Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows

Received: 25 July 2019 | Revised: 7 September 2019 | Accepted: 12 September 2019

DOI: 10.1111/jbg.12442

ORIGINAL ARTICLE

Journal of Animal Breeding and Genetics WILEY

Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows

Hadi Atashi[1,2] | **Mazdak Salavati**[3] | **Jenne De Koster**[1] | **Jim Ehrlich**[4] | **Mark Crowe**[5] | **Geert Opsomer**[1] | **the GplusE consortium*** | **Miel Hostens**[1]

[1]Department of Reproduction, Obstetrics and Herd Health, Ghent University, Merelbeke, Belgium

[2]Department of Animal Science, Shiraz University, Shiraz, Iran

[3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK

[4]Dairy Veterinarians Group, Argyle, New York

[5]University College Dublin, Dublin, Ireland

Correspondence
Miel Hostens, Department of Reproduction, Obstetrics and Herd Health, Ghent University, Salisburylaan 133, 9820 Merelbeke, Belgium.
Email: miel.hostens@ugent.be

Abstract

The aim of this study was to identify genomic regions associated with 305-day milk yield and lactation curve parameters on primiparous (n = 9,910) and multiparous (n = 11,158) Holstein cows. The SNP solutions were estimated using a weighted single-step genomic BLUP approach and imputed high-density panel (777k) genotypes. The proportion of genetic variance explained by windows of 50 consecutive SNP (with an average of 165 Kb) was calculated, and regions that accounted for more than 0.50% of the variance were used to search for candidate genes. Estimated heritabilities were 0.37, 0.34, 0.17, 0.12, 0.30 and 0.19, respectively, for 305-day milk yield, peak yield, peak time, ramp, scale and decay for primiparous cows. Genetic correlations of 305-day milk yield with peak yield, peak time, ramp, scale and decay in primiparous cows were 0.99, 0.63, 0.20, 0.97 and −0.52, respectively. The results identified three windows on BTA14 associated with 305-day milk yield and the parameters of lactation curve in primi- and multiparous cows. Previously proposed candidate genes for milk yield supported by this work include *GRINA, CYHR1, FOXH1, TONSL, PPP1R16A, ARHGAP39, MAF1, OPLAH and MROH1*, whereas newly identified candidate genes are *MIR2308, ZNF7, ZNF34, SLURP1, MAFA* and *KIFC2* (BTA14). The protein lipidation biological process term, which plays a key role in controlling protein localization and function, was identified as the most important term enriched by the identified genes.

KEYWORDS

genome-wide association study, Holstein, lactation curve, milk yield

1 | INTRODUCTION

Milk production is mainly dependent on the shape of the lactation curve, defined as the graphical representation of milk yield over the course of the lactating period (Do et al., 2017; Ehrlich, 2011; El-Awady, 2013). The shape of the lactation curve is characterized by the slope of the initial rise of the curve, peak yield, time to peak, the slope of the curve after peak yield (lactation persistency) and lactation length (El-Awady, 2013; López et al., 2015; Rekik, Gara, Hamouda, & Hammami, 2003). Although all these characteristics are responsible for total milk yield

*GplusE consortium members are listed in Acknowledgements

per lactation, peak yield and persistency are considered as among the economically most important production traits in dairy cows (Dekkers, Ten Hag, & Weersink, 1998; Do et al., 2017; Tekerli, Akinci, Dogan, & Akcan, 2000). There is a considerable variation between animals in terms of the shape of the lactation curve, which can be attributed to factors such as genetic background, parity, diet, health status and other environmental factors (Atashi, Zamiri, & Dadpasand, 2013; Hostens, Ehrlich, Van Ranst, & Opsomer, 2012; Rekaya, Carabano, & Toro, 2000; Tekerli et al., 2000). Although several linear and non-linear functions with different functional forms have been used to describe the relationship between daily milk yield and days in milk (DIM) in dairy cows (Sherchand, McNew, Kellogg, & Johnson, 1995; Silvestre, Petim-Batista, & Colaco, 2006), none has achieved widespread acceptance outside a few specialized applications which are directed primarily at improving estimates of actual production from incomplete data sets (Ehrlich, 2011). The MilkBot model proposed by Ehrlich (2011) is a non-linear lactation curve model in which parameter values can be interpreted by the effect that they have on the lactation curve. The MilkBot model is flexible enough to accommodate disease and management effect, and can provide more accurate estimates of dairy milk yield (Cole, Ehrlich, & Null, 2012).

Although genome-wide association studies (GWAS) carried out within a variety of cattle breeds identified many genomic regions explaining variation in milk yield, they are mainly based on the polygenic estimated breeding value (EBV), daughter yield deviation (DYD) or deregressed proof for accumulated 305-day milk yield (Cole et al., 2011; Iso-Touru, Sahana, Guldbrandtsen, Lund, & Vilkki, 2016; Jiang et al., 2010; Meredith et al., 2012; Nayeri et al., 2016). The accumulated 305-day milk yield is estimated by summing the test-day milk yield (TDMY) recorded every day during the lactation period or combining the weekly or monthly TDMY by linear interpolation (Schaeffer, 2016). Since the additive genetic variance for milk yield changes during lactation (Bignardi, El Faro, Cardoso, Machado, & de Albuquerque, 2009; Singh et al., 2016), the genetic effects of QTL related to 305-day milk yield are not constant during the lactation period; therefore, many QTL whose genetic effects change during lactation might not be detected in this approach (Lund, Sorensen, Madsen, & Jaffrézic, 2008; Ning et al., 2018). Strucken, Bortfeldt, De Koning, and Brockmann (2012) showed that lactation curve parameters provide a higher power to screen the whole genome for region whose effect change during lactation. Therefore, the objective of this study was to identify genomic region(s) associated with 305-day milk yield and the parameters of the MilkBot lactation curve model in Holstein dairy cows.

2 | MATERIALS AND METHODS

2.1 | Phenotypic data

Data in this study were collected as part of Work Package 4 from the Genotype plus Environment (GplusE) FP7-Project (http://www.gpluse.eu). The data were records of 21,068 lactations on primiparous (9,910) and multiparous (11,158) Holstein cows calving between 2010 and 2018, distributed among 118 herds in four countries (Belgium, The Netherlands, Great Britain and Denmark). To describe the lactation curve, the MilkBot model developed by Ehrlich (2011) was fitted on milk recording events from each animal. The MilkBot function is as follows:

$$y_t = a\left(1 - \frac{\exp\left(\frac{c-t}{b}\right)}{2}\right)\exp(-dt).$$

In which, a is the scale parameter, representing the theoretical maximum daily yield; b is the ramp parameter, controlling the rate of rise in milk production in early lactation; c is the offset parameter, describing the offset in time between parturition and the start of lactation; and d is the decay parameter, representing the rate of senescence of production capacity. The time at which peak lactation occurred (t_{peak}) was defined as: $t_{peak} = -b\ln\left(\frac{2db}{db+1}\right)+c$, and peak yield was calculated by substitution t_{peak} in the MilkBot equation. The 305-day milk, the cumulative milk yield between calving and day 305 of the lactation, was calculated as:

$$M_{305} = (a - a\exp(-305\,d))/d + (ab\exp(c/b)(-1+\exp(-305(1/b+d))))/(2+2bd).$$

2.2 | Genotypic data

Individuals ($n = 31{,}895$) were genotyped using the BovineLD ($n = 20{,}462$), BovineSNP50K ($n = 10{,}638$) or BovineHD SNP panel (795 animals). Genotypes of animals were imputed to HD with a reference population of 795 (46 M and 749 F) HD individuals using FImpute V2.2 software (Sargolzaei, Chesnais, & Schenkel, 2014). In total, 12,367 out of 31,895 genotyped individuals had either phenotypic data or were in the pedigree file which was used in the association analysis (the number of animals with records was 9,910, the number of animals with records and with genotypes was 8,172, the number of animals with records and no genotypes was 1,738, and the number of animals with genotypes and no records was 4,195). Quality control (QC) was performed on the imputed data. SNP markers with minor allele frequency (MAF) less than 5% were excluded. After genomic data QC, 566,345 out of 730,539 SNP were available for the association analysis.

