal growth factor family that influences behavioral responses to stress but also influences expression of immune and inflammatory genes (Marballi et al., 2010). Recently, Song et al. (2016) showed that *NRG1* methylation and expression in blood lymphocytes was strongly correlated with the progression of *Staphylococcus aureus* subclinical mastitis in cows, providing a connection to SCS.

In Holsteins, the largest number of genes associated with SCS was obtained in the second parity. Among these genes, *CPSF1* (cleavage and polyadenylation specific factor 1) and *DGAT1* were also identified by Kaupe et al. (2007) and Cochran et al. (2013). In Jerseys, *CACNA1A* (calcium voltage-gated channel, $\alpha 1$ subunit), *CTSD* (cathepsin D), *PTK2* (protein tyrosine kinase 2), and *IGF2* seemed to affect SCS throughout

the entire lactation. However, *BOLA-DMA* (major histocompatibility complex, class II, DM α), *ARCNI* (archain 1), *CD3E* (cluster of differentiation 3 ϵ), *CD3G* (cluster of differentiation 3 γ), *ITPR3* (inositol 1,4,5-trisphosphate receptor type 3), *PSMB9* (proteasome subunit β 9), and *BRD2* (bromodomain containing 2) were related to physiological processes at the beginning of lactation. Wang et al. (2013) reported that *PTK2* variants are associated with milk production traits but did not evaluate associations with SCS. The *PTK2* gene encodes a ubiquitously expressed protein tyrosine kinase (commonly called focal adhesion kinase) that is involved in cell motility responses to interactions with the extracellular matrix (Harburger and Calderwood, 2009). This kinase has been found to be important for survival, proliferation, and differentiation of mammary epithelial cells (Palmer et al., 2006) but it is also important for macrophage locomotion and inflammatory behavior (Owen et al., 2007). The *CD3E* and *CD3G* products are part of the CD3 complex that plays an important role in T-lymphocyte activation as part of cell-mediated immunity (van der Merwe and Dushek, 2011).

Comparison Between Breeds, Statistical Challenges, and Opportunities

In general, different sets of candidate genes were found for different breeds in this study. Especially for the Holstein breed, many of the candidate genes identified supported previous reports (Nayeri et al., 2016). However, well-known genes with major effects on milk production traits (e.g., *DGAT1*) had different effects between lactation stages, traits, and parities of different breeds. For instance, *DGAT1* explained a high proportion of variation in the Holstein traits and in FY in Jersey but we found no evidence of a substantial effect of *DGAT1* on milk production traits and SCS of the Ayrshire breed, even though the gene was segregating in all 3 breeds. This may explain the lower proportion of genetic variance explained by important windows in these breeds for these traits compared with Holsteins (that had *DGAT1* as candidate gene for the genomic windows that explained the highest proportion of genetic variance for the analyzed traits). However, in contrast to our results, Spelman et al. (2002) suggested that mutation in *DGAT1* has a great effect on milk production traits in all 3 breeds.

Several factors may affect the discrepancies between the sets of candidate genes identified in our study and previous studies reported in the literature. Among these factors is the fact that we used “partial accumulated yield” to represent traits at different lactation stages, whereas most studies in the literature used 305-d ac-

cumulated yield to identify genes (Spelman et al., 2002; Raven et al., 2014; Nayeri et al., 2016). This difference in phenotypes used in the GWAS probably influenced the set of candidate genes identified due to the non-constant effects of QTL over time for those traits, as suggested by the changes in additive genetic variance and heritability over the course of lactation (Miglior et al., 2009; Oliveira et al., 2017; Sasaki et al., 2017). Thus, our study allowed the identification of specific genes or genetic loci that were more pronounced during a specific lactation stage, and not necessarily with a high average QTL effect throughout lactation.

The discrepancies observed between the sets of candidate genes identified in our study and previous studies could be attributed to the statistical methods used. In our study, SNP effects were estimated by back-solving DGV instead of using a standard single-SNP regression (e.g., Buitenhuis et al., 2014; Nayeri et al., 2016). Thus, the method used in our study simultaneously estimated the effect of all SNP, which accounts for the LD among SNP. On the other hand, our findings rely on the use of arbitrary thresholds, such as the number of SNPs to create the genomic windows and the threshold used to identify the important windows. Thus, thresholds defined in this study (i.e., 20 consecutive SNP and 0.30% of total genetic variance) may affect the results. For instance, if we had considered fewer SNP per genomic window, we would likely have identified more windows explaining less of the total genetic variance (due to the level of LD found in the breeds studied here). However, if we had considered the same number of SNP per window and a higher proportion of the variance explained to identify the important windows, we would have identified fewer genomic windows, especially for the Ayrshire and Jersey breeds. To overcome these limitations and, therefore, to reduce the number of false-positive candidate genes, we performed a functional prioritization analysis after the GWAS analyses.

