Fine-mapping of BTA17 using imputed sequences for associations with de novo synthesized fatty acids in bovine milk

S. I. Duchemin,*[†] H. Bovenhuis,* H.-J. Megens,* J. A. M. Van Arendonk,* and M. H. P. W. Visker*

*Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands

[†]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden

ABSTRACT

The aim of this study was to fine-map a genomic region associated with milk fatty acids (FA) on *Bos taurus* autosome (BTA) 17. This genomic region has been discovered with 50,000 (50k) single nucleotide polymorphisms (SNP) imputed to 777,000 (777k) SNP. In this study, high-density genotypes were imputed to whole-genome sequences level to identify candidate gene(s) associated with milk FA composition on BTA17. Phenotypes and genotypes were available for 1,640 cows sampled in winter, and for 1,581 cows sampled in summer. Phenotypes consisted of gas chromatography measurements in winter and in summer milk samples of 6 individual FA and the indicator of de novo synthesis, C6:0–C14:0. Genotypes consisted of imputed 777k SNP, and 89 sequenced ancestors of the population of genotyped cows. In addition, 450 whole-genome sequences from the 1,000 Bull Genome Consortium were available. Using 495 Holstein-Friesian sequences as a reference population, the 777k SNP genotypes of the cows were imputed to sequence level. We then applied single-variant analyses with an animal model, and identified thousands of significant associations with C6:0, C8:0, C10:0, C12:0, C14:0, and C6:0–C14:0. For C8:0 in summer milk samples, the genomic region located between 29 and 34 Mbp on BTA17 revealed a total of 646 significant associations. The most significant associations [$-\log_{10}(P\text{-value}) = 7.82$] were 8 SNP in perfect linkage disequilibrium. After fitting one of these 8 SNP as a fixed effect in the model, and re-running the single-variant analyses, no further significant associations were found for any of the 6 FA or C6:0–C14:0. These findings suggest that one polymorphism underlying this QTL on BTA17 influences multiple de novo synthesized milk FA. Thirteen genes in the QTL region were identified and analyzed carefully. Six out of the 8 SNP that showed the strongest associations were located in

the La ribonucleoprotein domain family, member 1B (*LARP1B*) gene, and we suggest *LARP1B* as a primary candidate gene. Another gene of interest for this QTL region might be *PKL4*. None of these suggested candidate genes have previously been associated with milk fat synthesis or milk FA composition.

Key words: genome-wide association studies, imputation, quantitative trait loci, milk fat composition

INTRODUCTION

Bovine milk fat is an important source of fatty acids (FA) in human diets. Fatty acids from bovine milk are bioactive lipids with important biological activities regarding cell and tissue metabolism, as well as responsiveness to hormones and other signals in human cells (Calder, 2015). Many studies on bovine milk FA composition suggested that amounts of individual FA are heritable (e.g., Duchemin et al., 2013). In Dutch Holstein-Friesian cows, Stoop et al. (2008) showed that heritability estimates of individual milk FA range from 0.22 to 0.71, which indicates the presence of genetic variation for milk FA composition. Hence, it might be possible to identify genes underlying this genetic variation.

Supporting these findings, polymorphisms in the acyl-CoA:diacylglycerol acetyltransferase1 (*DGAT1*) and in the stearoyl-CoA desaturase1 (*SCD1*) genes have been associated with milk FA composition (e.g., Moioli et al., 2007; Schennink et al., 2007, 2008). In Dutch Holstein-Friesians, Bouwman et al. (2012) identified many promising genomic regions associated with individual FA in bovine milk, when performing a genome-wide association study (GWAS) with 50,000 (50k) SNP genotypes. One of these regions located on BTA17 was fine-mapped with imputed 777,000 (777k) SNP genotypes, and significant associations with multiple milk FA have been identified by Duchemin et al. (2014). In Danish Holsteins, Buitenhuis et al. (2014) performed a GWAS identifying a QTL on BTA17 associated with CLA milk FA. In the domestic water buffalo, *Bubalus bubalis*, Venturini et al. (2014) performed a GWAS on

Received March 31, 2017.

Accepted July 18, 2017.

[†]Corresponding author: sandrine.duchemin@slu.se

milk production traits and identified significant associations with milk yield, fat yield, and protein yield on BTA17 (note: BTA17 is used as a one-to-one correspondence to BBU17 in buffaloes). Nonetheless, the low density of SNP assays, which typically consist of genetic markers separated by more than 4 Mbp, have hampered the attempts of identifying causal variants underlying the identified QTL on BTA17.

With the advent of whole-genome sequences (WGS) in cattle, causal variants underlying QTL should be identified more effectively by GWAS. Whole-genome sequences should contain the polymorphisms causing the genetic differences between individuals. To overcome the high costs associated with WGS, an alternative consists of sequencing the influential ancestors of a population, and imputing the rest of this population to the sequence level. In dairy cattle, a GWAS using imputed sequences was first done by Daetwyler et al. (2014). Their study successfully fine-mapped a QTL region associated with milk fat composition on BTA27, suggesting *AGPAT6* as the most likely candidate gene underlying this QTL. Therefore, GWAS using imputed WGS can be successfully used in (fine) mapping complex traits.

The GWAS by Bouwman et al. (2012) identified a QTL region on BTA17 influencing C6:0, C8:0, and C10:0 milk FA. Duchemin et al. (2014) characterized this QTL region showing its influence on multiple short-chain FA (C6:0 to C10:0) at a similar location on BTA17. Furthermore, Duchemin et al. (2014) performed haplotype analyses, and concluded that one QTL was likely to influence multiple de novo synthesized milk FA at this location on BTA17. The aim of this study was to identify candidate gene(s) associated with de novo synthesized milk FA on BTA17 based on imputed sequences. The starting point of the present fine-mappings are the results and conclusions from Duchemin et al. (2014).

MATERIALS AND METHODS

Animals and Phenotypes

Morning milk (500 mL/cow) was sampled from 2,001 primiparous Holstein-Friesian cows belonging to 398 herds throughout the Netherlands. These samples were collected in 2 periods: February–March 2005 (referred to as winter samples) and May–June 2005 (referred to as summer samples). For each herd, most of the cows were sampled in both periods. However, some cows sampled in winter were no longer in lactation in summer. Consequently, additional cows were sampled in summer to ensure that at least 3 cows per herd were sampled in both periods. For winter milk samples, phenotypes

were available on 1,905 cows, and their lactation stages ranged from 63 to 282 d (see Stoop et al., 2008). For summer milk samples, phenotypes were available on 1,795 cows, and their lactation stages ranged from 97 to 335 d (see Duchemin et al., 2013). During the winter, all cows were kept indoors and fed silage, whereas in summer 50% of the cows could graze pasture (3.5 to 24 h/d). More information on the experimental design is available in Stoop et al. (2008).