2.3 | Variance components estimation

Pedigree information was collected for all phenotyped animals and contained a total of 43,181 individuals (12,367 and 9,910 out of 43,181 animals had genotype and phenotype data, respectively). The genetic analyses were carried out through the Average Information Restricted Maximum Likelihood (AIREML) method, using a linear single-trait animal model (for measurements on the primiparous cows). The linear model included fixed effect of country and herd-year-season of calving, covariate effects of age at first calving in both linear and quadratic forms, and animal and residual random effects. The complete model can be represented as follows:

$$y_{ijk} = \mu + \mathrm{HYS}_i + \mathrm{con}_j + b_1(\mathrm{age}_k) + b_2(\mathrm{age}_k)^2 + a_k + e_{ijk}$$

where y_{ijk} represents the response variable of animal k, μ is the overall mean, HYS_i is the fixed effect of ith herd-year-season of calving, con_j is the fixed effect of jth country, b_1 and b_2 are the linear and quadratic regression coefficients of the dependent variable on the age at first calving, age_k is the age at first calving of kth cows, a_k is the additive genetic effect, and e_{ijk} is the random residual error. The additive genetyic and residual variances were obtained as follows:

$$\mathrm{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{H}\sigma_a^2 & 0\\ 0 & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where σ_a^2 and σ_e^2 are, respectively, total additive genetic and residual variances, $\mathbf{a}$ is the vector of direct additive genetic effects, $\mathbf{e}$ is a vector of residual effects, and $\mathbf{H}$ is a matrix that combines pedigree and genomic relationships, and its inverse consists on the integration of additive and genomic relationship matrices, $\mathbf{A}$ and $\mathbf{G}$, respectively (Aguilar et al., 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix}0 & 0\\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\end{bmatrix}$$

where $\mathbf{A}$ is the numerator relationship matrix based on pedigree for all animals; $\mathbf{A}_{22}$ is the numerator relationship matrix for genotyped animals; and $\mathbf{G}$ is the genomic relationship matrix which was obtained using following function described by VanRaden et al. (2009).

$$\mathbf{G} = \frac{\mathbf{ZDZ}'}{\sum_{i=1}^{M} 2p_i(1-p_i)}$$

where $\mathbf{Z}$ is a matrix of gene content adjusted for allele frequencies (0, 1 or 2 for *aa*, *Aa* and *AA*, respectively); $\mathbf{D}$ is a diagonal matrix of weights for SNP variances (initially $\mathbf{D} = \mathbf{I}$); M is the number of SNP, and p_i is the MAF of ith SNP. The $\mathbf{H}$ matrix was built scaling $\mathbf{G}$ based on $\mathbf{A_{22}}$ considering that the average of the diagonal of $\mathbf{G}$ is equal to the average of the diagonal of $\mathbf{A_{22}}$ and, the average of off-diagonal $\mathbf{G}$ is equal to average of off-diagonal $\mathbf{A_{22}}$. Analyses were performed using AIREMLF90 (Misztal et al., 2002). The genetic analyses for the measurements on the multiparous cows were carried out using a linear single-trait repeatability animal model, which was the same as the model used for primiparous cows but here, the fixed effect of parity was included in the model. In addition, a third random effect representing the permanent effect associated with animals having repeated records was included in the model. This effect, assumed to be uncorrelated with additive genetic effects, allowed for the partitioning of the environmental variance into permanent and temporary components.

2.4 | Weighted single-step genome-wide association study

The analyses were performed using the weighted single-step genome-wide association study (WssGWAS) methodology (Wang, Misztal, Aguilar, Legarra, & Muir, 2012), considering the same linear animal model used to estimate the (co) variance components mentioned before. The animal effects were decomposed into those for genotyped ($\mathbf{a_g}$) and ungenotyped animals ($\mathbf{a_n}$). The animal effects of genotyped animals are a function of the SNP effects, $\mathbf{a_g} = \mathbf{Zu}$, where $\mathbf{Z}$ is a matrix relating genotypes of each locus and $\mathbf{u}$ is a vector of the SNP marker effect. The variance of animal effects was assumed as:

$$\mathrm{Var}(\mathbf{a_g}) = \mathrm{Var}(\mathbf{Zu}) = \mathbf{ZDZ}'\sigma_u^2 = \mathbf{G}^*\sigma_a^2$$

where $\mathbf{D}$ is a diagonal matrix of weights for variances of markers ($\mathbf{D} = \mathbf{I}$ for GBLUP), σ_u^2 is the genetic additive variance captured by each SNP marker when the weighted relationship matrix ($\mathbf{G}^*$) was built with no weight.

The SNP effects were obtained using following equation:

$$\hat{u} = \lambda\mathbf{DZ}'\mathrm{G}^{*-1}\hat{a}_g = \mathbf{DZ}'[\mathbf{ZDZ}']^{-1}\hat{a}_g$$

where λ was defined by VanRaden et al. (2009) as a normalizing constant, as described below:

$$\lambda = \frac{\sigma_u^2}{\sigma_a^2} = \frac{1}{\sum_{i=1}^{M} 2p_i(1-p_i)}$$

TABLE 1 Descriptive statistics for 305-day milk yield and lactation curve parameters[a]

	Primiparous	Multiparous
Trait	**Mean (*SD*)**	**Mean (*SD*)**
305-day milk (kg)	8,286 (1,471)	9,966 (1990)
Ramp	30.60 (0.24)	23.27 (2.62)
Scale (kg/day)	35.57 (6.03)	49.79 (9.76)
Decay[b]	1.42 (0.40)	2.59 (0.68)
Peak time (day)	72.66 (10.48)	46.91 (8.03)
Peak yield (kg/day)	30.67 (5.26)	41.28 (7.93)

[a]Calculated using the following model (MilkBot model): $y_t = a\left(1-\frac{\exp\left(\frac{c-t}{b}\right)}{2}\right)\exp(-dt)$, in this function, a is the scale parameter, representing the theoretical maximum daily yield; b is the ramp parameter, controlling the rate of rise in milk production in early lactation; c is the offset parameter, describing the offset in time between parturition and the start of lactation; and d is the decay parameter, representing the rate of senescence of production capacity. The time at which peak lactation occurred (t_{peak}) was defined as: $t_{peak} = -b\ln\left(\frac{2db}{db+1}\right)+c$, and peak milk production was calculated by substitution the t_{peak} in the MilkBot equation.
[b]The decay × 1,000.

The following iterative process described by Wang et al. (2012) was used to estimate the SNP effects. 1. $\mathbf{D} = \mathbf{I}$ in the first step; 2. to calculate the **G** matrix; 3. to calculate GEBVs for the entire data set using ssGBLUP; 4. to convert GEBVs to SNP effects ($\hat{u}$):$\hat{u} = \lambda \mathbf{DZ}'\mathbf{G}^{*-1}\hat{a}_g$; 5. to calculate the variance of each SNP:$d_i = \hat{u}_i^2 2p_i(1-p_i)$, where i is the ith SNP; and 6. to normalize SNP weights to keep the total genetic variance constant; exit or loop to step 2. The effects of markers were obtained by three iterations from step 2 to 6. Accuracies of GEBVs were obtained using the following formula:

$$\text{acc} = \sqrt{1-\frac{\text{PEV}}{\sigma_g^2}}$$

where PEV is the prediction error variance, and σ_g^2 is the additive genetic variance of the trait. The most accurate genomic evaluation was provided at iteration 2 which was used for estimating the percentage of genetic variance explained by ith genomic region as follow:

$$\frac{\text{Var}(a_i)}{\sigma_a^2}\times 100\% = \frac{\text{Var}\left(\sum_{j=1}^{50} Z_j\hat{u}_j\right)}{\sigma_a^2}\times 100$$

where a_i is the genetic value of the ith region that consists of 50 consecutive SNP, σ_a^2 is the total genetic variance, **Zj** is the vector of the SNP content of the jth SNP for all individuals, and $\hat{u}_j$ is the marker effect of the jth SNP within the ith region. The results were presented by the proportion of variance explained by each window of 50 consecutive SNP with an average of 165 Kb.

2.5 | Linkage disequilibrium analysis

The square of the correlation coefficient between two loci (r^2) was used to map linkage disequilibrium (LD) using PreG SF90 (Aguilar, Misztal, Tsuruta, Legarra, & Wang, 2014) with a sliding window size of 200 Kb across the same chromosome. LD blocks were overlaid with the coordinate of the association windows as a further annotation layer.

