Genome-wide association study for αS1- and αS2-casein phosphorylation in Dutch Holstein Friesian

Fang, Z. H., Bovenhuis, H., van Valenberg, H. J. F., Martin, P., Duchemin, S. I., Huppertz, T., & Visker, M. H. P. W.

Interpretive Summary

Fang

Proteins in cow's milk, particularly caseins, play an important role in human nutrition and producing dairy products, such as yogurt and cheese. These caseins are phosphorylated and interact with large amounts of calcium and phosphate. As a result, these minerals can be delivered efficiently to the neonate without damaging mammary epithelial cells. Moreover, several studies show that the phosphorylation levels of the caseins have impact on the cheese-making properties of milk. In this study, we investigated the genetic background of phosphorylation levels of α_{s1}- and α_{s2}-casein. These results can help us understand genetic control of variation in phosphorylation.

Genome-wide association study for α_{s1}- and α_{s2}-casein phosphorylation in Dutch Holstein Friesian

Z. H. Fang,*†1 H. Bovenhuis,† H. J. F. van Valenberg,‡ P. Martin,* S. I. Duchemin,# T. Huppertz§2, M. H. P. W. Visker,†3

*Génétique Animale et Biologie Intégrative, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

†Animal Breeding and Genomics, Wageningen University and Research, PO Box 338, 6700 AH Wageningen, the Netherlands

‡ Dairy Science and Technology Group, Wageningen University, PO Box 17, 6700 AA, Wageningen, the Netherlands

§ NIZO, PO Box 20, 6710 BA, Ede, the Netherlands

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

1 Current address: Animal Genomics, ETH Zürich, 8092 Zürich, Switzerland

2 Current address: FrieslandCampina, Amersfoort, The Netherlands

3 Corresponding author: Marleen Visker (marleen.visker@wur.nl)

ABSTRACT

Phosphorylation of caseins (CN) is a crucial post-translational modification that allows caseins to form colloid particles known as casein micelles. Both α_{s1}- and α_{s2}-CN show varying degrees of phosphorylation (isoforms) in cow's milk and were suggested to be more relevant for stabilizing internal micellar structure than β- and κ-CN. However, little is known about the genetic background of individual α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN (α_{s1}-CN PD and α_{s2}-CN PD) defined as the proportion of isoforms with higher degrees of phosphorylation in total α_{s1}- and α_{s2}-CN, respectively. We aimed to identify genomic regions associated with these traits using 50K SNP for 1,857 Dutch Holstein Friesian cows. A total of 10 QTL regions were identified for all studied traits on 10 *Bos taurus* autosomes (BTA1, 2, 6, 9, 11, 14, 15, 18, 24 and 28). Regions associated with multiple traits were found on BTA1, 6, 11, and 14. We showed two QTL regions on BTA1: one affects α_{s2}-CN production, and the other one harbors the *SLC37A1* gene that encodes a phosphorus antiporter and affects α_{s1}-CN PD and α_{s2}-CN PD. The QTL on BTA6 harbors the casein gene cluster and affects individual α_{s2}-CN phosphorylation isoforms. The QTL on BTA11 harbors the *PAEP* gene that encodes β-lactoglobulin (β-LG) and affects relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. The QTL on BTA14 harbors the *DGAT1* gene and affects relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. Our results suggest that effects of identified genomic regions on phosphorylation of α_{s1}-CN and α_{s2}-CN are related to changes in milk synthesis and phosphorus secretion in milk. The actual roles of *SLC37A1*, *PAEP* and *DGAT1* in α_{s1}- and α_{s2}-CN phosphorylation in Dutch Holstein Friesian require further investigation.

Key words: posttranslational modification, milk protein composition, quantitative trait loci

INTRODUCTION

Protein phosphorylation regulates nearly every aspect of cell life, including disease states, by altering the structural confirmation of proteins to either activate, deactivate or modify their function. Caseins from cow's milk are the most well-studied group of phosphoproteins. They play an important role in human nutrition and also affect manufacturing properties of dairy products (Wedholm et al., 2006; Hallén et al., 2008; Caroli et al., 2009). Phosphorylation of caseins is a crucial post-translational modification that affects the formation and stability of casein micelles as the structure of micelles partly relies on the interactions between calcium phosphate nanoclusters and phosphoserine residues of α_{s1}-, α_{s2}-, and β-casein (**CN**) (De Kruif and Holt, 2003; De Kruif et al., 2012). As a result of casein micelle formation, large amounts of calcium and phosphorus can be delivered efficiently to the neonate without damaging the mammary gland of the mother by evoking either pathological calcification or amyloidosis (Holt et al., 2013).