To reduce the influence of the thresholds used in this study on our findings, the same thresholds were used for all analyses. However, we observed substantial differences between the sets of candidate genes identified for each breed. Among the factors that likely contributed to the differences between sets of candidate genes among breeds are the differences in reliabilities of pseudo-phenotypes used, genetic backgrounds, and sample sizes.

In this study, we estimated SNP effects by back-solving the DGV estimated from the use of deregressed EBV as a response variable in the GBLUP. The use of deregressed EBV is important to avoid the double shrinkage of breeding values and the double counting of information in the analysis (Garrick et al., 2009). In summary, the deregressed EBV used here were calcu-

lated as functions of EBV and reliability of EBV for each DIM of each animal, in a way that animals with lower reliability of EBV had deregressed EBV farthest from the original EBV. Comparing the 3 breeds studied here, the increase in deregressed EBV variability was higher for the Ayrshire and Jersey breeds, because the threshold of minimum reliability of EBV to include animals in the analyses was lower for those breeds (the threshold was equal to 0.65 for Holstein and 0.50 for Ayrshire and Jersey breeds; Table 1). Thus, the Ayrshire and Jersey breeds had less accurate response variables than the Holstein breed, which may influence the detection of QTL (Beecher et al., 2010). A lower threshold was assigned for Ayrshire and Jersey than for Holsteins because of the smaller number of genotyped animals available.

Another fact that may influence the set of candidate genes is the different genetic backgrounds of the breeds, as these breeds have been genetically separated for many generations (Kantanen, 2000) and have undergone strong artificial selection (Buitenhuis et al., 2016). Thus, the different selection processes to which those breeds were subjected may have fixed different important alleles in each population and it may reflect in different LD structures between breeds, as shown in this study. In addition, differences in heritabilities could explain the difference in the number of QTL detected for each breed (Buitenhuis et al., 2016).

Differences in sample sizes; that is, differences in the number of genotyped animals, can also help to explain the difference between candidate genes identified for the different breeds in this study. The number of genotyped animals is related to the statistical power, and it is a major issue in any GWAS (Meredith et al., 2013). In our study, the power of detecting QTL was lower in the Ayrshire and Jersey breeds than in the Holstein breed; however, more regions were identified as important in Ayrshire and Jersey breeds. As suggested before, this higher number of detected windows in the Ayrshire and Jersey breeds is likely due to artifacts from a smaller sample size. For instance, a smaller sample size might be associated with a larger number of animals carrying variants with low MAF, which will consequently have a larger effect on the estimated SNP allele substitution effects (and number of important genomic regions identified). These artifacts may be most influential for the SCS trait in the Jersey breed, where the number of genotyped animals was small and the number of candidate genes was large. However, the number of genotyped animals is not under control of researchers and, in many instances, it cannot be increased. In this case, further studies using alternative methods and populations should be used to validate previous results.

Alternative methodologies such as the use of a meta-analysis to combine information from independent studies (Evangelou and Ioannidis, 2013), the use of single-step GWAS considering different weights for different SNP (Wang et al., 2014a), or the joint consideration of a set of correlated traits (Pei et al., 2009) can be used in posterior studies to validate the findings from this study. The reason we adopted a single-trait analyses herein is that the model used by CDN to estimate the EBV was already a multiple-trait model (Interbull, 2018); that is, the EBV already contained information about the other traits. In addition, we opted for a multiple-step method (i.e., for estimating and deregressing the EBV to be used as a pseudo-phenotype to estimate the DGV) instead of the single-step method because the single-step method is not currently used for official genomic evaluations in Canada (Interbull, 2017).