2.6 | Gene prospection

In a post-GWAS study, gene ontology (GO) enrichment analysis can be performed to investigate pathways and biological processes that are shared by candidate genes related to associated regions identified in GWAS (Verardo et al., 2015). In this study, the chromosome segments that explained more than 0.50% of the additive genetic variance were selected to explore and determine potential quantitative trait loci (QTL). The Map Viewer tool of the bovine genome available at the National Center for Biotechnology Information (NCBI—http://www.ncbi.nlm.nih.gov) in the UMD3.1 bovine genome assembly and Ensembl Genome Browser (http://www.ensembl.org/index.html) was used for identification of genes. The list of genes inside the chromosome segments that explained more than 0.50% of additive genetic variance for each trait, considered as positional candidate genes, was uploaded to Enrichr for GO enrichment analysis (Chen et al., 2013; Kuleshov et al., 2016). Significantly enriched terms with at least four genes from the input gene list were identified based on the retrieved adjusted p value.

3 | RESULTS

3.1 | Variance components

Descriptive statistics for 305-day milk yield and lactation curve parameters in both primi- and multiparous cows are presented in Table 1. Variance components, calculated using the AIREML method, for additive, permanent environmental (for multiparous cows) and residual variances are in Table 2. Estimated heritabilities (*SD*) were 0.37 (0.01), 0.34 (0.01), 0.17 (0.01), 0.12 (0.01), 0.30 (0.01) and 0.19 (0.01), respectively, for 305-day milk yield, peak yield, peak time, ramp, scale and decay for primiparous cows. Genetic correlations (*SD*) of 305-day milk yield with peak yield, peak time, ramp, scale and decay in primiparous cows were 0.99 (0.001), 0.63 (0.001), 0.20 (0.001), 0.97 (0.001) and −0.52 (0.001), respectively.

TABLE 2 Variance components for 305- milk yield and lactation curve parameters[a]

	Primiparous		Multiparous		
Trait	σ_a^2 **(*SE*)**	σ_e^2 **(*SE*)**	σ_a^2 **(*SE*)**	σ_p^2 **(*SE*)**	σ_e^2 **(*SE*)**
305-day milk (kg)	526,630 (32,046)	892,670 (21,683)	685,000 (57,726)	414,700 (49,863)	1,498,600 (34,048)
Scale (kg/day)	7.58 (0.52)	17.19 (0.39)	12.42 (1.19)	4.85 (1.11)	42.92 (0.93)
Ramp	0.0056 (0.00071)	0.0408 (0.00080)	0.3021 (0.06775)	0.0145 (0.08288)	5.6433 (0.1010)
Decay[b]	0.02228 (0.002076)	0.09256 (0.001944)	0.07120 (0.006737)	0.01143 (0.006347)	0.27119 (0.005747)
Peak time (day)	14.39 (1.43)	69.86 (1.43)	7.04 (0.86)	1.43 (0.94)	46.08 (0.94)
Peak yield (kg/day)	6.36 (0.40)	11.99 (0.28)	9.08 (0.83)	4.58 (0.75)	26.23 (0.58)

[a]Calculated using the following model (MilkBot model): $y_t = a\left(1 - \frac{\exp\left(\frac{c-t}{b}\right)}{2}\right)\exp(-dt)$, in this function, a is the scale parameter, representing the theoretical maximum daily yield; b is the ramp parameter, controlling the rate of rise in milk production in early lactation; c is the offset parameter, describing the offset in time between parturition and the start of lactation; and d is the decay parameter, representing the rate of senescence of production capacity. The time at which peak lactation occurred (t_{peak}) was defined as: $t_{peak} = -b\ln\left(\frac{2db}{db+1}\right) + c$, and peak milk production was calculated by substitution t_{peak} in the MilkBot equation.

[b]The decay × 1,000.

TABLE 3 Identification of genes based on additive genetic variance explained by windows of 50 adjacent SNP for 305-day milk yield and lactation curve parameters[a] in primiparous cows

Chromosome	Position (bp)	Genes[b]	Trait (% variance explained)
BTA14	1480260–1683767	*MIR2308, LOC104973955, CYHR1, FOXH1, COMMD5, TONSL, PPP1R16A, MFSD3, LRRC24, C14H8orf33, RECQL4, ARHGAP39, RPL8, GPT, LRRC14, ZNF34, C14H8orf82, ZNF7, KIFC2*	305-day milk (1.21), peak yield (1.04), peak time (0.47), ramp (0.37), scale (0.81), decay (0.57)
BTA14	1855090–2118405	*LOC100141215, MIR2309, MIR1839, LOC101907640, LOC101908059, LOC104968841, LOC104973958, LOC104973959, LOC104973960, LOC104973961, OPLAH, HGH1, LOC509114, GRINA, PARP10, MAF1, SHARPIN, CYC1, GPAA1, MROH1, LOC523023, EXOSC4, SPATC1, LOC786966*	305-day milk (1.09), peak yield (0.94), peak time (0.38), ramp (0.28), scale (0.73), decay (0.44)
BTA14	2676321–2940147	*PSCA, LY6K, LOC100848939, LOC101904969, LOC101905222, LOC104973965, LOC104973966, THEM6, LYNX1, JRK, ARC, SLURP1, LY6D, GML, LOC787628, LYPD2*	305-day milk (0.94), peak yield (0.82), peak time (0.55), ramp (0.37), scale (0.64), decay (0.50)

[a]Calculated using the following model (MilkBot model): $y_t = a\left(1 - \frac{\exp\left(\frac{c-t}{b}\right)}{2}\right)\exp(-dt)$, in this function, a is the scale parameter, representing the theoretical maximum daily yield; b is the ramp parameter, controlling the rate of rise in milk production in early lactation; c is the offset parameter, describing the offset in time between parturition and the start of lactation; and d is the decay parameter, representing the rate of senescence of production capacity. The time at which peak lactation occurred (t_{peak}) was defined as: $t_{peak} = -b\ln\left(\frac{2db}{db+1}\right) + c$, and peak milk production was calculated by substitution t_{peak} in the MilkBot equation.

[b]Official gene symbol (assembly UMD_3.1, annotation release 103).

Estimated heritabilities (*SD*) were 0.26 (0.02), 0.23 (0.02), 0.13 (0.02), 0.05 (0.02), 0.21 (0.02) and 0.20 (0.02), respectively, for 305-day milk yield, peak yield, peak time, ramp, scale and decay for multiparous cows. Corresponding values for repeatability estimates were 0.42 (0.01), 0.34 (0.01), 0.15 (0.01), 0.05 (0.01), 0.29 (0.01) and 0.24 (0.01). Genetic correlations of 305-day milk yield with peak yield, peak time, ramp, scale and decay for multiparous cows were 0.95 (0.005), 0.52 (0.005), 0.15 (0.005), 0.89 (0.005) and −0.49 (0.005), respectively.

3.2 | Genome-wide association study

3.2.1 | Primiparous cows

General information about the results of ssGWAS for primiparous cows is described in Data S1-S6. The windows associated with 305-day milk yield and lactation curve parameters in primiparous cows along with the genes inside them are presented in Table 3. The results identified three windows associated with the 305-day milk yield or the parameters of