Milk FA composition was measured by GC at the COKZ laboratory (Qlip, Leudsen, the Netherlands). Milk fat was extracted from milk samples, and FAME were prepared from fat fractions, as described by Schenink et al. (2007). The FA were identified and quantified by comparing the methyl ester chromatograms of the milk fat samples with the chromatograms of pure FAME standards (Stoop et al., 2008). The milk FA included in this study were C4:0, C6:0, C8:0, C10:0, C12:0, and C14:0, and they were expressed as weight proportion of total fat (%wt/wt). In addition, an indicator of de novo synthesized milk FA was created by combining C6:0 through C14:0 individual FA in the index designated C6:0–C14:0.

Genotypes, Imputation, and Variant Calling

Blood from cows and semen from bulls were sampled to extract DNA for genotyping purposes. First, a total of 55 sires [Milk Genomics Initiative (MGI) sires, or simply MGI sires] and 1,813 cows (MGI cows) were genotyped with a 50k SNP chip designed by CRV (Arnhem, the Netherlands) with the Illumina Infinium assay (Illumina Inc., San Diego, CA). Second, MGI cows were imputed from 50k to 777k SNP genotypes using a reference population of 1,333 animals including the 55 MGI sires genotyped with the 777k SNP chip (Illumina). As described by Duchemin et al. (2014), the imputation resulted in 1,736 MGI cows imputed to 777k SNP genotypes. From these 1,736 cows, some were removed from the data: 12 cows because of pedigree inconsistencies, and subsequently, 3 cows that did not meet the criterion of a minimum of 3 cows sampled per herd. After these exclusions, 777k SNP genotypes were available for 1,721 MGI cows. For BTA17, the target of the present study, the data consisted of a total of 22,240 imputed SNP genotypes for each of the 1,721 MGI cows.

Third, the 55 MGI sires and 34 influential ancestors (MGI grand-sires) of MGI cows were sequenced with the HiSeq 2000 Sequencing System (Illumina). All downstream analyses were performed according to the protocols described by Daetwyler et al. (2014). Multi-sample variant calling was done using the Unified Genotyper implemented in GATK, following procedures explained by Daetwyler et al. (2014). The resulting

raw VCF files were filtered for exclusion of duplicates, resulting in 854,779 called sites on BTA17 for each of the 89 sequenced MGI sires and grand-sires.

Fourth, 450 sequences from Holstein cows and bulls were available from the 5th Run of the 1000 Bull Genome Consortium (**RUN5**; Daetwyler et al., 2014). These 450 RUN5 sequenced animals included sequences from 44 out of the 55 MGI sires. All positions of the variants on sequences were aligned to the bovine genome assembly UMD3.1 (Zimin et al., 2009). The SNP and indels at the same base-pair positions were excluded because of alignment and sequencing problems. For further details on alignment, variant calling, and filtering, see Daetwyler et al. (2014). For BTA17, a total of 1,157,678 sites were called for each of the 450 sequenced animals of RUN5.

Imputation to Sequence Level

Imputation of BTA17 to sequence level consisted of 2 distinct steps. First, we created a reference population combining sequenced animals from MGI and RUN5. This reference population consisted of imputing the 45 sequenced MGI sires and grandsires, which were not present in RUN5, to equalize their number of sites with the 450 RUN5 sequences. Comparisons between called sites from the 45 MGI and the 450 RUN5 sequences showed the following differences: (1) 359,053 sites were called in MGI, but not in RUN5 sequences; (2) more sites were called in RUN5 than in MGI sequences (302,899 called sites more); and (3) the overlap between MGI and RUN5 sequences was equal to 495,726 called sites. When accounting for the total difference in sites between MGI and RUN5, a total of 661,952 sites were imputed based on the 450 RUN5 sequences. Imputation was done using Beagle version 4.0 (Browning and Browning, 2007). After imputation, the 45 MGI sequences were combined with the 450 RUN5 sequences, resulting in a reference population of 495 sequenced Holstein cows and bulls with 1,157,678 called sites for BTA17 (Supplemental Figure S1A; <https://doi.org/10.3168/jds.2017-12965>).

Second, we used Beagle version 4.0 (Browning and Browning, 2007) to impute the 1,721 MGI cows from imputed 777k SNP to sequence level for all sites on BTA17 (Supplemental Figure S1B; <https://doi.org/10.3168/jds.2017-12965>). This imputation can only occur after running a technical check for inconsistencies between the imputed 777k SNP and the sequenced reference population. Inconsistencies were checked using the Conform-gt software (<https://faculty.washington.edu/browning/conform-gt.html>), and resulted in the following: (1) 383 SNP had strand problems, and (2) 1,481 SNP showed different positions between the 777k SNP

and the reference population. All inconsistencies were set to missing and imputed to sequence level based on the reference population composed of 495 sequenced Holstein cows and bulls. The accuracy of imputation for each variant was provided by Beagle as the biallelic squared correlation coefficient (**AR2**). According to Browning and Browning (2007), AR2 is an upper bound on the loss of power when imputed posterior genotype probabilities are used in place of the true genotypes for a marker, and the AR2 interpretation does not depend on allele frequency. Following the same approach as Duchemin et al. (2016), who identified low minor allele frequency (**MAF**) variants associated with noncoagulation of milk in Swedish Red cows, only polymorphic variants with an AR2 ≥ 0.8 were retained for the remaining analyses (Supplemental Figure S1B; <https://doi.org/10.3168/jds.2017-12965>).

Fine-Mapping of BTA17 with Imputed Sequences

The fine-mapping of BTA17 with imputed sequences was performed separately for the winter and summer milk samples. For each season, animals were included in the analyses if both phenotypic and genotypic data were available. Therefore, a total of 1,640 animals with phenotypes and genotypes were available for winter milk, and a total of 1,581 animals with phenotypes and genotypes were available for summer milk samples.