Although α_{s1}-, α_{s2}-, β-, and κ-CN are all phosphorylated, α_{s1}- and α_{s2}-CN are more heavily phosphorylated, possess multiple phosphoserine clusters, and show varying degrees of phosphorylation (isoforms) in cow's milk. Previous studies suggest that α_{s1}- and α_{s2}-CN might be more important for stabilizing internal micellar structure than β- and κ-CN (Dalgleish and Corredig, 2012; Huppertz et al., 2017). α_{s1}-CN has been observed to carry 8 to 9 phosphate groups (**P**), and α_{s1}-CN-8P is the predominant isoform (Holland and Boland, 2014). α_{s2}-CN has been observed to carry 9 to 15 phosphate groups, and α_{s2}-CN-11P is the predominant isoform (Fang et al., 2016).

Relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms vary considerably among milk of individual cows (Bijl et al., 2014; Fang et al., 2016), and exploitable genetic variation for these isoforms exists in French Montbéliarde (Fang et al., 2017a), Danish Holstein and Danish Jersey (Buitenhuis et al., 2016), and Dutch Holstein

77 Friesian (Bijl et al., 2014; Fang et al., 2017b). Furthermore, the phosphorylation degrees (**PD**)
78 of α_{s1}-CN and α_{s2}-CN, defined as the proportion of isoforms with higher degrees of
79 phosphorylation, are heritable in French Montbéliarde (Fang et al., 2017a) and highly
80 heritable in Dutch Holstein Friesian (Fang et al., 2017b). This indicates that the difference in
81 the phosphorylation process is to a great extent determined by genetic factors. Additionally,
82 Bijl et al. (2014) showed that α_{s1}-CN-8P and α_{s1}-CN-9P are largely regulated by different sets
83 of genes. Our recent work also suggests that α_{s1}-and α_{s2}-CN phosphorylated at lower degrees
84 is regulated differently from α_{s1}- and α_{s2}-CN phosphorylated at higher degrees (Fang et al.,
85 2016; Fang et al., 2017b). To date, little is known about the genetic backgrounds of individual
86 α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD. Therefore, this study aimed to
87 identify genomic regions associated with these traits.

MATERIALS AND METHODS

Animals

Test-day morning milk samples were collected from approximately 2,000 primiparous Dutch Holstein-Friesian cows as part of the Dutch Milk Genomic Initiative. Cows were located on 398 herds in the Netherlands, and at least 3 cows per herd were sampled. The pedigree of the cows was supplied by cattle improvement organization CRV (Arnhem, the Netherlands). Detailed description of the experimental design is provided by Schopen et al. (2009).

Phenotypes

Milk production traits, phosphorous, and milk protein composition from 1,857 milk samples collected in winter (February and March 2005) were available for the current study.

Milk production traits and phosphorus. Protein percentage was determined by infrared spectroscopy using MilkoScan FT 6000 (Foss Electric, Hillerød, Denmark) at the milk control station laboratory (Qlip, Zutphen, the Netherlands). Phosphorus concentration was determined by inductively coupled plasma-atomic emission spectrometry (Vista Axial, Varian, Australia) from whole milk as described in van Hulzen et al. (2009). Test-day morning milk yield was available for 1,721 cows and was obtained from CRV. Yields of protein and phosphorus were calculated by multiplying the respective content traits by the observed test-day morning milk yield.