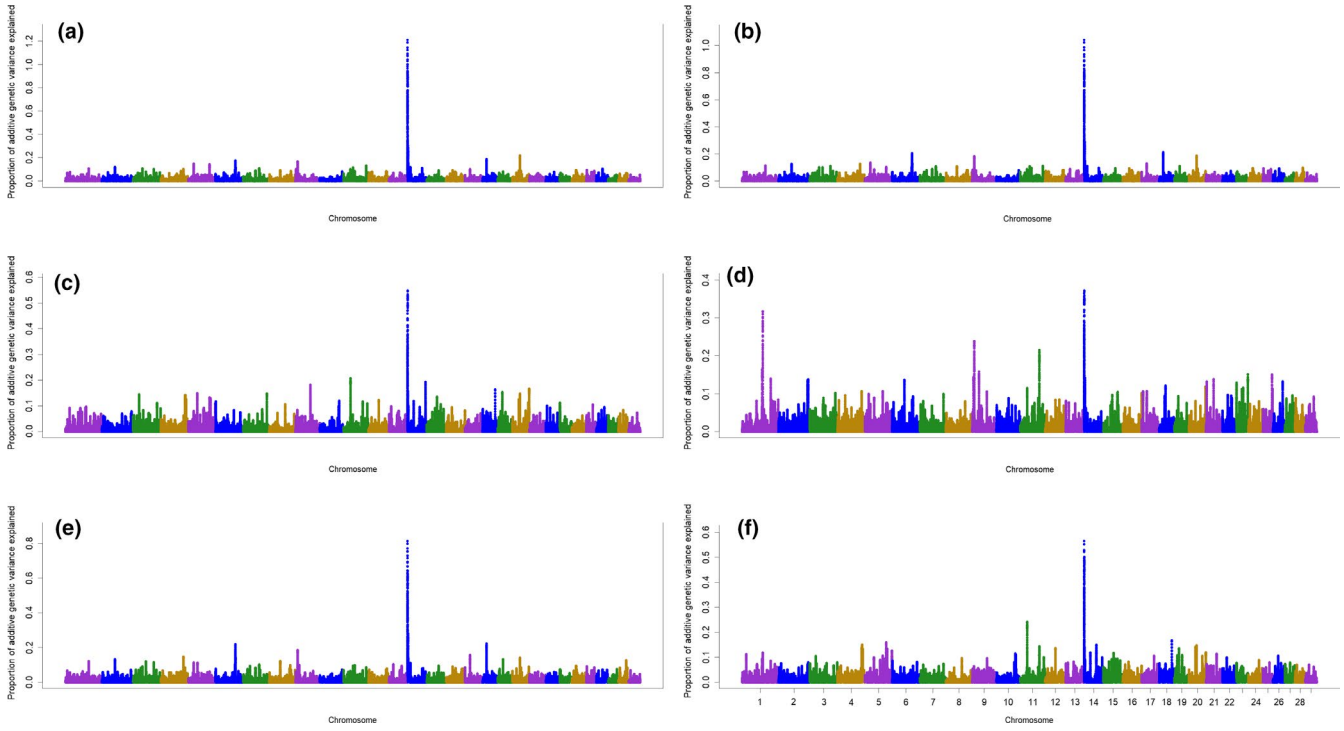

FIGURE 1 Additive genetic variance explained by windows of 50 adjacent SNP across chromosomes for 305-day milk yield (a), peak yield (b), peak time (c), ramp (d), scale (e) and decay (f) in primiparous cows

TABLE 4 Identification of genes based on additive genetic variance explained by windows of 50 adjacent SNP for 305-day milk yield and lactation curve parameters[a] in multiparous cows

Chromosome	Position (bp)	Genes[b]	Trait (% variance explained)
BTA14	1480260–1683767	*MIR2308, LOC104973955, CYHR1, FOXH1, COMMD5, TONSL, PPP1R16A, MFSD3, LRRC24, C14H8orf33, RECQL4, ARHGAP39, RPL8, GPT, LRRC14, ZNF34, C14H8orf82, ZNF7, KIFC2*	305-day milk (1.10), peak yield (0.89), peak time (0.25), scale (0.71), decay (0.32)
BTA14	1855090–2118405	*LOC100141215, MIR2309, MIR1839, LOC101907640, LOC101908059, LOC104968841, LOC104973958, LOC104973959, LOC104973960, SPATC1, LOC786966, LOC104973961, OPLAH, HGH1, LOC509114, GRINA, PARP10, MAF1, SHARPIN, CYC1, GPAA1, MROH1, LOC523023, EXOSC4*	305-day milk (0.88), peak yield (0.69), peak time (0.20), scale (0.56), decay (0.22)
BTA14	2676321–2940147	*PSCA, LY6K, LOC100848939, LOC101904969, LOC101905222, JRK, LYPD2, LOC104973965, LOC104973966, THEM6, LYNX1, ARC, SLURP1, LY6D, GML, LOC78762*	305-day milk (0.79), peak yield (0.60), peak time (0.36),, scale (0.48), decay (0.36)

[a]Calculated using the following model (MilkBot model): $y_t = a\left(1 - \frac{\exp\left(\frac{c-t}{b}\right)}{2}\right)\exp(-dt)$, in this function, a is the scale parameter, representing the theoretical maximum daily yield; b is the ramp parameter, controlling the rate of rise in milk production in early lactation; c is the offset parameter, describing the offset in time between parturition and the start of lactation; and d is the decay parameter, representing the rate of senescence of production capacity. The time at which peak lactation occurred (t_{peak}) was defined as:$t_{peak} = -b\ln\left(\frac{2db}{db+1}\right) + c$, and peak milk production was calculated by substitution t_{peak} in the MilkBot equation.

[b]Official gene symbol (assembly UMD_3.1, annotation release 103).

the lactation curve in primiparous cows (Table 3; Figure 1). These three regions combined explained more than 3.24%, 2.80%, 1.40%, 2.18% and 1.51% of the total genetic variances of 305-day milk yield, peak yield, peak time, scale and decay parameter of the lactation curve, respectively. However, no window was found to explain more than 0.50% of additive genetic variance of ramp. A region was found on BTA14 position 2.67–2.94 Mb which was associated with 305-day milk

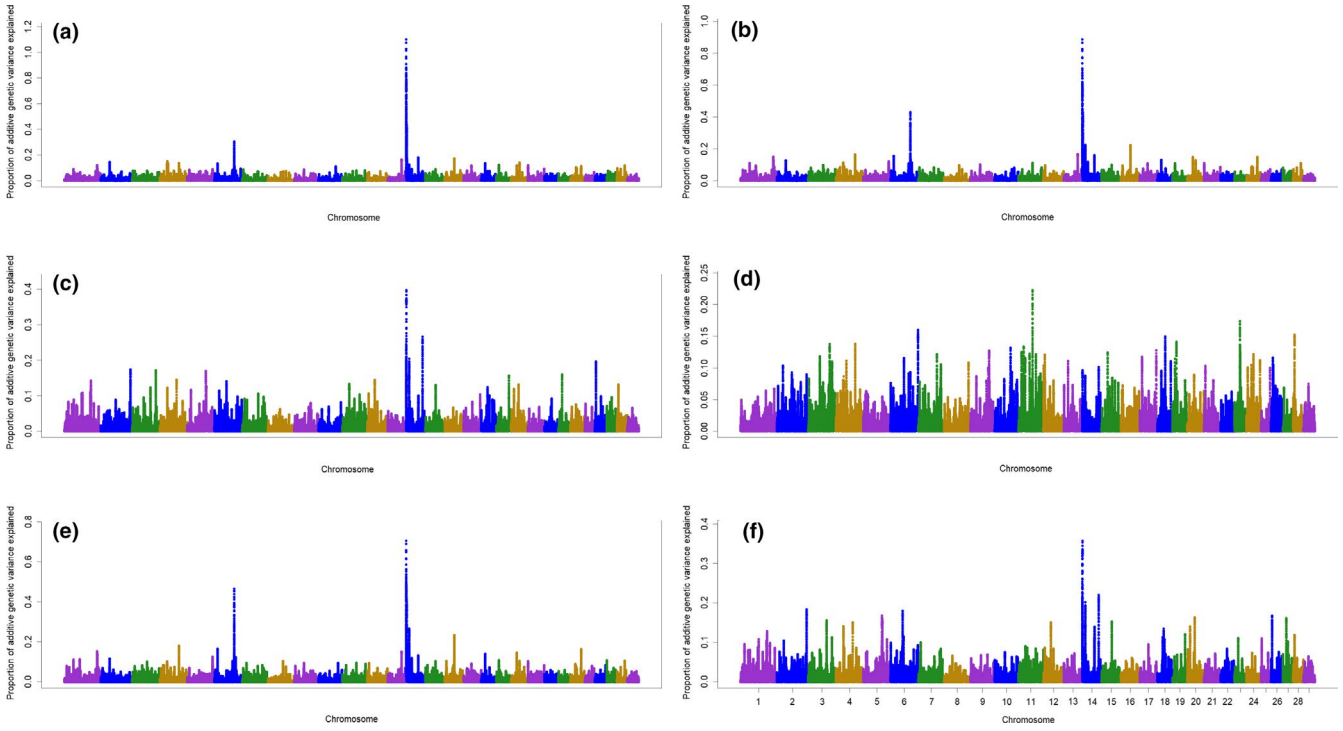

FIGURE 2 Additive genetic variance explained by windows of 50 adjacent SNP across chromosomes for 305-day milk yield (a), peak yield (b), peak time (c), ramp (d), scale (e) and decay (f) in multiparous cows

yield, peak yield, peak time, scale and decay parameter of the lactation curve. A window was found on BTA14 in position 1.48–1.68 Mb which explained more than 1.21%, 1.04%, 0.81% and 0.57% of additive genetic variances of 305-day milk yield, peak yield, scale and decay, respectively. A region on BTA14 in position 1.85–2.11 Mb was identified to be associated with 305-day milk yield, peak yield and scale. A total of 59 genes were found to be associated with 305-day milk yield and lactation curve parameters in primiparous cows.