The fine-mapping of BTA17 with imputed sequences was performed in Asreml 4.1 (Gilmour et al., 2015), and consisted of 2 steps. For the first step, we ran single-variant analyses for each of the 6 individual FA and C6:0–C14:0 with all polymorphic variants imputed with an AR2 ≥ 0.8 , using the following animal model:

$$y_{ijkl} = \mu + b_1 \times \text{dim} + b_2 \times e^{-0.05 \times \text{dim}} + b_3 \times \text{afc} + b_4 \times \text{afc}^2 + \text{season}_i + \text{scode}_j + b_5 \times \text{variant} + \text{herd}_k + a_l + e_{ijkl}, \quad [1]$$

where y_{ijkl} is the phenotype; b_1 and b_2 are the regression coefficients regarding *dim*; *dim* is the fixed effect of DIM modeled by a Wilmink's curve (Wilmink, 1987); b_3 and b_4 are the regression coefficients regarding *afc*; *afc* is the fixed effect of age at first calving; *season_i* is the fixed effect of calving season (June–August 2004, September–November 2004, or December 2004–February 2005); *scode_j* is the fixed effect accounting for genetic differences between groups of proven bull daughters and young bull daughters; b_5 is the regression coefficient representing the allele substitution effect of *variant*; *variant* is the covariate representing 0, 1, or 2 copies of an allele; *herd_k* is the random effect of herd assumed to be distributed as $N(0, \mathbf{I}\sigma_{\text{herd}}^2)$, where $\mathbf{I}$ is

the identity matrix and σ_{herd}^2 is the herd variance; a_l is the random additive genetic effect of the l th animal assumed to be distributed as $N(0, \mathbf{A}\sigma_a^2)$, where $\mathbf{A}$ is the additive relationship matrix based on 12,548 animals and σ_a^2 is the additive genetic variance; and e_{ijkl} is the random residual effect assumed to be distributed as $N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{I}$ is the identity matrix and σ_e^2 is the residual variance.

Variance components were estimated based on model [1] before the inclusion of information on variants, and these estimates were subsequently fixed in model [1]. The strongest association found in the first step was named TagSNP1. For the second step, TagSNP1 was added as a fixed effect in model [1], and single-variant analyses were re-run for each FA and for C6:0–C14:0 with all polymorphic variants imputed with an $AR2 \geq 0.8$.

Heritability estimates for each of the 6 individual FA and C6:0–C14:0 were calculated from univariate analyses based on model [1], without the effects of variants as follows: $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$. These estimates were calculated for genotyped animals only and separately for winter and summer milk samples.

The variants' variance ($\sigma_{variant}^2$) was computed as 2 times the major allele frequency times MAF times the squared allele substitution effect. The genetic variance explained by a variant was computed as the division between $\sigma_{variant}^2 / \sigma_a^2$, and the phenotypic variance explained by a variant was computed as the division between $\sigma_{variant}^2 / \sigma_p^2$.

The genetic correlation between C6:0–C14:0 measured in winter and in summer milk samples was calculated as follows: $r = \frac{\sigma_{Tw, Ts}}{\sqrt{\sigma_{Tw}^2 \times \sigma_{Ts}^2}}$, where $\sigma_{Tw, Ts}$ = covariance between C6:0–C14:0 measured in winter and in summer milk samples; σ_{Tw}^2 = variance of C6:0–C14:0 in winter samples, and σ_{Ts}^2 = variance of C6:0–C14:0 in summer samples.

Manhattan plots illustrating significance of associations were produced in R (R Core Team, 2015). In addition, linkage disequilibrium (**LD**) between TagSNP1 and all polymorphic SNP imputed with an $AR2 \geq 0.8$ was calculated using Plink (Purcell et al., 2007; Supplemental Figure S3, <https://doi.org/10.3168/jds.2017-12965>).

Candidate Genes

Candidate genes were reported based on National Center for Biotechnology Information (**NCBI**; accessed

October 2016 at <http://ncbi.nlm.nih.gov>) and on Ensembl (version 86, accessed October 2016 at <http://www.ensembl.org>; Yates et al., 2016) public databases. We chose an automated method to analyze the possible coding consequences of candidate genes. For this, we used the online tool variant effect predictor (**VEP**; McLaren et al., 2016) available through Ensembl. Transcripts analyzed by VEP were based on the Ensembl transcripts database only. This tool determines the coding consequences of SNP, insertions, deletions, copy number variants, and structural variants on genes, transcripts, proteins, or regulatory regions.

RESULTS

Descriptive Statistics

Phenotypic means and heritability estimates of the 6 individual FA and C6:0–C14:0 are presented in Table 1. In both samples, C14:0 was the most abundant individual FA. For C6:0–C14:0, larger fractions of total fat were observed in winter (22.35%) as compared with summer milk (21.31%). Heritability estimates were higher in winter in comparison with summer milk samples, especially for C8:0, C10:0, C12:0, and C6:0–C14:0. Phenotypic means and heritability estimates of the 6 individual FA in winter and summer milk samples have been discussed in detail by Duchemin et al. (2013).

Imputation

To enable combining the 45 MGI with the 450 RUN5 sequences into one reference population, the 661,952 sites that differed between these 2 sets of sequences were imputed based on the 450 RUN5 sequences. The average accuracy of imputation for these 661,952 sites was equal to 0.97 (Supplemental Figure S1A; <https://doi.org/10.3168/jds.2017-12965>). Based on the reference population of 495 sequenced Holstein cows and bulls, all 1,157,678 sites on BTA17 were imputed for the 1,721 MGI cows. Supplemental Figure S1B (<https://doi.org/10.3168/jds.2017-12965>) shows that 58.6% of imputed sites in our data set were monomorphic sites, hence excluded from fine-mapping. The remaining 41.4% imputed sites were polymorphic variants, of which 356,044 (30.8% of the total number of sites) were imputed with an average accuracy equal to 0.96, and were used for fine-mapping.