Milk protein composition. Relative concentrations (% wt/wt) of individual milk proteins and their isoforms were determined by capillary zone electrophoresis (CZE) by Heck et al. (2008) and Fang et al (2017b). Yields (in grams) of individual milk proteins and their isoforms were calculated by multiplying relative concentrations (% wt/wt) by protein yield (in grams). Relative concentrations of α_{s1}- and α_{s2}-CN phosphorylation isoforms are the result of two distinct processes: the production of α_{s1}- and α_{s2}-CN and the posttranslational modification of both caseins. To specifically characterize the phosphorylation process, we defined the

phosphorylation degrees of α_{s1}-CN and α_{s2}-CN as the proportion of isoforms with higher degree of phosphorylation (Fang et al., 2017b), which were calculated as

$$\alpha_{s1}\text{-CN PD} = \left(\frac{\alpha_{s1}\text{-CN-9p}}{\alpha_{s1}\text{-CN-8p}+\alpha_{s1}\text{-CN-9p}}\right) \times 100\%$$

$$\alpha_{s2}\text{-CN PD} = \left(\frac{\alpha_{s2}\text{-CN-12p}}{\alpha_{s2}\text{-CN-10p}+\alpha_{s2}\text{-CN-11p}+\alpha_{s2}\text{-CN-12p}}\right) \times 100\%$$

***Casein Phosphate*.** The phosphorus distribution that is bound to casein was quantified by estimating the content of phosphate groups attached to caseins in milk (i.e. molar concentration of casein phosphate, **P_{CN})** and the total amount of phosphate groups attached to caseins in milk (i.e. output of casein phosphate, P_{CN} yield). To derive P_{CN}, we first calculated the molar concentration (**C_{molar}**) of each casein (α_{s1}-, α_{s2}-, κ- and β-CN) as its concentration in milk (g/L), calculated as protein percentage (% wt/wt) × 10, divided by its respective molecular weight (Da). As κ-CN carried 1 phosphate group and β-CN carried 5 in our milk samples, C_{molar} of phosphate groups attached to κ-CN and β-CN were approximated by multiplying C_{molar} of κ-CN and β-CN by 1 and 5, respectively. The C_{molar} of phosphate groups attached to α_{s1}-CN was the sum of C_{molar} of α_{s1}-CN-8P multiplied by 8 and C_{molar} of α_{s1}-CN-9P multiplied by 9. The C_{molar} of phosphate groups attached to α_{s2}-CN was the sum of C_{molar} of α_{s2}-CN-10P multiplied by 10, C_{molar} of α_{s2}-CN-11P multiplied by 11, and C_{molar} of α_{s2}-CN-12P multiplied by 12. Therefore, P_{CN} was the sum of C_{molar} of α_{s1}-, α_{s2}-, κ- and β-CN. Subsequently, P_{CN} yield was approximated by multiplying P_{CN} by test-day morning milk yield.

Genotypes

DNA was isolated from blood samples of 1,868 cows for genotyping. As described in detail by Schopen et al. (2011), a 50K (~50,000) SNP chip developed by CRV was used to genotype cows with the Infinium assay technology (Illumina Inc., San Diego, CA). The map positions of the SNP were based on bovine genome assembly BTAU 4.0 (Liu et al., 2009).

Monomorphic SNP, SNP with a genotyping rate < 80%, and SNP with less than 10 observations for one of the genotype classes were discarded (SNP with only two genotype classes instead of three were kept in the final marker set in case at least 10 observations per genotype class). After filtering, 44,669 SNP were retained for the genome-wide association study (**GWAS**). The data set used in the association study consisted of 1,667 animals with both phenotypes and genotypes. Protein variants A and B for β-lactoglobulin (β-LG) were genotyped for 1,671 cows as described by Ganai et al. (2009). Genotypes for the diacylglycerol acyltransferase 1 (DGAT1) K232A polymorphism were obtained for 1,702 cows as described by Schennink et al. (2007).