3.2.2 | Multiparous cows

General information about all results of ssGWAS for multiparous cows is described in Data S7–S12. The windows associated with 305-day milk yield and lactation curve parameters in multiparous cows along with the genes inside them are presented in Table 4. The results identified three windows associated with 305-day milk yield and the parameters of lactation curve in multiparous cows (Table 4; Figure 2). The identified windows were quiet similar to those identified for primiparous cows. These three regions combined explained more than 2.69%, 2.18%, 0.86%, 1.75% and 0.90% of the total genetic variances of 305-day milk yield, peak yield, peak time, scale and decay, respectively. However, no window was found to explain more than 0.50% of additive genetic variance of the peak time, decay or ramp in multiparous cows.

3.2.3 | Linkage disequilibrium (LD) analysis

In total, four LD blocks consisting of 15, 5, 3 and 3 SNP were found in the region identified on BTA14 position 1.48–1.68 Mb (Figure 3), while the biggest block (across a 37 Kb region) contains genes including *ZNF7, ZNF34, RPL8, COMMD5* and *C14H8orf33*. Three LD blocks consisted of 7, 6 and 4 SNP were found in the identified region on BTA14 position 1.85–2.11 Mb, while the biggest LD block contains genes including *MIR2309* and *LOC786966*. The biggest LD block found in the window identified on BTA14 in position 2.67–2.94 Mb, consisted of nine SNP and contains genes including *LOC101905222*, *JRK* and *ARC* (results not shown).

3.2.4 | Gene ontology enrichment analysis

Significantly enriched biological processes with at least four genes from the input gene list are shown in Table 5. The C-terminal protein lipidation (GO: 0006501), C-terminal protein amino acid modification (GO: 0018410) and protein lipidation (GO: 0006497) were identified as the enriched biological processes. The protein lipidation, the covalent attachment of lipid groups to an amino acid in a protein, and C-terminal protein amino acid modification, the alteration of the C-terminal amino acid residue in a protein, are child terms of the C-terminal protein lipidation, the covalent attachment of a lipid group to the carboxy-terminus of a protein.

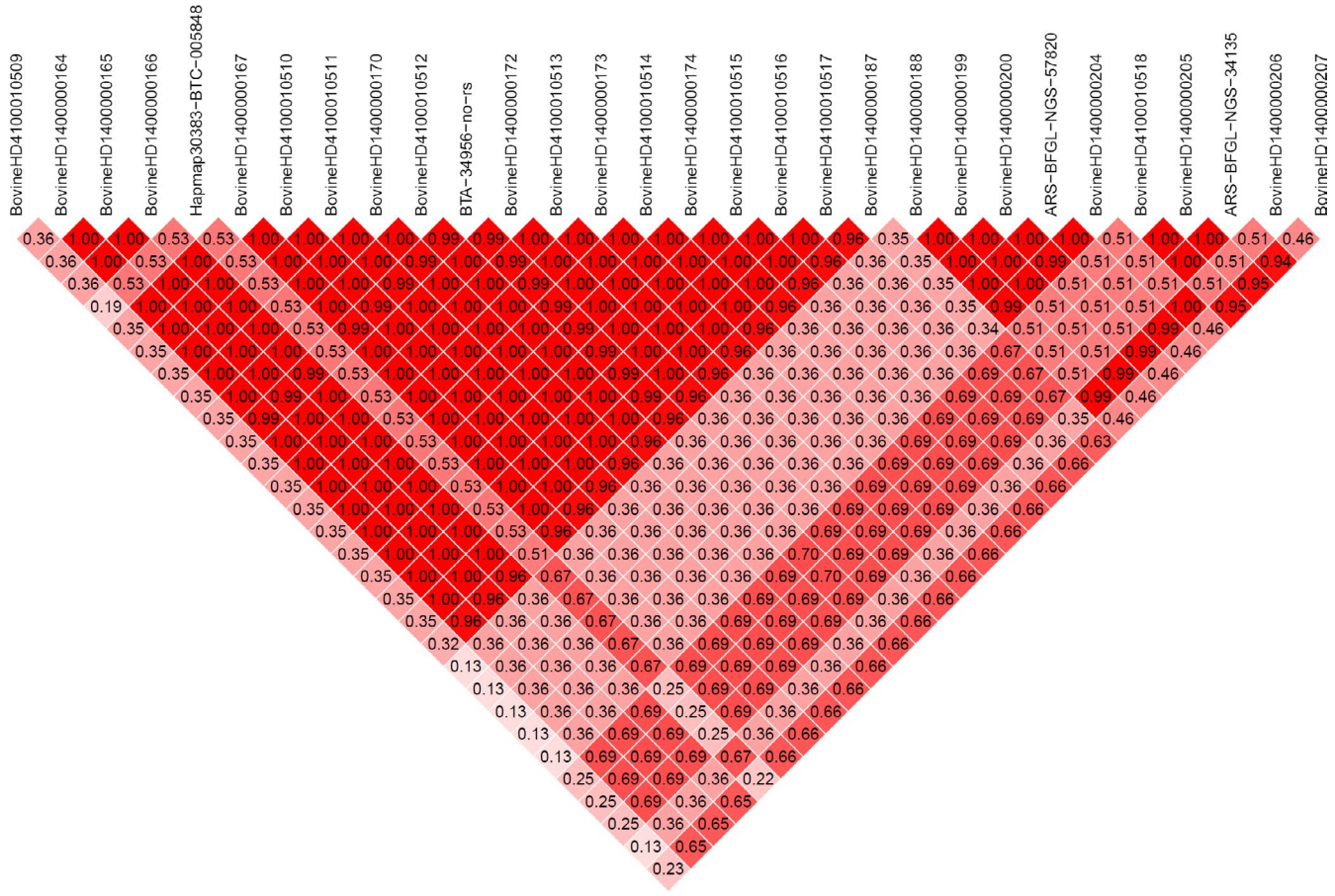

FIGURE 3 Linkage disequilibrium between 30 SNP inside the genomic region on BTA14 in position 1.48–1.68 associated with 305-day milk and lactation curve parameters in both primiparous and multiparous cows. The colour scale ranges from red to white (colour intensity decreases with decreasing r^2 value). Strong LD was detected across a 37 kb region between SNP BovineHD1400000167 and BovineHD1400000167. This region contains genes including *ZNF7, ZNF34, RPL8, COMMD5* and *C14H8orf33*

TABLE 5 Gene ontologies (GO) terms enriched by the genes inside the chromosomal region of associated with milk production and lactation curve parameters

Trait	GO term description	Genes
305-day milk, Scale, peak yield, peak time, ramp	C-terminal protein lipidation (GO:0006501)	*LY6H, LY6K, LY6D, LYPD2, PSCA, GPIHBP1*
305-day milk, scale, peak yield, peak time, ramp	C-terminal protein amino acid modification (GO:0018410)	*LY6H, LY6K, LY6D, LYPD2, PSCA, GPIHBP1*
305-day milk, scale, peak yield, peak time, ramp	protein lipidation (GO:0006497)	*LY6H, LY6K, LY6D, LYPD2, PSCA, GPIHBP1*

4 | DISCUSSION

The heritability of 305-day milk yield was 0.37 and 0.26, respectively, in primiparous and multiparous cows. The heritability for lactation curve parameters in primiparous cows ranged from 0.12 (ramp) to 0.34 (peak yield). The corresponding values in multiparous cows were 0.05–0.23, respectively, for ramp and peak yield. Previous researchers have reported that peak yield is more heritable compared with other parameters of the lactation curve (Gebreyohannes, Koonawootrittriron, Elzo, & Suwanasopee, 2013; Saghanezhad, Atashi, Dadpasand, Zamiri, & Shokri-Sangari, 2017; Shanks, Berger, Freeman, & Dickinson, 1981). Genetic correlations of 305-day milk yield with lactation curve parameters ranged from −0.52 (decay) to 0.99 (peak yield). The corresponding values in multiparous cows were −0.49 to 0.95, respectively, for decay and peak yield. The higher correlation between 305-day milk yield and peak yield compared to decay, as an indicator for lactation persistency, indicates that peak yield is more important in determining the lactation yield than persistency and can be used as a management tool to

monitor milk production performance of the herd (Ali & Schaeffer, 1987; Atashi, Zamiri, & Sayyadnejad, 2012; Gebreyohannes et al., 2013; Tekerli et al., 2000).