Fine-Mapping of BTA17 with Imputed Sequences

The fine-mapping of BTA17 with imputed sequences consisted of separate association analyses for winter and summer milk samples. Fatty acid composition of

winter and summer milk samples can be largely considered as genetically the same trait (Duchemin et al., 2013). For each season, we analyzed 6 individual FA and C6:0–C14:0 as phenotypes, in combination with 356,044 imputed sequence variants (BTA17) as genotypes. For winter and summer milk samples, significant associations with the 6 FA and C6:0–C14:0 were found at a similar genomic region (Supplemental Figures S2A and S2B; <https://doi.org/10.3168/jds.2017-12965>). We chose C8:0 to describe our findings because for this genomic region the strongest associations were found with this trait in both samples. For C8:0 in winter and summer milk, an association was significant at $-\log_{10}(P\text{-value}) = 3.63$ (Figures 1A–1C; Supplemental Figures S2A–S2D, <https://doi.org/10.3168/jds.2017-12965>). This threshold was defined by Bouwman et al. (2012) and corresponds to the genome-wide significance level based on the 50k SNP genotypes at a false discovery rate of 5%. For C8:0, we identified 1,182 significant associations in winter milk, and 1,737 significant associations in summer milk samples on BTA17. A total of 323 significant associations in summer milk overlap with associations in winter milk samples. In addition, 302 out of these 323 associations are in the QTL region located between 29 and 34 Mbp (Figures 1A and 1C).

For C8:0 in summer milk samples, 646 out of 1,737 significant associations were located in the QTL region (Figure 1A). A set of 8 SNP in perfect LD represented the strongest associations with C8:0 in summer milk samples at $-\log_{10}(P\text{-value}) = 7.82$ (Figures 1A and 1C; Supplemental Figure S3, <https://doi.org/10.3168/jds.2017-12965>). One of these 8 SNP, *rs110127535*, was defined as TagSNP1 (Figures 1A and 1C). The TagSNP1 segregates at MAF of 0.44, and was imputed with an $AR^2 = 0.98$. Estimated effects of TagSNP1 regarding the 6 individual FA and C6:0–C14:0 are shown in Table 2. TagSNP1 was added as fixed effect in model 1, and associations were re-analyzed. After adjusting for TagSNP1, no significant associations were

found neither for the 6 individual FA nor for C6:0–C14:0 in winter and in summer milk samples (Figure 1B, Supplemental Figures S2C–S2D; <https://doi.org/10.3168/jds.2017-12965>).

Candidate Genes

The QTL region located between 29 and 34 Mbp on BTA17 contains 30 annotated genes based on the NCBI database, whereas this region contains 22 annotated genes based on the Ensembl database. The lists of annotated genes based on NCBI and on Ensembl databases are shown in Figure 1D. Despite the 22 annotated genes based on the Ensembl database, VEP only predicted the coding consequences of 13 genes (i.e., 13 genes that were annotated and validated). The 9 other genes were not used in VEP predictions for the following reasons: (1) no significant associations were found in *HSPA4L* and *ANKRD50* genes; (2) *ENSBTAG00000004743*, *ENSBTAG000000045423*, *ENSBTAG000000040242*, and *ENSBTAG000000005435* are not validated genes yet; (3) *COMMD6* is not a validated gene although assigned to BTA17 in the Ensembl database. Furthermore, this gene is assigned to BTA12 in the NCBI database; and (4) *U6* gene has 2 distinct assigned locations on BTA17 in the Ensembl database.

Therefore, we report here the results of the 13 validated genes analyzed using VEP (Table 3). These 13 genes are chromosome 4 open reading frame 33 (*C4orf33*) gene; sodium channel and clathrin linker 1 (*SLCT1*) gene; jade family PHD finger 1 (*JADE1*) gene; progesterone receptor membrane component 2 (*PGRMC2*) gene; La ribonucleoprotein domain family, member 1B (*LARP1B*) gene; U2 spliceosomal RNA (*U2*) gene; the abhydrolase domain containing 18 (*ABHD18*, former *C4orf29*) gene; small nucleolar RNA SNORA42/SNO-RA80 family (*SNORA42*) gene; major facilitator superfamily domain containing 8 (*MFSD8*) gene; polo-like kinase 4 (*PLK4*) gene; solute carrier family 25 member

Table 1. Phenotypic means (SD) and heritability estimates (h^2)¹ for individual fatty acids and the indicator of de novo synthesis C6:0–C14:0 based on 1,640 winter milk and 1,581 summer milk samples

Phenotype (% wt/wt)	Winter		Summer	
	Mean (SD)	h^2	Mean (SD)	h^2
C4:0	3.51 (0.27)	0.47	3.52 (0.35)	0.41
C6:0	2.23 (0.16)	0.46	2.17 (0.21)	0.39
C8:0	1.36 (0.14)	0.59	1.32 (0.17)	0.35
C10:0	3.02 (0.43)	0.73	2.87 (0.45)	0.48
C12:0	4.12 (0.70)	0.62	3.79 (0.72)	0.48
C14:0	11.62 (0.92)	0.62	11.16 (1.05)	0.54
C6:0–C14:0	22.35 (2.01)	0.71	21.31 (2.26)	0.50

¹ $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$, where σ_a^2 is the additive genetic variance and σ_e^2 is the residual variance; SE between 0.01 and 0.12 for winter samples, and between 0.02 and 0.08 for summer samples.

Figure 1. Fine-mapping of BTA17 showing associations with C8:0 on summer milk samples identified with variants imputed at accuracy of imputation (AR2) ≥ 0.8 presented as follows: (A) overlaid with imputed 777,000 SNP genotypes. Dotted line shows the genome-wide significance level based on 50,000 SNP genotypes at a false discovery rate of 0.05 [$-\log_{10}(P\text{-value}) = 3.63$]. Vertical lines define the QTL region. TagSNP1 is the SNP with the highest significance. (B) After fitting TagSNP1. (C) Magnification of the QTL region located between 29 and 34 Mbp. (D) Genes listed based on National Center for Biotechnology Information (NCBI) and Ensembl databases. For the NCBI database, validated genes are in red, labeled genes not yet validated are in black, and genes not labeled but present in the region are in light gray. For Ensembl database, validated genes based on the Ensembl transcripts database are in blue, and genes not yet validated are in black. Color version available online.