Statistical Analyses

GWAS. Single-SNP associations were analyzed using the following animal model:

$$y_{klmno} = \mu + \beta_1 dim_{klmno} + \beta_2 e^{-0.05*dim_{klmno}} + \beta_3 ca_{klmno} + \beta_4 ca^2_{klmno} + season_k + scode_l + animal_m + herd_n + SNP_o + e_{klmno}, \quad [1]$$

where y_{klmno} is the observation of the trait of interest; μ is the overall mean of the trait; $\beta_{1,2}$ are the regression coefficients for dim_{klmno}; $\beta_{3,4}$ are the regression coefficients for ca_{klmno}; dim_{klmno} is a covariate describing the effect of days in lactation, modeled with a Wilmink curve (Wilmink, 1987); ca_{klmno} is a covariate describing the effect of age at first calving; $season_k$ is the fixed effect for calving season (June–August 2004, September–November 2004, and December 2004–February 2005); $scode_l$ is the fixed effect accounting for possible differences in genetic level between proven bull daughters and young bull daughters; $animal_m$ is the random additive genetic effect assumed to be distributed as $N(\mathbf{0}, \mathbf{A}\sigma_a^2)$, where **A** is the additive genetic relationships matrix consisting of 26,300 animals, and σ_a^2 is the additive genetic variance; $herd_n$ is the random herd effect assumed to be distributed as $N(\mathbf{0}, \mathbf{I}\sigma_{herd}^2)$, where **I** is the identity matrix, and σ_{herd}^2 is the herd variance; SNP_o is the fixed effect of the

SNP modeled as a class variable; e_{klmno} is the random residual effect assumed to be distributed as $N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where **I** is the identity matrix, and σ_e^2 is the residual variance. Variance components were fixed to estimates obtained from model [1] without the SNP effect. The effects of β-LG protein variants and DGAT1 genotypes were estimated using model [1] by replacing the SNP effect by protein variant and genotype effects, respectively. All statistical analyses were performed using ASReml 4.1 (Gilmour et al., 2015).

Significance Thresholds. The genome-wide false discovery rate (**FDR**) was calculated based on the P-values obtained from the single-SNP analyses using the R package qvalue (Dabney et al., 2010; R Core Team, 2015). The FDR was calculated for each trait separately. Associations with an FDR < 0.01 were considered significant. Obtained results are shown as Manhattan plots constructed by qqman R package (Turner, 2014).

QTL regions. Because of strong linkage disequilibrium between neighboring SNP, significant SNP located close to each other might be associated with the same causal variant. Therefore, we defined QTL regions as follows: a QTL region starts with the first significant SNP on a chromosome that is followed by an additional significant SNP within 10 Mega-base pairs (**Mbp**), extends as long as another significant SNP occurs within 10 Mbp from the previous one, and ends at the last significant SNP that is not followed by another significant SNP within the next 10 Mbp.

RESULTS AND DISCUSSION

In this study, we explored the genetic background of individual α_{s2}-CN phosphorylation isoforms (% wt/wt), and the phosphorylation degrees of α_{s1}-CN (α_{s1}-CN PD) and α_{s2}-CN (α_{s2}-CN PD). Phenotypic means, standard deviations, heritability estimates and proportions of variance explained by herd for all studied traits are given in Table 1. For α_{s2}-CN, the predominant isoform was α_{s2}-CN-11P. The proportion of isoforms with higher degree of phosphorylation was 26% for α_{s1}-CN and 34% for α_{s2}-CN. Heritability estimates were moderate to very high for all traits. Results have been discussed in detail by Fang et al. (2017b).

The GWAS showed significant associations for all studied traits, and a total of 10 QTL regions were identified (FDR < 0.01) on 10 different chromosomes (BTA1, 2, 6, 9, 11, 14, 15, 18, 24 and 28, see Figure 1). Recently, FAM20C was discovered as the kinase that phosphorylates secretory pathway proteins with S-X-E/pS motifs (X represents any amino acid residue, and p indicates phosphorylation) including the caseins found in milk as well as several other proteins implicated in biomineralization (Tagliabracci et al., 2012). We did not detect a QTL signal on BTA25 where the *FAM20C* gene is located (between 43.86 to 43.90 Mbp (BTAU 4.0)), neither for individual α_{s2}-CN phosphorylation isoforms, nor for α_{s1}-CN PD and α_{s2}-CN PD. This is in line with results reported by Bijl et al. (2014) and Buitenhuis et al. (2016) and suggests that no *FAM20C* variants are segregating in the Dutch Holstein population or in the Danish Holstein and Danish Jersey populations. Regions associated with multiple traits were found on BTA1, 6, 11, and 14, and their effects will be discussed in detail. Furthermore, producing casein phosphorylation isoforms is a function of casein synthesis and their subsequent phosphorylation. Little is known about genes regulating the phosphorylation process. This process might be interlinked with different pathways of milk production, including milk protein synthesis and phosphorus secretion in milk. To investigate