CONCLUSIONS

Genomic regions associated with MY, FY, PY, and SCS in the first 3 parities were identified for different lactation stages in Ayrshire, Holstein, and Jersey breeds. Despite the small sample sizes available in the Ayrshire and Jersey breeds, our results suggest evidence of differential sets of candidate genes underlying the phenotypic expression of the analyzed traits analyzed across breeds, parities, and lactation stages. Further studies are needed to validate our findings in independent populations.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support for this research from the Canada First Research Excellence Fund (Ottawa, ON, Canada), Agriculture and (Guelph, ON, Canada) and additional contributions from the Dairy Farmers of Canada (Ottawa, ON, Canada), the Canadian Dairy Network (Guelph, ON, Canada), and the Canadian Dairy Commission (Ottawa, ON, Canada) under the Agri-Science Clusters Initiative. The first author acknowledges the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; Brasília, DF, Brazil).

REFERENCES

- Adams, W. T., and T. R. Skopek. 1987. Statistical test for the comparison of samples from mutational spectra. *J. Mol. Biol.* 194:391–396. [https://doi.org/10.1016/0022-2836\(87\)90669-3](https://doi.org/10.1016/0022-2836(87)90669-3).
- Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Pages 1–3 in Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, BC, Canada. Am. Soc. Anim. Sci., Champaign, IL. <https://doi.org/10.13140/2.1.4801.5045>.
- Akers, R. M. 2017. A 100-Year Review: Mammary development and lactation. *J. Dairy Sci.* 100:10332–10352. <https://doi.org/10.3168/jds.2017-12983>.
- Ali, S., Y. Jenkins, M. Kirkley, A. Dagkalis, A. Manivannan, I. J. Crane, and J. A. Kirby. 2008. Leukocyte extravasation: A role for α -1-fucosidase? *J. Immunol.* 181:2407–2413. <https://doi.org/10.4049/jimmunol.181.4.2407>.
- Almeida, A. M., and P. D. Eckersall. 2018. Proteomics and mammary gland research in dairy species. Pages 255–280 in *Proteomics in Domestic Animals: from Farm to Systems Biology*. Springer International Publishing, New York, NY.
- Bauman, D. E., I. H. Mather, R. J. Wall, and A. L. Lock. 2006. Major advances associated with the biosynthesis of milk. *J. Dairy Sci.* 89:1235–1243. [https://doi.org/10.3168/jds.S0022-0302\(06\)72192-0](https://doi.org/10.3168/jds.S0022-0302(06)72192-0).
- Baumgard, L. H., R. J. Collier, and D. E. Bauman. 2017. A 100-Year Review: Regulation of nutrient partitioning to support lactation. *J. Dairy Sci.* 100:10353–10366. <https://doi.org/10.3168/jds.2017-13242>.
- Beecher, C., M. Daly, S. Childs, D. P. Berry, D. A. Magee, T. V. McCarthy, and L. Giblin. 2010. Polymorphisms in bovine immune genes and their associations with somatic cell count and milk production in dairy cattle. *BMC Genet.* 11:99. <https://doi.org/10.1186/1471-2156-11-99>.
- Bionaz, M., and J. J. Loo. 2008a. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. *J. Nutr.* 138:1019–1024. <https://doi.org/10.1093/jn/138.6.1019>.
- Bionaz, M., and J. J. Loo. 2008b. Gene networks driving bovine milk fat synthesis during the lactation cycle. *BMC Genomics* 9:366. <https://doi.org/10.1186/1471-2164-9-366>.
- Brito, L. F., F. G. Silva, H. R. Oliveira, N. Souza, G. Caetano, E. V. Costa, G. R. O. Menezes, A. L. P. Melo, M. T. Rodrigues, and R. A. Torres. 2017. Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. *Can. J. Anim. Sci.* 98:73–83. <https://doi.org/10.1139/CJAS-2017-0019>.
- Buitenhuis, B., L. L. Janss, N. A. Poulsen, L. B. Larsen, M. K. Larsen, and P. Sørensen. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. *BMC Genomics* 15:1112. <https://doi.org/10.1186/1471-2164-15-1112>.
- Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen. 2016. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post-translational modifications in Danish Holstein and Danish Jersey cattle. *BMC Genet.* 17:114. <https://doi.org/10.1186/s12863-016-0421-2>.
- Cecchinato, A., C. Ribeca, S. Chessa, C. Cipolat-Gotet, F. Maretto, J. Casellas, and G. Bittante. 2014. Candidate gene association analysis for milk yield, composition, urea nitrogen and somatic cell scores in Brown Swiss cows. *Animal* 8:1062–1070. <https://doi.org/10.1017/S1751731114001098>.
- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4:7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Cochran, S. D., J. B. Cole, D. J. Null, and P. J. Hansen. 2013. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC Genet.* 14:49. <https://doi.org/10.1186/1471-2156-14-49>.
- Evangelou, E., and J. P. A. Ioannidis. 2013. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* 14:379–389. <https://doi.org/10.1038/nrg3472>.
- Fragomeni, B. O., I. Misztal, D. L. Lourenco, I. Aguilar, R. Okimoto, and W. M. Muir. 2014. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. *Front. Genet.* 5:332. <https://doi.org/10.3389/fgene.2014.00332>.
- 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>.
- Hambruch, N., S. Kumstel, J. D. Haeger, and C. Pfarrer. 2017. Bovine placental heparanase and syndecan expression is related to placental maturation. *Placenta* 57:42–51. <https://doi.org/10.1016/j.placenta.2017.06.006>.
- Harburger, D. S., and D. A. Calderwood. 2009. Integrin signalling at a glance. *J. Cell Sci.* 122:159–163. <https://doi.org/10.1242/jcs.018093>.
- Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38:226–231. <https://doi.org/10.1007/BF01245622>.
- Interbull. 2017. Description of national genomic evaluation systems. Canada. Accessed May 2, 2018. <http://www.interbull.org/ib/nationalgenoforms>.
- Interbull. 2018. Description of national genetic evaluation systems. Accessed May 2, 2018. <http://www.interbull.org/ib/geforms>.
- Jamrozik, J., L. R. Schaeffer, and G. B. Jansen. 2000. Approximate accuracies of prediction from random regression models. *Livest. Prod. Sci.* 66:85–92. [https://doi.org/10.1016/S0301-6226\(00\)00158-5](https://doi.org/10.1016/S0301-6226(00)00158-5).
- Jin, Y., and T. M. Penning. 2007. Aldo-keto reductases and bioactivation/detoxication. *Annu. Rev. Pharmacol. Toxicol.* 47:263–292. <https://doi.org/10.1146/annurev.pharmtox.47.120505.105337>.
- Kantanen, J. 2000. Genetic diversity and population structure of 20 north European cattle breeds. *J. Hered.* 91:446–457. <https://doi.org/10.1093/jhered/91.6.446>.
- Kaupe, B., H. Brandt, E. M. Prinzenberg, and G. Erhardt. 2007. Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. *J. Anim. Sci.* 85:11–21. <https://doi.org/10.2527/jas.2005-753>.
- Kinsella, R. J., A. Kahari, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek. 2011. Ensembl BioMart: A hub for data retrieval across taxonomic space. *Database (Oxford)* 2011:bar030. <https://doi.org/10.1093/database/bar030>.
- Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124:979–993.
- Larmer, S. G., M. Sargolzaei, and F. Schenkel. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. *J. Dairy Sci.* 97:3128–3141. <https://doi.org/10.3168/jds.2013-6826>.
- Lefebvre, R., M. Lo, and S. Suarez. 1997. Bovine sperm binding to oviductal epithelium involves fucose recognition. *Biol. Reprod.* 56:1198–1204.
- Magiera, M. M., P. Singh, S. Gadadhar, and C. Janke. 2018. Tubulin posttranslational modifications and emerging links to human disease. *Cell* 173:1323–1327. <https://doi.org/10.1016/j.cell.2018.05.018>.
- Marballi, K., M. P. Quinones, F. Jimenez, M. A. Escamilla, H. Raventos, M. C. Soto-Bernardini, S. S. Ahuja, and C. Walss-Bass. 2010. In vivo and in vitro genetic evidence of involvement of neuregulin 1 in immune system dysregulation. *J. Mol. Med. (Berl.)* 88:1133–1141. <https://doi.org/10.1007/s00109-010-0653-y>.
- Matena, A., E. Rehic, D. Hönig, B. Kamba, and P. Bayer. 2018. Structure and function of the human parvulins Pin1 and Par14/17. *Biol. Chem.* 399:101–125. <https://doi.org/10.1515/hsz-2017-0137>.
- Meredith, B. K., D. P. Berry, F. Kearney, E. K. Finlay, A. G. Fahey, D. G. Bradley, and D. J. Lynn. 2013. A genome-wide association study for somatic cell score using the Illumina high-density bovine beadchip identifies several novel QTL potentially related to mastitis susceptibility. *Front. Genet.* 4:229. <https://doi.org/10.3389/fgene.2013.00229>.