31 (*SLC25A31*) gene; inturnd planar cell polarity protein (*INTU*) gene; and FAT atypical cadherin 4 (*FAT4*) gene. In addition, VEP was used to assess the coding

consequences of the 646 significant variants including TagSNP1, and these variants were distributed as indicated in Figure 2. Two variants were predicted to cause

Table 2. Estimated effects of TagSNP1 (CC, CG, or GG genotypes; SE in parentheses), significance of TagSNP1, and genetic¹ and phenotypic² variances explained by TagSNP1 for each of the 6 individual fatty acids and C6:0–C14:0 in summer milk samples

Item	Summer sample			-Log ₁₀ (<i>P</i> -value)	$\frac{\sigma_{TagSNP1}^2}{\sigma_a^2}$	$\frac{\sigma_{TagSNP1}^2}{\sigma_p^2}$
	CC	CG	GG			
C4:0	0.000 (0.000)	-0.047 (0.012)	-0.083 (0.014)	0.33	0.2%	0.1%
C6:0	0.000 (0.000)	-0.046 (0.014)	-0.091 (0.018)	6.48	7.6%	2.4%
C8:0	0.000 (0.000)	-0.047 (0.012)	-0.083 (0.014)	7.82	8.9%	2.9%
C10:0	0.000 (0.000)	-0.102 (0.032)	-0.206 (0.039)	6.84	5.5%	2.4%
C12:0	0.000 (0.000)	-0.145 (0.044)	-0.235 (0.055)	4.48	4.2%	1.2%
C14:0	0.000 (0.000)	-0.163 (0.066)	-0.278 (0.083)	3.09	2.6%	0.8%
C6:0-C14:0	0.000 (0.000)	-0.520 (0.148)	-0.920 (0.185)	5.91	5.4%	1.9%

¹ $\sigma_{\text{TagSNP1}}^2 / \sigma_a^2$ = genetic variance explained by TagSNP1, where $\sigma_{\text{TagSNP1}}^2$ is the variance of TagSNP1 and σ_a^2 is the genetic variance. TagSNP1 is the SNP with the highest significance.

$\sigma_{\text{TagSNP1}}^2 / \sigma_p^2$ = phenotypic variance explained by TagSNP1, where $\sigma_p^2 = \sigma_a^2 + \sigma_{\text{herd}}^2 + \sigma_e^2$, and σ_{herd}^2 is the herd variance, and σ_e^2 is the residual variance.

Figure 2. Distribution of the 646 significant variants according to their functions and their coding consequences. UTR = untranslated region. Color version available online.

AA changes, one in *PLK4* [$-\log_{10}(P\text{-value}) = 4.39$] and one in *SLC25A31* [$-\log_{10}(P\text{-value}) = 3.83$].

Regarding the 8 SNP that showed the strongest association with C8:0 in summer milk samples, 2 SNP were intergenic, 3 SNP were upstream gene variants of the *LARP1B* gene, and 3 SNP were intron variants in the *LARP1B* gene. In addition, a splice-region variant (*rs110862734*) was located at 29.94 Mbp in the *LARP1B* gene. This splice-region variant is located in intron 19–20, at 5 bp from exon 19, and at 1 bp apart from a highly conserved sequence. This splice-region variant was associated with C8:0 in summer milk at $-\log_{10}(P\text{-value}) = 6.31$, and its LD with the 8 most significantly associated SNP was equal to 0.92.

DISCUSSION

A QTL associated with individual milk FA on BTA17 was discovered by Bouwman et al. (2012). Duchemin et al. (2014) fine-mapped this QTL to a genomic region located between 29 and 34 Mbp on BTA17. Furthermore, Duchemin et al. (2014) characterized this QTL region with haplotype analyses and concluded that one QTL was likely to influence multiple de novo synthesized milk FA at this location on BTA17. In the present work, we used the same 6 individual FA studied by Duchemin et al. (2014), and re-analyzed BTA17 with imputed sequences. Although phenotypic means and heritability estimates of the 6 individual FA are

Table 3. Details about candidate genes identified in the QTL region located between 29 and 34 Mbp on BTA17

Gene	Identifier	Location	No. of variants
<i>C4orf33</i>	ENSBTAG00000044159	Chr17:29105309–29116531	2
<i>SCLT1</i>	ENSBTAG00000013611	Chr17:29190572–29354595	32
<i>JADE1</i>	ENSBTAG00000017493	Chr17:29368416–29421827	5
<i>PGRMC2</i>	ENSBTAG00000010843	Chr17:29872406–29890867	8
<i>LARP1B</i>	ENSBTAG00000012135	Chr17:29938416–30073786	83
<i>U2</i>	ENSBTAG00000043806	Chr17:30096344–30096524	4
<i>ABHD18</i>	ENSBTAG00000010630	Chr17:30106834–30143868	13 ¹
<i>SNORA42</i>	ENSBTAG00000042423	Chr17:30118538–30118671	2
<i>MFSD8</i>	ENSBTAG00000044058	Chr17:30144105–30181831	13 ²
<i>PLK4</i>	ENSBTAG00000039552	Chr17:30185756–30202777	4
<i>SLC25A31</i>	ENSBTAG00000012826	Chr17:30291318–30319495	2
<i>INTU</i>	ENSBTAG00000012824	Chr17:30324842–30404702	3
<i>FAT4</i>	ENSBTAG00000003345	Chr17:32712712–32889849	119

¹Two out of 13 variants in the *ABHD18* gene overlap with variants in the *SNORA42* gene. These are 2 intron variants in the *ABHD18* gene corresponding to 1 downstream and 1 upstream variant in the *SNORA42* gene.

²One out of 13 variants in the *MFSD8* gene overlaps with a variant in the *PLK4* gene. This variant is a downstream variant in the *MFSD8* gene corresponding to a downstream variant in the *PLK4* gene.

comparable with Stoop et al. (2008) and Duchemin et al. (2013), these heritability estimates are higher than previous published studies (e.g., Soyeurt et al., 2008). These higher values might be caused because previous studies either used other methods to measure milk FA (e.g., infrared predicted versus GC) or studied different breeds.

A novelty compared with Duchemin et al. (2014) is the fine-mapping of the indicator of de novo synthesis, namely C6:0–C14:0. Our goal was to characterize this QTL region with more variants, ideally all, including causal variants, and recent information on candidate genes. In short, new results were obtained by the fine-mapping of BTA17 with imputed sequences, besides confirming previously published results. These new results are (1) the QTL region influences 2 extra de novo synthesized milk FA at a similar location on BTA17; (2) 8 SNP in perfect LD are the most significant SNP associated with this QTL region; and (3) the identification of potential candidate gene for this QTL region.