In this study, the weighted single-step genomic BLUP (WssGWAS) approach described by Wang et al. (2012) was used to identify genomic region(s) associated with 305-day milk yield and lactation curve parameters in Holstein dairy cows. The WssGWAS approach integrates all phenotypic, genotypic and pedigree data simultaneously; therefore, there is no need to calculate pseudo-phenotypes. In addition, this approach allows the use of different weights for SNP according to their importance, which is a deviation from the non-realistic GBLUP assumption of the infinitesimal model and improves the precision of estimates of SNP effects. In this procedure, the $\mathbf{H}^{-1}$ matrix, calculated by combining all known pedigree and genotype information, is used in the ssGBLUP to estimate genomic estimated breeding values (GEBV) for all animals. Then, the GEBVs of the genotyped animals are used to estimate effects for the SNP. Finally, SNP effects are used to calculate the percentage of genetic variance explained by sets of consecutive SNP (SNP windows). In this study, the proportion of additive genetic variance explained by windows of 50-adjacent SNP was calculated and the regions that accounted for more than 0.50% of the additive genetic variance were identified as potential QTL. The present study, identified three windows on BTA14 (in position 1.48–1.68, 1.85–2.11 and 2.67–2.94 Mb, respectively) in both primi- and multiparous cows associated with 305-day milk yield or lactation curve parameters. The identified regions overlap with QTL regions for multiple traits including milk yield and milk composition, somatic cell count, calving ease and average daily gain in cattle (Bennewitz et al., 2003; Boichard et al., 2003; Buitenhuis et al., 2014; Iso-Touru et al., 2016; Jiang et al., 2010; Lund et al., 2008; Nayeri et al., 2016; Rupp & Boichard, 2003). Iso-Touru et al. (2016) identified 755 SNP in six different chromosomes (BTA5, BTA14, BTA16, BTA19, BTA20 and BTA25) associated with milk yield in Nordic Red cattle with the highest number of significant SNP on BTA14. Nayeri et al. (2016) using a single SNP regression mixed linear model, identified 292 SNP associated with milk yield in Canadian Holsteins, with the highest number of significant SNP on BTA14. Meredith et al. (2012), using a single SNP regression approach, identified 370 SNP associated with milk yield in Irish Holsteins with the highest number of significant SNP on BTA14.

A region on BTA14 in position 1.85–2.11 Mb was identified to be associated with 305-day milk yield, peak yield and scale in both primi- and multiparous cows. This region contains several genes including *MIR2309, MIR1839, OPLAH, HGH1, GRINA, PARP10, MAF1, SHARPIN, CYC1, GPAA1, MROH1, EXOSC4* and *SPATC1*. The association of genes including *OPLAH, GRINA* and *MF1* with milk yield and lactation performance has been reported in previous studies (Kolbehdari et al., 2009; Nayeri & Stothard, 2016; Wang, Ning, Liu, Zhang, & Jiang, 2019); however, the association of the remaining identified genes inside this region with lactation performance in dairy cows has not been reported previously.

A region on BTA14 in position 1.48–1.68 Mb was identified to be associated with 305-day milk yield, peak yield, scale and decay in primiparous cows. This region was also associated with 305-day milk yield, peak yield and the scale in multiparous cows. Previous studies have reported this region to be associated with milk fat yield and milk protein percentage (Bagnato et al., 2008; Rodriguez-Zas, Southey, Heyen, & Lewin, 2002). This region contains several genes including *MIR2308, CYHR1, FOXH1, COMMD5, TONSL, PPP1R16A, MFSD3, LRRC24, C14H8orf33, RECQL4, ARHGAP39, RPL8, GPT, LRRC14, ZNF34, C14H8orf82, ZNF7* and *KIFC2*. The association of genes including *TONSL, PPP1R16A, FOXH1, ARHGAP39, CYHR1* and *ARHGAP39* with milk yield and milk composition has been reported in previous studies (Buitenhuis et al., 2014; Nayeri & Stothard, 2016; Ning et al., 2017; Wang et al., 2019). Nayeri et al. (2016) reported that highly significant SNP for milk yield in Canadian Holsteins were mapped inside genes including *CPSF1*, *DGAT1*, *TONSL*, *CYHR1*, *FOXH1* and *PPP1R16A*.

It is documented that cows which reach peak production later, produce more 305-day milk, more milk at peak, and show a higher milk yield persistency (Saghanezhad et al., 2017). The only windows associated with 305-day milk yield, peak yield, peak time and decay was identified on BTA14 in position 2.67–2.94 Mb. This region which contains several genes, including *PSCA, LY6K, THEM6, LYNX1, JRK, ARC, SLURP1, LY6D, GML* and *LYPD2,* was also associated with 305-day milk yield and peak yield in multiparous cows. Buitenhuis et al. (2014) reported that the *GML* is associated with milk fat and protein percentage; however, the association of the remaining identified genes inside this region with lactation performance in dairy cows has not been reported before.

The decay parameter represents the rate of decline in milk production after peak and can be considered as a measure for lactation persistency. The genomic regions explaining variation in lactation persistency have been investigated in several GWAS (Kolbehdari et al., 2008; Nayeri et al., 2017). Nayeri et al. (2017) identified 83 SNP in four different chromosomes (BTA6, BTA13, BTA20 and BTA27) to be associated with lactation persistency (defined as the average of expected milk yield at day 280 in lactation compared with that at day 60 in lactation) in Canadian Holsteins with the highest number of significant SNP on BTA20. Do et al. (2017) reported eight SNP on BTA2, 5, 9, 14, 19 and 20 to be associated with lactation persistency in Canadian Holsteins. However, the present

study identified two regions (BTA14 position 1.48–1.68 Mb and 2.67–2.94 Mb) associated with decay parameter which has not been previously reported to be associated with lactation persistency in dairy cattle.

The square of the correlation coefficient between two loci (r^2) was used to map LD in the identified windows. In total, four LD blocks were found in the region identified on BTA14 in position 1.48–1.68 Mb. These strong LD blocks reflect the strong selection pressure towards allele fixation that has been carried out in this part of the bovine genome (Khatkar et al., 2008; McKay et al., 2007). The C-terminal protein lipidation (GO: 0006501), C-terminal protein amino acid modification (GO: 0018410) and protein lipidation (GO: 0006497) were identified as the enriched biological process terms. The protein lipidation, the covalent attachment of lipid groups to an amino acid in a protein, and C-terminal protein amino acid modification, the alteration of the C-terminal amino acid residue in a protein, are child terms of the C-terminal protein lipidation, the covalent attachment of a lipid group to the carboxy-terminus of a protein. Protein lipidation is an important co- or post-translational modification which occurs in many proteins in eukaryotic cells and regulates numerous biological pathways such as membrane trafficking, protein secretion, signal transduction and apoptosis (Chen, Sun, Niu, Jarugumilli, & Wu, 2018; Jiang et al., 2018). Protein lipidation is essential for binding and partitioning in different membrane microdomains, and for the interaction with effectors and the regulation of signalling processes, thereby playing a key role in controlling protein localization and function (Triola, 2011).