- Miglior, F., W. Gong, Y. Wang, G. J. Kistemaker, A. Sewalem, and J. Jamrozik. 2009. Short communication: Genetic parameters of production traits in Chinese Holsteins using a random regression test-day model. *J. Dairy Sci.* 92:4697–4706. <https://doi.org/10.3168/jds.2009-2212>.
- Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. Lee. 2002. BLUPF90 and related programs (BGF90). Pages 21–22 in *Proc. 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France*. Editions Quae, Montpellier, France.
- Morota, G., T. M. Beissinger, and F. Peñagaricano. 2016. MeSH-informed enrichment analysis and MeSH-guided semantic similarity among functional terms and gene products in chicken. *G3 (Bethesda)* 6:2447–2453. <https://doi.org/10.1534/g3.116.031096>.
- Morota, G., F. Peñagaricano, J. L. Petersen, D. C. Ciobanu, K. Tsuyuzaki, and I. Nikaido. 2015. An application of MeSH enrichment analysis in livestock. *Anim. Genet.* 46:381–387. <https://doi.org/10.1111/age.12307>.
- Nayeri, S., M. Sargolzaei, M. K. Abo-Ismael, N. May, S. P. Miller, F. S. Schenkel, S. Moore, and P. Stothard. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* 17:75. <https://doi.org/10.1186/s12863-016-0386-1>.
- Nayeri, S., and P. Stothard. 2016. Tissues, metabolic pathways and genes of key importance in lactating dairy cattle. *Springer Sci. Rev.* 4:49–77. <https://doi.org/10.1007/s40362-016-0040-3>.
- Nelson, S. J., M. Schopen, A. G. Savage, J. L. Schulman, and N. Arluk. 2004. The MeSH translation maintenance system: Structure, interface design, and implementation. *Stud. Health Technol. Inform.* 107:67–69.
- Nyman, A. K., T. Ekman, U. Emanuelson, A. H. Gustafsson, K. Holtenius, K. P. Waller, and C. H. Sandgren. 2007. Risk factors associated with the incidence of veterinary-treated clinical mastitis in Swedish dairy herds with a high milk yield and a low prevalence of subclinical mastitis. *Prev. Vet. Med.* 78:142–160. <https://doi.org/10.1016/j.prevetmed.2006.10.002>.
- Oliveira, C. S. F., H. Hogeveen, A. M. Botelho, P. V. Maia, S. G. Coelho, and J. P. A. Haddad. 2015. Cow-specific risk factors for clinical mastitis in Brazilian dairy cattle. *Prev. Vet. Med.* 121:297–305. <https://doi.org/10.1016/j.prevetmed.2015.08.001>.
- Oliveira, H. R., L. F. Brito, F. F. Silva, D. A. L. Lourenco, J. Jamrozik, and F. S. Schenkel. 2019. Genomic prediction of lactation curves for milk, fat, protein and somatic cell score in Holstein cattle. *J. Dairy Sci.* 102:452–463. <https://doi.org/10.3168/jds.2018-15159>.
- Oliveira, H. R., F. F. Silva, L. F. Brito, A. R. Guarini, J. Jamrozik, and F. S. Schenkel. 2018. Comparing deregression methods for genomic prediction of test-day traits in dairy cattle. *J. Anim. Breed. Genet.* 135:97–106. <https://doi.org/10.1111/jbg.12317>.
- Oliveira, H. R., F. F. Silva, M. V. G. B. Silva, O. H. G. B. D. Siqueira, M. A. Machado, J. C. C. Panetto, L. S. Glória, and L. F. Brito. 2017. Bayesian Models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. *Livest. Sci.* 201:78–84. <https://doi.org/10.1016/j.livsci.2017.05.007>.
- Owen, K. A., F. J. Pixley, K. S. Thomas, M. Vicente-Manzanares, B. J. Ray, A. F. Horwitz, J. T. Parsons, H. E. Beggs, E. R. Stanley, and A. H. Bouton. 2007. Regulation of lamellipodial persistence, adhesion turnover, and motility in macrophages by focal adhesion kinase. *J. Cell Biol.* 179:1275–1287.
- Palmer, C. A., M. C. Neville, S. M. Anderson, and J. L. McManaman. 2006. Analysis of lactation defects in transgenic mice. *J. Mammary Gland Biol. Neoplasia* 11:269–282. <https://doi.org/10.1007/s10911-006-9023-3>.
- Pei, Y. F., L. Zhang, J. Liu, and H. W. Deng. 2009. Multivariate association test using haplotype trend regression. *Ann. Hum. Genet.* 73:456–464. <https://doi.org/10.1111/j.1469-1809.2009.00527.x>.
- Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, and B. J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J. Dairy Sci.* 93:3331–3345. <https://doi.org/10.3168/jds.2009-2893>.
- R Core Team. 2011. R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing*, editor. R Core Team, Vienna, Austria.
- Raven, L. A., B. G. Cocks, and B. J. Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative vari-