Fine-Mapping of BTA17 with Imputed Sequences

The BTA17 QTL region influences C6:0, C8:0, and C10:0 individual milk FA but not C4:0, congruent with Duchemin et al. (2014). With imputed sequences, we found that this QTL region also influences C12:0 and C14:0, results not previously found by Duchemin et al. (2014) with imputed 777k SNP genotypes. Although the associations were weaker for C12:0 and C14:0, their associations overlap with the associations of the other FA in the QTL region for both winter and summer milk samples (Supplemental Figures S2A and S2B; <https://doi.org/10.3168/jds.2017-12965>). Govignon-Gion et al. (2014) found a QTL region on BTA17 associated with multiple milk FA when running GWAS with imputed 500k SNP genotypes for 3 different breeds of dairy cattle. Their study mainly differs from our study as follows: (1) their studied FA were measured by mid-infrared spectrometry, whereas ours were measured by gas chromatography; (2) their strongest association was identified with C4:0, whereas associations with C4:0 were not significant in our study; and (3) their QTL region does not overlap with ours. Based on these differences, it is likely that the QTL regions in the 2 studies are in fact distinct QTL regions.

With imputed sequences, the (fine) mapping resolution of the QTL region on BTA17 was substantially improved in the present work. Previously for C8:0 in summer milk samples, Bouwman et al. (2012) identified 10 SNP with 50k SNP genotypes, and Duchemin et al. (2014) identified 83 SNP with imputed 777k SNP genotypes. No overlap of associations was observed between the 50k and the imputed 777k SNP genotypes. From

imputed 777k SNP genotypes to sequence level, more than a thousand variants became significantly associated with multiple individual FA and with C6:0–C14:0. Specifically for C8:0 in summer milk samples, 70 SNP identified by the imputed 777k SNP genotypes were found among the 646 significant associations with imputed sequences. Among these 646 associations, 8 SNP in perfect LD represent the strongest associations with C8:0 in summer milk samples. These 8 SNP include the strongest associated SNP (*rs109290136*) identified by Duchemin et al. (2014) with imputed 777k SNP genotypes.

The significance of the top SNP in Duchemin et al. (2014) was higher [$-\log_{10}(P\text{-value}) = 7.93$] than the significance of the 8 SNP in perfect LD [$-\log_{10}(P\text{-value}) = 7.82$] identified with imputed sequences. It is important to realize that Duchemin et al. (2014) modeled each SNP effect as a class variable, for which each predicted SNP effect was estimated with 2 df. In the present study, each variant was modeled as a co-variable, hence an allele-substitution effect for each variant was estimated. For comparison purposes, we re-modeled the data used by Duchemin et al. (2014) with a co-variable approach (Figure 1A). As a result, the significance of the top SNP in Duchemin et al. (2014) was identical to the significance of the 8 SNP in perfect LD [$-\log_{10}(P\text{-value}) = 7.82$] identified with imputed sequences.

The QTL region is associated with multiple individual milk FA at a similar location on BTA17. These multiple FA are FA synthesized de novo in the mammary gland of a cow as previously described by Duchemin et al. (2014). In this study, we fine-mapped BTA17 not only for individual FA but also for the indicator of de novo milk FA synthesis C6:0–C14:0 based on imputed sequences (Supplemental Figures S2A–S2D; <https://doi.org/10.3168/jds.2017-12965>). In summer milk, associations with C6:0–C14:0 were more pronounced than in the winter samples. The C6:0–C14:0 in summer milk samples is strongly genetically correlated with C6:0–C14:0 in winter milk samples ($r = 0.92$) suggesting that C6:0–C14:0 in winter and summer milk samples is influenced by mainly the same genes. The TagSNP1 explained 5.4% of the genetic variance and 1.9% of the phenotypic variance in C6:0–C14:0 in summer milk. After correcting for TagSNP1, no remaining genetic variation was found associated with C6:0–C14:0 in winter and summer milk samples. These findings suggest that one polymorphism on BTA17 underlying this QTL region influences multiple de novo synthesized milk FA.

Candidate Genes

The VEP online tool available through Ensembl predicted the coding consequences for 13 out of 22

potential candidate genes present in the QTL region located between 29 and 34 MBP on BTA17. Among these 13 genes, we prioritized a gene if it contained variants with significant associations, followed by if these significant associations could potentially alter the encoded protein. Hence, the 13 candidate genes were narrowed to 3: *LARP1B*, *PKL4*, and *SLC25A31* genes. According to Ron and Weller (2007) and Weller (2009), a gene can be declared as a candidate gene if at least 1 out of the 4 following characteristics is present: (1) the gene has a known physiological role in the trait of interest, (2) the gene affects the trait of interest based on knockout, mutation, or transgenic studies in other species, (3) the gene is preferentially expressed in organs related to the trait of interest, and (4) the gene is preferentially expressed during developmental stages related to the trait of interest. Because no knockout, mutations, or transgenic studies in other species have been found for *LARP1B*, *PKL4* and *SLC25A31* genes, we investigated their gene ontology (GO) terms, their expression in the mammary gland, and their expression during developmental stages of lactation.

Six of the 8 strongest associations are located in the *LARP1B* gene. According to genecards (<http://www.genecards.org/>), the *LARP1B* gene encodes a protein containing domains found in the La related protein of *Drosophila melanogaster*. According to Stavraka and Blagden (2015), La related proteins family 1 (*LARP1*) genes have 2 paralogues in humans: *LARP1A* and *LARP1B*, which are similar (60% homology and 73% of positivity). The *LARP1A* (or simply *LARP1*) is positioned at chromosome 5q34, encoding an 1,096 AA protein. The *LARP1B* (or *LARP2*) is positioned at chromosome 4q28, encoding an 914 AA protein. In humans, Burrows et al. (2010) showed that *LARP1A* is more expressed than *LARP1B*, and for that matter, *LARP1B* has been less studied. According to Uniprot (www.uniprot.org; accessed March 30, 2017), the GO term regarding the molecular function of the *LARP1B* gene in humans is poly(A) RNA binding, which is also the function of *LARP1A*. Bousquet-Antonelli and Deragon (2009) suggested that members of the same family are functional homologs, share a common molecular mode of action on different RNA baits, or both. Glisovic et al. (2008) concluded that RNA binding proteins are present throughout the RNA biology (from transcription, polyadenylation to RNA transport and translation, to name some processes). With interconnections within and between their diverse processes, RNA binding proteins form networks of events that are essential for posttranscriptional regulation, and in case of malfunctioning, these network events can lead to phenotypic differences, such as diseases. In this regard, the *LARP1B* gene might be involved in post-

transcriptional gene regulation. In mammalian cells, Tcherkezian et al. (2014) found that the *LARP1A* gene associates with the mTOR complex 1 (*mTORC1*) and is required for global protein synthesis as well as cell growth and proliferation. This implicates the *LARP1A* gene (as well as the *LARP1B* gene) as an important regulator of cell growth and proliferation. In the present study, one association in the *LARP1B* gene was identified as a splice-region variant (i.e., located in the first base pair of an intron). Because it is not located in the splice donor/acceptor site, it is hard to predict the effect of this splice-region variant on splicing. Although no further information is available on the possible transcripts generated by the *LARP1B* gene, this gene is highly expressed in bovine mammary tissue (Bionaz et al., 2012), and it is expressed in all stages of lactation in humans (Lemay et al., 2013). Yet, the *LARP1B* gene has not been associated with milk FA composition or milk fat synthesis.