5 | CONCLUSION

The objective of this study was to identify genomic regions associated with milk yield and the shape of lactation curve in Holstein dairy cows. The lactation curve parameters, the slope of the initial rise of the curve, peak yield, time to peak and the slope of the curve after peak yield were used as new phenotypic variables in the GWAS. Among the parameters of lactation curve, scale and peak yield showed the highest heritability and the highest genetic correlation with 305-day milk yield which can explain the overlapping regions among these traits. The genomic regions were found to be associated with 305-day milk yield, scale and peak yield in both primi- and multiparous cows. Although during the last decade many animals have been genotyped using high-density SNP chip panels, no significant impact has been observed on genetic improvement programme (Erbe et al., 2012; VanRaden et al., 2013). However, Abo-Ismail et al. (2017) reported that combining the significant SNPs or SNPs within or nearby gene(s) from the HD panel with the BovineSNP50 panel yielded a marginal increase in the accuracy of prediction of genomic estimated breeding values compared to the use of the BovineSNP50 panel alone.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement no. 613689. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

List of authors within the GplusE consortium are Niamh McLoughlin, Alan Fahey, Elizabeth Matthews, Andreia Santoro, Colin Byrne, Pauline Rudd, Roisin O'Flaherty, Sinead Hallinan, Claire Wathes, Zhangrui Cheng, Ali Fouladi, Geoff Pollott, Dirk Werling, Beatriz Sanz Bernardo, Alistair Wylie, Matt Bell, Mieke Vaneetvelde, Kristof Hermans, Geert Opsomer, Sander Moerman, Jenne De Koster, Hannes Bogaert, Jan Vandepitte, Leila Vandevelde, Bonny Vanranst, Johanna Hoglund, Susanne Dahl, Soren Ostergaard, Janne Rothmann, Mogens Krogh, Else Meyer, Charlotte Gaillard, Jehan Ettema, Tine Rousing, Federica Signorelli, Francesco Napolitano, Bianca Moioli, Alessandra Crisà, Luca Buttazzoni, Jennifer McClure, Daragh Matthews, Francis Kearney, Andrew Cromie, Matt McClure, Shujun Zhang, Xing Chen, Huanchun Chen, Junlong Zhao, Liguo Yang, Guohua Hua, Chen Tan, Guiqiang Wang, Michel Bonneau, Andrea Pompozzi, Armin Pearn, Arnold Evertson, Linda Kosten, Anders Fogh, Thomas Andersen, Matthew Lucey, Chris Elsik, Gavin Conant, Jerry Taylor, Nicolas Gengler, Michel Georges, Frédéric Colinet, Marilou Ramos Pamplona, Hedi Hammami, Catherine Bastin, Haruko Takeda, Aurelie Laine, Anne-Sophie Van Laere, Martin Schulze, Sergio Palma Vera, Conrad Ferris and Cinzia Marchitelli.

CONFLICT OF INTEREST

The authors declare that they do not have any conflicts of interest.

DATA AVAILABILITY STATEMENT

Relevant information supporting the results not presented in the manuscript is given in additional files (Data S1–S12).

ORCID

Hadi Atashi https://orcid.org/0000-0002-6853-6608
Miel Hostens https://orcid.org/0000-0001-5376-976X

REFERENCES

Abo-Ismail, M. K., Brito, L. F., Miller, S. P., Sargolzaei, M., Grossi, D. A., Moore, S. S., … Schenkel, F. S. (2017). Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle. *Genetics Selection Evolution*, *49*(1), 82. https://doi.org/10.1186/s12711-017-0356-8

Aguilar, I., Misztal, I., Johnson, D., Legarra, A., Tsuruta, S., & Lawlor, T. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, *93*(2), 743–752. https://doi.org/10.3168/jds.2009-2730

Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A., & Wang, H. (2014). *PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs*. Paper presented at the Proceedings of the 10th world congress of genetics applied to livestock production.

Ali, T., & Schaeffer, L. (1987). Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*, *67*(3), 637–644. https://doi.org/10.4141/cjas87-067

Atashi, H., Zamiri, M., & Dadpasand, M. (2013). Association between dry period length and lactation performance, lactation curve, calf birth weight, and dystocia in Holstein dairy cows in Iran. *Journal of Dairy Science*, *96*(6), 3632–3638. https://doi.org/10.3168/jds.2012-5943

Atashi, H., Zamiri, M. J., & Sayyadnejad, M. B. (2012). Effect of twinning and stillbirth on the shape of lactation curve in Holstein dairy cows of Iran. *Archives Animal Breeding*, *55*(3), 226–233. https://doi.org/10.5194/aab-55-226-2012

Bagnato, A., Schiavini, F., Rossoni, A., Maltecca, C., Dolezal, M., Medugorac, I., … Lipkin, E. (2008). Quantitative trait loci affecting milk yield and protein percentage in a three-country Brown Swiss population. *Journal of Dairy Science*, *91*(2), 767–783. https://doi.org/10.3168/jds.2007-0507

Bennewitz, J., Reinsch, N., Grohs, C., Levéziel, H., Malafosse, A., Thomsen, H., … Kalm, E. (2003). Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genetics Selection Evolution*, *35*(3), 319. https://doi.org/10.1186/1297-9686-35-3-319

Bignardi, A. B., El Faro, L., Cardoso, V. L., Machado, P. F., & de Albuquerque, L. G. (2009). Random regression models to estimate test-day milk yield genetic parameters Holstein cows in Southeastern Brazil. *Livestock Science*, *123*(1), 1–7. https://doi.org/10.1016/j.livsci.2008.09.021

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., … Levéziel, H. (2003). Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution*, *35*(1), 77. https://doi.org/10.1186/1297-9686-35-1-77

Buitenhuis, B., Janss, L. L., Poulsen, N. A., Larsen, L. B., Larsen, M. K., & Sørensen, P. (2014). Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. *BMC Genomics*, *15*(1), 1112. https://doi.org/10.1186/1471-2164-15-1112

Chen, B., Sun, Y., Niu, J., Jarugumilli, G. K., & Wu, X.. (2018). Protein lipidation in cell signaling and diseases: function, regulation, and therapeutic opportunities. *Cell Chemical Biology*, *25*, 817–831.

Chen, H., Zhang, L., Li, X., Li, X., Sun, G., Yuan, X., … Liu, G. (2013). Adiponectin activates the AMPK signaling pathway to regulate lipid metabolism in bovine hepatocytes. *Journal of Steroid Biochemistry and Molecular Biology*, *138*, 445–454. https://doi.org/10.1016/j.jsbmb.2013.08.013

Cole, J., Ehrlich, J., & Null, D. (2012). Projecting milk yield using best prediction and the MilkBot lactation model. *Journal of Dairy Science*, *95*(7), 4041–4044.

Cole, J. B., Wiggans, G. R., Ma, L. I., Sonstegard, T. S., Lawlor, T. J., Crooker, B. A., … Da, Y. (2011). Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genomics*, *12*(1), 408. https://doi.org/10.1186/1471-2164-12-408

Dekkers, J., Ten Hag, J., & Weersink, A. (1998). Economic aspects of persistency of lactation in dairy cattle. *Livestock Production Science*, *53*(3), 237–252. https://doi.org/10.1016/S0301-6226(97)00124-3

Do, D., Bissonnette, N., Lacasse, P., Miglior, F., Sargolzaei, M., Zhao, X., & Ibeagha-Awemu, E. (2017). Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. *Journal of Dairy Science*, *100*(3), 1955–1970. https://doi.org/10.3168/jds.2016-11910

Ehrlich, J. (2011). Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. *Bovine Practitioner*, *45*(1), 88.

El-Awady, H. (2013). Genetic aspects of lactation curve traits and persistency indices in Friesian cows. *Archiva Zootechnica*, *16*(1), 15-29.

Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., … Goddard, M. E. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, *95*(7), 4114–4129. https://doi.org/10.3168/jds.2011-5019

Gebreyohannes, G., Koonawootrittriron, S., Elzo, M. A., & Suwanasopee, T. (2013). Variance components and genetic parameters for milk production and lactation pattern in an Ethiopian multibreed dairy cattle population. *Asian-Australasian Journal of Animal Sciences*, *26*(9), 1237. https://doi.org/10.5713/ajas.2013.13040

Hostens, M., Ehrlich, J., Van Ranst, B., & Opsomer, G. (2012). On-farm evaluation of the effect of metabolic diseases on the shape of the lactation curve in dairy cows through the MilkBot lactation model. *Journal of Dairy Science*, *95*(6), 2988–3007. https://doi.org/10.3168/jds.2011-4791

Iso-Touru, T., Sahana, G., Guldbrandtsen, B., Lund, M., & Vilkki, J. (2016). Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. *BMC Genetics*, *17*(1), 55. https://doi.org/10.1186/s12863-016-0363-8

Jiang, H., Zhang, X., Chen, X., Aramsangtienchai, P., Tong, Z., & Lin, H. (2018). Protein lipidation: Occurrence, mechanisms, biological functions, and enabling technologies. *Chemical Reviews*, *118*(3), 919–988. https://doi.org/10.1021/acs.chemrev.6b00750