- ants underlying milk production in dairy cattle. *BMC Genomics* 15:62. <https://doi.org/10.1186/1471-2164-15-62>.
- Rezamand, P., and M. A. McGuire. 2011. Short communication: Effects of trans fatty acids on markers of inflammation in bovine mammary epithelial cells. *J. Dairy Sci.* 94:316–320. <https://doi.org/10.3168/jds.2010-3289>.
- Ricklin, D., G. Hajishengallis, K. Yang, and J. D. Lambris. 2010. Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* 11:785–797. <https://doi.org/10.1038/ni.1923>.
- Ron, M., G. Israeli, E. Seroussi, J. I. Weller, J. P. Gregg, M. Shani, and J. F. Medrano. 2007. Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. *BMC Genomics* 8:183. <https://doi.org/10.1186/1471-2164-8-183>.
- Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2011. FImpute—An efficient imputation algorithm for dairy cattle populations. *J. Dairy Sci.* 94(E-Suppl. 1):421. (Abstr.)
- Sasaki, O., M. Aihara, A. Nishiura, and H. Takeda. 2017. Genetic correlations between the cumulative pseudo-survival rate, milk yield, and somatic cell score during lactation in Holstein cattle in Japan using a random regression model. *J. Dairy Sci.* 100:7282–7294. <https://doi.org/10.3168/jds.2016-12311>.
- Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and B. J. Van Doormaal. 2000. Experience with a test-day model. *J. Dairy Sci.* 83:1135–1144. [https://doi.org/10.3168/jds.S0022-0302\(00\)74979-4](https://doi.org/10.3168/jds.S0022-0302(00)74979-4).
- Sengar, G. S., R. Deb, U. Singh, T. V. Raja, R. Kant, B. Sajjanar, R. Alex, R. R. Alyethodi, A. Kumar, S. Kumar, R. Singh, S. J. Jakhesara, and C. Joshi. 2018. Differential expression of microRNAs associated with thermal stress in Frieswal (*Bos taurus* x *Bos indicus*) crossbred dairy cattle. *Cell Stress Chaperones* 23:155–170. <https://doi.org/10.1007/s12192-017-0833-6>.
- Silva, R. M. O., N. B. Stafuzza, B. O. Fragomeni, G. M. F. Camargo, T. M. Ceacero, J. N. S. G. Cyrillo, F. Baldi, A. A. Boligon, M. E. Z. Mercadante, D. L. Lourenco, I. Misztal, and L. G. Albuquerque. 2017. Genome-wide association study for carcass traits in an experimental Nelore cattle population. *PLoS One* 12:e0169860. <https://doi.org/10.1371/journal.pone.0169860>.
- Soares, A. C. C., S. E. F. Guimarães, M. J. Kelly, M. R. S. Fortes, F. F. Silva, L. L. Verardo, R. Mota, and S. Moore. 2017. Multiple-trait genomewide mapping and gene network analysis for scrotal circumference growth curves in Brahman cattle. *J. Anim. Sci.* 95:3331–3345. <https://doi.org/10.2527/jas.2017.1409>.
- Song, M., Y. He, H. Zhou, Y. Zhang, X. Li, and Y. Yu. 2016. Combined analysis of DNA methylome and transcriptome reveal novel candidate genes with susceptibility to bovine *Staphylococcus aureus* subclinical mastitis. *Sci. Rep.* 6:29390. <https://doi.org/10.1038/srep29390>.
- Spelman, R. J., C. A. Ford, P. McElhinney, G. C. Gregory, and R. G. Snell. 2002. Characterization of the DGAT1 gene in the New Zealand dairy population. *J. Dairy Sci.* 85:3514–3517. [https://doi.org/10.3168/jds.S0022-0302\(02\)74440-8](https://doi.org/10.3168/jds.S0022-0302(02)74440-8).
- Stachowicz, K., M. Sargolzaei, F. Miglior, and F. S. Schenkel. 2011. Rates of inbreeding and genetic diversity in Canadian Holstein and Jersey cattle. *J. Dairy Sci.* 94:5160–5175. <https://doi.org/10.3168/jds.2010-3308>.