Among other potential candidate genes in our QTL region, the polo like kinase 4 (*PLK4*) and the solute carrier family 25 member 31 (*SLC25A31*) genes might be of interest. The preferred name of the *PLK4* gene is serine/threonine-protein kinase *PLK4*. The *PLK4* gene ontology (provided by NCBI) suggests that this gene is involved in ATP binding and protein serine/threonine kinase activity regarding the centriole replication, protein phosphorylation, trophoblast giant cell differentiation, or a combination. The preferred name of the *SLC25A31* gene is ADP/ATP translocase 4. The *SLC25A31* gene ontology (provided by NCBI) suggests that this gene is involved in the transmembrane transport of ADP/ATP in the mitochondria. Each of these genes contains 1 missense variant. A missense variant changes the codon of one AA into the codon of another AA, which can affect the stability and aggregation of the encoded protein (DePristo et al., 2005). The sorting intolerant from tolerant (SIFT) algorithm implemented in VEP scored 0.98 for the *PLK4* gene and scored 0.58 for the *SLC25A31* gene. Both scores indicate that these missense variants are well tolerated and have low consequences at the protein level. Regarding expression studies, the *PLK4* gene is differentially expressed in bovine mammary tissue 15 d before parturition and at around 60 d after parturition (Bionaz et al., 2012), and it is expressed in all stages of lactation in humans (Lemay et al., 2013). Conversely, the *SLC25A31* gene is neither expressed in bovine mammary tissue (Bionaz et al., 2012) nor during all stages of lactation in humans (Lemay et al., 2013). Although the *SLC25A31* gene can no longer be a candidate gene, up to this moment, the *PLK4* gene has neither been associated with milk FA composition nor with milk fat synthesis yet.

Previously, the candidate gene identified by Duchemin et al. (2014) was the *PGRMC2* gene. The *PGRMC2* gene is still among the genes associated with the QTL region that influences multiple FA on BTA17 (Table 3). The *PGRMC2* gene was assigned as the most likely candidate gene previously because it was the closest annotated gene to the strongest association found by Duchemin et al. (2014). In addition, *LOC515517* was even closer to the strongest association on BTA17. However, because of limited annotation available for BTA17 at that time, *LOC515517* was identified as a suggestive candidate whereas *PGRMC2* was suggested as the primary candidate gene. Since then, *LOC515517* has been annotated in the cattle genome as the *LARP1B* gene. Therefore, in the present study, the *LARP1B* gene is proposed as the primary candidate gene underlying the QTL region on BTA17 associated with de novo synthesized FA. This demonstrates that interpretation of (fine) mapping results depends on the annotation of the bovine genome, which is still progressing.

Reviewing the results for this QTL region suggest that (1) a panel of 50k SNP genotypes is efficient in identifying QTL correctly, (2) a panel of imputed 777k SNP genotypes is efficient in identifying the most significant association with the phenotypes, and (3) imputed sequences are efficient in identifying the candidate gene(s) underlying a QTL.

CONCLUSIONS

The fine-mapping of BTA17 with imputed sequences revealed that a QTL region influencing C6:0, C8:0, and C10:0 individual FA also influences C12:0 and C14:0 in winter and summer milk samples. The fine-mapping of BTA17 with imputed sequences identified thousands of significant associations with C6:0, C8:0, C10:0, C12:0, C14:0, and C6:0–C14:0, resulting in improved resolution of the QTL region in comparison with previous studies. Our strongest associations were found with C8:0 in summer milk samples, and consisted of 8 SNP in perfect LD. These 8 SNP in perfect LD explain 5.4% of the genetic variance and 1.9% of the phenotypic variance in C6:0–C14:0. In this study, the number of candidate genes in this QTL region was reduced from 29 to 13. Within these 13 candidate genes, 6 out of the 8 strongest SNP in perfect LD were found in the *LARP1B* gene. It is possible that the *LARP1B* gene is involved in posttranscriptional gene regulation. The *LARP1B* gene is highly expressed in bovine mammary tissue, and during all phases of lactation in humans. Another gene of interest might be the *PLK4* gene. None of these genes have been associated with milk FA composition or milk fat synthesis yet.

ACKNOWLEDGMENTS

This study is part of the Dutch Milk genomics Initiative, funded by Wageningen University (WUR; Wageningen, the Netherlands), the Dutch Dairy Association (NZO; Den Haag, the Netherlands), Cooperative Cattle Improvement Organization (CRV; Arnhem, the Netherlands), and the Dutch Technology Foundation (STW; Utrecht, the Netherlands). We thank Jérémie Vandenplas (WUR) for his help with coding for parallel computing, and Juan Manuel Herrero-Medrano (WUR) for his help with variant calling. The first author currently benefits from a joint-grant from the European Commission (Brussels, Belgium; within the framework of the Erasmus-Mundus joint doctorate “EGS-ABG”) and Breed4food (a public-private partnership in the domain of animal breeding and genomics, and CRV).