if the detected QTL specifically affect the phosphorylation, we extended the analyses for the QTL on BTA1, 6, 11 and 14 which were associated with multiple traits. Estimated genotypic effects on a range of traits, including content and yield of phosphorus and phosphate groups attached to caseins (P_{CN} and P_{CN} yield) are expected to provide insight in the nature of the observed QTL. Genotype effects of the most significantly associated (lead) SNP in each QTL region are reported in Table 2.

BTA1

The QTL region between 145.55 and 152.18 Mbp on BTA1 was significantly associated with relative concentrations of α_{s2}-CN-11P and α_{s2}-CN-12P, and with α_{s1}-CN PD. However, the lead SNP differed between traits: ARS-BFGL-NGS-8140 at 149.19 Mbp was the lead SNP for α_{s2}-CN-11P concentration [$-\log_{10}(P) = 8.01$], ARS-BFGL-NGS-91705 (rs43282015) at 149.65 Mbp was the lead SNP for α_{s2}-CN-12P concentration [$-\log_{10}(P) = 8.51$] and ARS-BFGL-NGS-24811 at 146.63 Mbp was the lead SNP for α_{s1}-CN PD [$-\log10(P) = 5.50$]. Interestingly, significant associations on BTA1 were found for all studied traits except for α_{s1}-CN-8P concentration. For both α_{s2}-CN-10P and α_{s2}-CN PD, only one SNP reached the significant threshold (BTB-00068200 as the lead SNP for α_{s2}-CN-10P and ULGR_BTA-55413 as the lead SNP for α_{s2}-CN PD), so they did not qualify as a QTL region.

To investigate if this region harbors multiple QTL (see Figure 2A for associations of α_{s2}-CN-12P and α_{s1}-CN PD as examples), associations for all studied traits were reanalyzed after adjusting for the lead SNP for α_{s2}-CN-12P concentration (ARS-BFGL-NGS-91705) for all studied traits (see Figure 2B for α_{s2}-CN-12P and α_{s1}-CN PD as examples). This analysis resulted in no significant associations for α_{s2}-CN-11P and α_{s2}-CN-12P except for one isolated SNP for α_{s2}-CN-12P. However, significant associations remained for α_{s2}-CN-10P, α_{s1}-CN PD and α_{s2}-CN PD. To identify which of the lead SNP of the different traits tags this region for all studied traits (see Figure 2C for the most significant SNP of each trait), associations were

reanalyzed after adjusting for the lead SNP of α_{s1}-CN PD for all studied traits. The same analyses were repeated with the respective lead SNP of α_{s2}-CN-10P and α_{s2}-CN PD. Only after adjusting for the lead SNP for α_{s2}-CN-10P concentration (BTB-00068200), the QTL signal for α_{s2}-CN-10P, α_{s1}-CN PD and α_{s2}-CN PD disappeared, whereas the signals for α_{s2}-CN-11P and α_{s2}-CN-12P were hardly affected (Figure 2D). These results suggest that BTA1 harbors two QTL affecting α_{s1}- and α_{s2}-CN phosphorylation: QTL1 located in the region between 147.5 and 152.1 Mbp and represented by ARS-BFGL-NGS-91705 and QTL2 located in the region between 144.41 and 147.3 Mbp and represented by BTB-00068200. The low level of linkage disequilibrium between ARS-BFGL-NGS-91705 and BTB-00068200 (r^2=0.09) supports the presence of two QTL in this region.