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y., & Zhang, Q. (2010). Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS ONE*, *5*(10), e13661. https://doi.org/10.1371/journal.pone.0013661

Khatkar, M. S., Nicholas, F. W., Collins, A. R., Zenger, K. R., Cavanagh, J. A. L., Barris, W., … Raadsma, H. W. (2008). Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle

based on a high-density SNP panel. *BMC Genomics*, *9*(1), 187. https://doi.org/10.1186/1471-2164-9-187

Kolbehdari, D., Wang, Z., Grant, J. R., Murdoch, B., Prasad, A., Xiu, Z., … Moore, S. S. (2008). A whole-genome scan to map quantitative trait loci for conformation and functional traits in Canadian Holstein bulls. *Journal of Dairy Science*, *91*(7), 2844–2856. https://doi.org/10.3168/jds.2007-0585

Kolbehdari, D., Wang, Z., Grant, J. R., Murdoch, B., Prasad, A., Xiu, Z., … Moore, S. S. (2009). A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *Journal of Animal Breeding and Genetics*, *126*(3), 216–227. https://doi.org/10.1111/j.1439-0388.2008.00793.x

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., … Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. https://doi.org/10.1093/nar/gkw377

López, S., France, J., Odongo, N. E., McBride, R. A., Kebreab, E., AlZahal, O., … Dijkstra, J. (2015). On the analysis of Canadian Holstein dairy cow lactation curves using standard growth functions. *Journal of Dairy Science*, *98*(4), 2701–2712. https://doi.org/10.3168/jds.2014-8132

Lund, M. S., Sorensen, P., Madsen, P., & Jaffrézic, F. (2008). Detection and modelling of time-dependent QTL in animal populations. *Genetics Selection Evolution*, *40*(2), 177. https://doi.org/10.1186/1297-9686-40-2-177

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., … Moore, S. S. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genetics*, *8*(1), 74. https://doi.org/10.1186/1471-2156-8-74

Meredith, B. K., Kearney, F. J., Finlay, E. K., Bradley, D. G., Fahey, A. G., Berry, D. P., & Lynn, D. J. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genetics*, *13*(1), 21. https://doi.org/10.1186/1471-2156-13-21

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., & Lee, D. (2002). *BLUPF90 and related programs (BGF90)*. Paper presented at the Proceedings of the 7th world congress on genetics applied to livestock production.

Nayeri, S., Sargolzaei, M., Abo-Ismail, M. K., May, N., Miller, S. P., Schenkel, F., … Stothard, P. (2016). Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genetics*, *17*(1), 75. https://doi.org/10.1186/s12863-016-0386-1

Nayeri, S., Sargolzaei, M., Abo-Ismail, M., Miller, S., Schenkel, F., Moore, S., & Stothard, P. (2017). Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. *Journal of Dairy Science*, *100*(2), 1246–1258. https://doi.org/10.3168/jds.2016-11770

Nayeri, S., & Stothard, P. (2016). Tissues, metabolic pathways and genes of key importance in lactating dairy cattle. *Springer Science Reviews*, *4*(2), 49–77. https://doi.org/10.1007/s40362-016-0040-3

Ning, C., Kang, H., Zhou, L., Wang, D., Wang, H., Wang, A., … Liu, J. (2017). Performance gains in genome-wide association studies for longitudinal traits via modeling time-varied effects. *Scientific Reports*, *7*(1), 590. https://doi.org/10.1038/s41598-017-00638-2

Ning, C., Wang, D., Zheng, X., Zhang, Q., Zhang, S., Mrode, R., & Liu, J.-F. (2018). Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. *Genetics Selection Evolution*, *50*(1), 12. https://doi.org/10.1186/s12711-018-0383-0

Rekaya, R., Carabano, M., & Toro, M. (2000). Bayesian analysis of lactation curves of Holstein-Friesian cattle using a nonlinear model. *Journal of Dairy Science*, *83*(11), 2691–2701. https://doi.org/10.3168/jds.S0022-0302(00)75163-0

Rekik, B., Gara, A. B., Hamouda, M. B., & Hammami, H. (2003). Fitting lactation curves of dairy cattle in different types of herds in Tunisia. *Livestock Production Science*, *83*(2–3), 309–315. https://doi.org/10.1016/S0301-6226(03)00028-9

Rodriguez-Zas, S., Southey, B., Heyen, D., & Lewin, H. (2002). Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle. *Journal of Dairy Science*, *85*(11), 3081–3091. https://doi.org/10.3168/jds.S0022-0302(02)74395-6

Rupp, R., & Boichard, D. (2003). Genetics of resistance to mastitis in dairy cattle. *Veterinary Research*, *34*(5), 671–688. https://doi.org/10.1051/vetres:2003020

Saghanezhad, F., Atashi, H., Dadpasand, M., Zamiri, M., & Shokri-Sangari, F. (2017). Estimation of genetic parameters for lactation curve traits in Holstein Dairy Cows in Iran. *Iranian Journal of Applied Animal Science*, *7*(4), 559-566.

Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, *15*(1), 478. https://doi.org/10.1186/1471-2164-15-478

Schaeffer, L. (2016). *Random regression models*. Retrieved from http://animalbiosciences.uoguelph.ca/~lrs/BOOKS/rrmbook.pdf

Shanks, R., Berger, P., Freeman, A., & Dickinson, F. (1981). Genetic aspects of lactation curves. *Journal of Dairy Science*, *64*(9), 1852–1860. https://doi.org/10.3168/jds.S0022-0302(81)82775-0

Sherchand, L., McNew, R., Kellogg, D., & Johnson, Z. (1995). Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. *Journal of Dairy Science*, *78*(11), 2507–2513. https://doi.org/10.3168/jds.S0022-0302(95)76880-1

Silvestre, A., Petim-Batista, F., & Colaco, J. (2006). The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *Journal of Dairy Science*, *89*(5), 1813–1821. https://doi.org/10.3168/jds.S0022-0302(06)72250-0

Singh, A., Singh, A., Singh, M., Prakash, V., Ambhore, G., Sahoo, S., & Dash, S. (2016). Estimation of genetic parameters for first lactation monthly test-day milk yields using random regression test day model in Karan Fries cattle. *Asian-Australasian Journal of Animal Sciences*, *29*(6), 775. https://doi.org/10.5713/ajas.15.0643

Strucken, E., Bortfeldt, R., De Koning, D., & Brockmann, G. (2012). Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. *Animal Genetics*, *43*(4), 375–382. https://doi.org/10.1111/j.1365-2052.2011.02278.x

Tekerli, M., Akinci, Z., Dogan, I., & Akcan, A. (2000). Factors affecting the shape of lactation curves of Holstein cows from the Balikesir province of Turkey. *Journal of Dairy Science*, *83*(6), 1381–1386. https://doi.org/10.3168/jds.S0022-0302(00)75006-5

Triola, G. (2011). The protein lipidation and its analysis. *Journal of Glycomics and Lipidomics*, *2*, 2153–0637.

VanRaden, P. M., Null, D. J., Sargolzaei, M., Wiggans, G. R., Tooker, M. E., Cole, J. B., … Doak, G. A. (2013). Genomic imputation and evaluation using high-density Holstein genotypes. *Journal of Dairy Science*, *96*(1), 668–678. https://doi.org/10.3168/jds.2012-5702

VanRaden, P., Van Tassell, C., Wiggans, G., Sonstegard, T., Schnabel, R., Taylor, J., & Schenkel, F. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, *92*(1), 16–24. https://doi.org/10.3168/jds.2008-1514

Verardo, L., Silva, F., Varona, L., Resende, M., Bastiaansen, J., Lopes, P., & Guimarães, S. (2015). Bayesian GWAS and network analysis revealed new candidate genes for number of teats in pigs. *Journal of Applied Genetics*, *56*(1), 123–132. https://doi.org/10.1007/s13353-014-0240-y

Wang, D., Ning, C., Liu, J.-F., Zhang, Q., & Jiang, L. (2019). Replication of genome-wide association studies for milk production traits in Chinese Holstein by an efficient rotated linear mixed model. *Journal of Dairy Science*, *102*(3), 2378–2383.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research*, *94*(2), 73–83. https://doi.org/10.1017/S0016672312000274

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Atashi H, Salavati M, De Koster J, et al; the GplusE consortium. Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. *J Anim Breed Genet*. 2019;00:1–13. https://doi.org/10.1111/jbg.12442