REFERENCES

- Bionaz, M., W. Hurley, and J. Loo. 2012. Milk protein synthesis in the lactating mammary gland: Insights from transcriptomics analyses. INTECH Open Access Publisher, Rijeka, Croatia.
- Bousquet-Antonelli, C., and J. M. Deragon. 2009. A comprehensive analysis of the La-motif protein superfamily. *RNA* 15:750–764.
- Bouwman, A., M. H. P. W. Visker, J. A. M. van Arendonk, and H. Bovenhuis. 2012. Genomic regions associated with bovine milk fatty acids in both summer and winter milk samples. *BMC Genet.* 13:93.
- Browning, S. R., and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084–1097.
- Buitenhuis, B., L. L. G. 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.
- Burrows, C., N. Abd Latip, S. J. Lam, L. Carpenter, K. Sawicka, G. Tzolovsky, H. Gabra, M. Bushell, D. M. Glover, A. E. Willis, and S. P. Blagden. 2010. The RNA binding protein Larp1 regulates cell division, apoptosis and cell migration. *Nucleic Acids Res.* 38:5542–5553.
- Calder, P. C. 2015. Functional roles of fatty acids and their effects on human health. *J. Parenter. Enteral Nutr.* 39:18S–32S.
- Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R. F. Brondum, X. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerre, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chamberlain, C. Anderson, C. P. VanTassell, I. Hulsege, M. E. Goddard, B. Guldbrandtsen, M. S. Lund, R. F. Veerkamp, D. A. Boichard, R. Fries, and B. J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865.
- DePristo, M. A., D. M. Weinreich, and D. K. Hartl. 2005. Missense meanderings in sequence space: A biophysical view of protein evolution. *Nat. Rev. Genet.* 6:678–687.
- Duchemin, S. I., H. Bovenhuis, W. M. Stoop, A. C. Bouwman, J. A. M. van Arendonk, and M. H. P. W. Visker. 2013. Genetic correlation between composition of bovine milk fat in winter and summer, and DGAT1 and SCD1 by season interactions. *J. Dairy Sci.* 96:592–604.
- Duchemin, S. I., M. Glantz, D.-J. de Koning, M. Paulsson, and W. F. Fikse. 2016. Identification of QTL on chromosome 18 associated with non-coagulating milk in Swedish Red cows. *Front. Genet.* 7:57 <https://doi.org/10.3389/fgene.2016.00057>.

- Duchemin, S. I., M. H. P. W. Visker, J. A. M. Van Arendonk, and H. Bovenhuis. 2014. A quantitative trait locus on *Bos taurus* autosome 17 explains a large proportion of the genetic variation in de novo synthesized milk fatty acids. *J. Dairy Sci.* 97:7276–7285.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2015. ASReml User Guide Release 4.1 Functional Specification. VSN International Ltd., Hemel Hempstead, UK.
- Glisovic, T., J. L. Bachorik, J. Yong, and G. Dreyfuss. 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 582:1977–1986.
- Govignon-Gion, A., S. Fritz, H. Larroque, M. Brochard, C. Chantry, F. Lahalle, and D. Boichard. 2014. QTL detection for milk fatty acids in French dairy cattle. In 10th World Congress on Genetics Applied to Livestock Production. ASAS, Champaign, IL.
- Lemay, D. G., O. A. Ballard, M. A. Hughes, A. L. Morrow, N. D. Horseman, and L. A. Nommsen-Rivers. 2013. RNA sequencing of the human milk fat layer transcriptome reveals distinct gene expression profiles at three stages of lactation. *PLoS One* 8:e67531.
- McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. The Ensembl variant effect predictor. *Genome Biol.* 17:122.
- Moioli, B., G. Contarini, A. Avalli, G. Catillo, L. Orru, G. De Matteis, G. Masoero, and F. Napolitano. 2007. Short communication: Effect of stearyl-coenzyme A desaturase polymorphism on fatty acid composition of milk. *J. Dairy Sci.* 90:3553–3558.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Ron, M., and J. I. Weller. 2007. From QTL to QTN identification in livestock—Winning by points rather than knock-out: A review. *Anim. Genet.* 38:429–439.
- Schennink, A., J. M. L. Heck, H. Bovenhuis, M. H. P. W. Visker, H. J. F. van Valenberg, and J. A. M. van Arendonk. 2008. Milk fatty acid unsaturation: Genetic parameters and effects of stearyl-CoA desaturase (*SCD1*) and acyl CoA: diacylglycerol acyltransferase 1 (*DGAT1*). *J. Dairy Sci.* 91:2135–2143.
- Schennink, A., W. M. Stoop, M. H. P. W. Visker, J. M. L. Heck, H. Bovenhuis, J. J. Van Der Poel, H. J. F. Van Valenberg, and J. A. M. Van Arendonk. 2007. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. *Anim. Genet.* 38:467–473.
- Soyeurt, H., P. Dardenne, F. Dehareng, C. Bastin, and N. Gengler. 2008. Genetic parameters of saturated and monounsaturated fatty acid content and the ratio of saturated to unsaturated fatty acids in bovine milk. *J. Dairy Sci.* 91:3611–3626.
- Stavraka, C., and S. Blagden. 2015. The La-related proteins, a family with connections to cancer. *Biomolecules* 5:2701–2722.
- Stoop, W. M., J. A. M. van Arendonk, J. M. L. Heck, H. J. F. van Valenberg, and H. Bovenhuis. 2008. Genetic parameters for major milk fatty acids and milk production traits of Dutch Holstein-Friesians. *J. Dairy Sci.* 91:385–394.
- Tcherkezian, J., M. Cargnello, Y. Romeo, E. L. Huttlin, G. Lavoie, S. P. Gygi, and P. P. Roux. 2014. Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5' TOP mRNA translation. *Genes Dev.* 28:357–371.
- Venturini, G. C., D. F. Cardoso, F. Baldi, A. C. Freitas, R. R. Aspilcueta-Borquis, D. J. Santos, and H. Tonhati. 2014. Association between single-nucleotide polymorphisms and milk production traits in buffalo. *Genet. Mol. Res.* 13:10256.
- Weller, J. I. 2009. Quantitative trait loci analysis in animals. 2nd ed. CABI, Oxfordshire, UK.
- Wilmink, J. B. M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16:335–348.
- Yates, A., W. Akanni, M. R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva, C. Cummins, P. Clapham, S. Fitzgerald, L. G. C. G. Girón, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, S. Keenan, I. Lavidas, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, M. Nuhn, A. Parker, M. Patricio, M. Pignatelli, M. Rahtz, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, E. Birney, J. Harrow, M. Muffato, E. Perry, M. Ruffier, G. Spudich, S. J. Trevanion, F. Cunningham, B. L. Aken, D. R. Zerbino, and P. Flicek. 2016. Ensembl 2016. *Nucleic Acids Res.* 44:D710–D716.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, and T. S. Sonstegard. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42.