The effects of SNP ARS-BFGL-NGS-91705 (QTL1) and BTB-00068200 (QTL2) on relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD are given in Table 2. For ARS-BFGL-NGS-91705, the *G* allele was associated with lower α_{s2}-CN-11P and α_{s2}-CN-12P concentrations but not with α_{s1}-CN PD or α_{s2}-CN PD. For BTB-00068200, the *G* allele was associated with lower α_{s1}-CN-9P concentration and higher α_{s2}-CN-10P and -11P concentrations. This results in lower degrees of phosphorylation of α_{s1}-CN and α_{s2}-CN as shown by the negative association of the GG genotype with both α_{s1}-CN PD and α_{s2}-CN PD. Taken together, our results suggest that QTL1 affects α_{s2}-CN production, and QTL2 affects the phosphorylation degrees of α_{s1}- and α_{s2}-CN. Furthermore, combining our results with those reported by Bijl et al. (2014) and Schopen et al. (2011) indicates that QTL1 on BTA1 is involved only in α_{s2}-CN production but not in α_{s1}-CN production, suggesting α_{s1}-CN and α_{s2}-CN are regulated differently. The lead SNP (ARS-BFGL-NGS-91705) of the QTL1 region is an intergenic variant. The gene closest to the lead SNP is F-box protein 25 (*FBXO25*) that is located at 149.56-149.59 Mbp on BTA1. In cattle, FBXO25 is involved in the pathway of post-translational protein modification as adding ubiquitin to the

substrate protein according to UniProt (http://www.uniprot.org/) but has not been associated with milk characteristics. The QTL2 region harbors the *SLC37A1* gene (145.72-145.80 Mbp) encoding for a protein functioning as a phosphorus antiporter that translocates inorganic phosphate in exchange of glucose-6-phosphate (Pan et al., 2011). Furthermore, a QTL associated with phosphorus concentration has been identified in this region in Danish Jersey (Buitenhuis et al., 2016) and Australian Holstein (Kemper et al., 2016).

We detected significant effects of ARS-BFGL-NGS-91705 (QTL1) on yields of α_{s1}-CN-8P, α_{s2}-CN-10P, -11P and -12P, protein, phosphorus and P_{CN} (Table 2). These consistent negative associations of the *G* allele with the yield traits confirm that QTL1 might affect only the production of α_{s2}-CN. This is supported by the fact that we detected the significant effect of QTL1 on test-day morning protein yield but did not detect significant effects on α_{s1}-CN PD and α_{s2}-CN PD in the current study. The effect on protein yield is relatively small probably because α_{s2}-CN contributes only about 10 % to the total milk protein. Furthermore, this QTL has been reported to be associated with protein yield in Chinese Holstein (Jiang et al., 2010).

For BTB-00068200 (QTL2), we detected significant effects on yields of α_{s2}-CN-10P and -11P and phosphorus as well as phosphorus content. The *G* allele was associated with higher yields of α_{s2}-CN-10P and α_{s2}-CN -11P as well as higher content and yield of phosphorus. Furthermore, we did not detect significant effects of BTB-00068200 on yields of milk and protein. Therefore, the highly significant effect of QTL2 on phosphorus content [$-\log_{10}(P)$ = 17.40] might be mainly due to the change of total phosphorus output in milk rather than a change of milk volume. Similarly, significant effects of QTL2 on relative concentrations of α_{s2}-CN-10P and -11P are probably mainly due to the change of yields of α_{s2}-CN-10P and α_{s2}-CN -11P rather than a change of protein yield. Taken together, these associations suggest that QTL2 has direct effects on phosphorylation degree by increasing the amount of the less phosphorylated isoforms, which might be related to the regulation of phosphorus output in

milk. This is also supported by significant associations of BTB-00068200 with α_{s1}-CN PD and α_{s2}-CN PD. Furthermore, the *SLC37A1* gene located in this region plays a role in translocating inorganic phosphate (Pan et al., 2011), and it has been associated with the phosphorus content in cows' milk (Kemper et al., 2016). Here, we show that this gene might have a direct effect on total phosphorus output in milk, especially on the inorganic phosphorus because we detected fairly small effects of BTB-00068200 on the content of phosphate groups attached to caseins (P_{CN}) and no significant effect on the total amount of phosphate groups attached to caseins (P_{CN} yield). Furthermore, the route of secreting inorganic phosphorus has been shown to be similar to that of casein phosphate, which is via the Golgi apparatus (Shennan and Peaker, 2000). We, therefore, hypothesize that the effect of QTL2 on phosphorylation degrees of α_{s1}- and α_{s2}-CN might be because the secretion of inorganic phosphate is interlinked with phosphorylation of caseins in the Golgi apparatus (Bingham and Farrell, Jr., 1974; Moore et al., 1985) .