- Strabel, T., and I. Misztal. 1999. Genetic parameters for first and second lactation milk yields of Polish Black and White cattle with random regression test-day models. *J. Dairy Sci.* 82:2805–2810. [https://doi.org/10.3168/jds.S0022-0302\(99\)75538-4](https://doi.org/10.3168/jds.S0022-0302(99)75538-4).
- Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of random regression test-day models for Polish Black and White cattle. *J. Dairy Sci.* 88:3688–3699. [https://doi.org/10.3168/jds.S0022-0302\(05\)73055-1](https://doi.org/10.3168/jds.S0022-0302(05)73055-1).
- Strucken, E. M., R. H. Bortfeldt, D. J. Koning, and G. A. Brockmann. 2012a. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. *Anim. Genet.* 43:375–382. <https://doi.org/10.1111/j.1365-2052.2011.02278.x>.
- Strucken, E. M., R. H. Bortfeldt, J. Tetens, G. Thaller, and G. A. Brockmann. 2012b. Genetic effects and correlations between production and fertility traits and their dependency on the lactation-stage in Holstein Friesians. *BMC Genet.* 13:108. <https://doi.org/10.1186/1471-2156-13-108>.
- Strucken, E. M., D. J. Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. *J. Dairy Sci.* 94:442–449. <https://doi.org/10.3168/jds.2009-2932>.
- Summer-Thomson, J. M., J. L. Vierk, and J. P. McNamara. 2011. Differential expression of genes in adipose tissue of first-lactation dairy cattle. *J. Dairy Sci.* 94:361–369. <https://doi.org/10.3168/jds.2010-3447>.
- Tsuyuzaki, K., G. Morota, M. Ishii, T. Nakazato, S. Miyazaki, and I. Nikaido. 2015. MeSH ORA framework: R/Bioconductor packages to support MeSH over-representation analysis. *BMC Bioinformatics* 16:45. <https://doi.org/10.1186/s12859-015-0453-z>.
- van der Merwe, P. A., and O. Dushek. 2011. Mechanisms for T cell receptor triggering. *Nat. Rev. Immunol.* 11:47–55. <https://doi.org/10.1038/nri2887>.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggins, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24. <https://doi.org/10.3168/jds.2008-1514>.
- Wang, H., L. Jiang, X. Liu, J. Yang, J. Wei, J. Xu, Q. Zhang, and J. F. Liu. 2013. A post-GWAS replication study confirming the PTK2 gene associated with milk production traits in Chinese Holstein. *PLoS One* 8:e83625. <https://doi.org/10.1371/journal.pone.0083625>.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W. M. Muir. 2014a. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* 5:134. <https://doi.org/10.3389/fgene.2014.00134>.
- Wang, R. N., J. Green, Z. Wang, Y. Deng, M. Qiao, M. Peabody, Q. Zhang, J. Ye, Z. Yan, S. Denduluri, O. Idowu, M. Li, C. Shen, A. Hu, R. C. Haydon, R. Kang, J. Mok, M. J. Lee, and L. L. Shi. 2014b. Bone morphogenetic protein (BMP) signaling in development and human diseases. *Genes Dis.* 1:87–105.
- Wiggins, 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>.
- Zhang, Z., J. Liu, X. Ding, P. Bijma, D. J. de Koning, and Q. Zhang. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS One* 5:e12648. <https://doi.org/10.1371/journal.pone.0012648>.
- Zhou, Z., F. Ferdous, P. Montagner, D. N. Luchini, M. N. Corrêa, and J. J. Loo. 2018. Methionine and choline supply during the periparturient period alter polymorphonuclear leukocyte immune response and immunometabolic gene expression in Holstein cows. *J. Dairy Sci.* 101:10374–10382. <https://doi.org/10.3168/jds.2018-14972>.