BTA6

The QTL region between 46.52 and 103.18 Mbp on BTA 6 was significantly associated with relative concentrations of α_{s2}-CN-10P, α_{s2}-CN-11P, and α_{s2}-CN-12P. This region harbors the casein gene cluster (around 87 Mbp). The SNP ARS-BFGL-NGS-94898 at 87.66 Mbp was the lead SNP for α_{s2}-CN-10P concentration [$-\log_{10}(P) = 5.44$]. The SNP ULGR_BTC-053514 at 83.57 Mbp was the lead SNP for both α_{s2}-CN-11P concentration [$-\log_{10}(P) = 38.49$] and α_{s2}-CN-12P concentration [$-\log_{10}(P) = 46.04$]. This SNP was also previously reported as the lead SNP for α_{s1}-CN-9P concentration (Bijl et al., 2014) and for total α_{s2}-CN concentration (Schopen et al., 2011). No significant association on BTA6 was found with α_{s1}-CN PD and α_{s2}-CN PD (Figure 1), suggesting this region is only involved in casein production but not in the phosphorylation process. As shown by Fang et al. (2017b), the proportion of isoforms

with higher degrees of phosphorylation is hardly affected when more α_{s1}- and α_{s2}-CN are produced, indicating that phosphorylation might not be an important rate limiting step.

The estimated effects of the lead SNP for α_{s2}-CN-12P concentration on relative concentrations of individual α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD show that the *G* allele was associated with lower concentrations of individual α_{s2}-CN phosphorylation isoforms (Table 2). Highly significant effects on yields of individual α_{s2}-CN phosphorylation isoforms confirm that this QTL affects α_{s2}-CN production. Note that this SNP did not pass the genome-wide significance threshold for α_{s1}-CN PD and α_{s2}-CN PD.

BTA11

The QTL region between 95.06 and 109.41 Mbp on BTA11 was significantly associated with relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. This region harbors the *PAEP* gene encoding for β-LG. The SNP ULGR_SNP_X14710_1740 (rs41255679) at 107.2 Mbp was the lead SNP for α_{s2}-CN-10P concentration [$-\log_{10}(P)$ = 4.96], α_{s2}-CN-11P concentration [$-\log_{10}(P)$ = 10.42], α_{s1}-CN PD [$-\log_{10}(P)$ = 6.33] and α_{s2}-CN PD [$-\log_{10}(P)$ = 10.42]. This SNP was previously reported as the lead SNP on BTA11 for α_{s1}-CN-8P concentration (Bijl et al., 2014). It is located in the promoter region of the *PAEP* gene and is in linkage disequilibrium with β-LG protein variants A and B (Ganai et al., 2009). The estimated effects of β-LG genotypes on all studied traits (Table 2) show that the BB genotype was associated with higher α_{s1}-CN-8P, α_{s2}-CN-10P and α_{s2}-CN-11P concentrations (isoforms with lower degrees of phosphorylation). This results in lower degrees of phosphorylation of α_{s1}-CN and α_{s2}-CN as shown by the negative association of the BB genotype with both α_{s1}-CN PD and α_{s2}-CN PD. Buitenhuis et al. (2016) and Fang et al. (2017a) did not detect significant effects of β-LG genotypes on individual α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD in Danish Holstein and Danish Jersey, and in French Montbéliarde, respectively. Differences between studies might be due to the

329 genetic differences between studied breeds (Holstein, Montbéliarde and Jersey) such as
330 differences in linkage disequilibrium between β-LG genotypes and other variants that affect
331 the traits of interest, limited sample size of Buitenhuis et al. (2016) and Fang et al. (2017a),
332 and the use of different analytical methods. Regarding differences in linkage disequilibrium
333 between β-LG genotypes and other variants across breeds, Bijl et al. (2014) detected a
334 significant effect of β-LG genotypes only on α_{s1}-CN-8P concentration in Dutch Holstein
335 Friesian, whereas Fang et al. (2017a) detected significant effects of β-LG genotypes on both
336 α_{s1}-CN-8P and α_{s1}-CN-9P concentrations in French Montbéliarde. The reported effect of β-
337 LG BB genotype on β-LG concentration in French Montbéliarde by Fang et al. (2017a) was
338 about 1.5 times larger than the one in Dutch Holstein Friesian reported by Heck et al. (2009).
339 Therefore, the observed genotype effect of β-LG in French Montbéliarde might be the result
340 of multiple linked variants. This might explain the differences in effects of β-LG genotypes
341 on individual α_{s1}- and α_{s2}-CN isoforms, α_{s1}-CN PD and α_{s2}-CN PD in different breeds.
342 Differences between CZE used by Bijl et al. (2014) and LC-ESI/MS used by Fang et al.
343 (2017a) seem to be negligible for the measurement of α_{s1}-CN isoforms and α_{s1}-CN PD
344 according to Fang et al. (2017b). However, protein fractions measured with the same
345 analytical method, such as LC (as used by Buitenhuis et al., 2016 and by Fang et al., 2017a),
346 may still differ because of differences in separation conditions.
347 Significant effects of β-LG genotypes on yields of α_{s1}-CN-9P, α_{s2}-CN-10P and -11P were
348 detected (Table 2), but the effects on α_{s1}-CN-9P yield were relatively small. The BB genotype
349 was associated with higher yields of α_{s2}-CN-10P and -11P. Surprisingly, we did not detect a
350 significant effect on the yield of α_{s1}-CN-8P, whereas we detected a highly significant effect on
351 α_{s1}-CN-8P concentration. Furthermore, we detected highly significant effects of β-LG
352 genotypes on P_{CN} content [$-\log_{10}(P) = 6.49$] but no effect on P_{CN} yield. This might be due to
353 the fact that β-LG genotype is associated with higher proportion of caseins. Previous studies

have shown that the β-LG B variant decreases the proportion of β-LG, which results in increased proportions of caseins (Bobe et al., 1999; Hallén et al., 2008; Heck et al., 2009; Bonfatti et al., 2010; Fang et al., 2017a). We show that the β-LG B variant increases only the proportions of α_{s1}- and α_{s2}-CN isoforms phosphorylated at a lower degree and thus decreases α_{s1}-CN PD and α_{s2}-CN PD. Three possible explanations could be that either phosphorylation is a rate limiting step due to increased casein production, the interaction between the amount of phosphorus available and increased α_{s1}- and α_{s2}-CN production, or β-LG has a role in the phosphorylation. As discussed above, phosphorylation might not be an important rate limiting step for the production of α_{s1}- and α_{s2}-CN phosphorylation isoforms, thus, increased α_{s1}- and α_{s2}-CN production that is associated with the β-LG B variant in itself should not affect α_{s1}-CN PD and α_{s2}-CN PD. Furthermore, we did not detect significant interactions between β-LG genotypes and QTL2 on BTA1 for α_{s1}-CN PD ($P = 0.26$) and α_{s2}-CN PD ($P = 0.56$), suggesting the amount of phosphorus available is not rate limiting. Taken together, β-LG seems to play a role in regulating milk protein composition, proportion of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, and the phosphorylation process. Several roles have been suggested for β-LG but its true biological function remains elusive (Kontopidis et al., 2002). Therefore, the actual mechanism causing the effects of β-LG genotypes on the concentrations of caseins is currently unknown as well as the role of β-LG in the phosphorylation process.

BTA14

The QTL region between 0.2 and 19.36 Mbp on BTA14 was significantly associated with relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD, and α_{s2}-CN PD. This region harbors the *DGAT1* gene. The SNP ULGR_ SNP_AJ318490_1c (rs109234250) at 0.44 Mbp was the lead SNP for α_{s2}-CN-10P concentration [$-\log_{10}(P) = 17.04$], α_{s2}-CN-11P concentration [$-\log_{10}(P) = 21.98$], α_{s1}-CN PD [$-\log_{10}(P) = 48.79$] and α_{s2}-CN PD [$-\log_{10}(P) = 21.55$]. This SNP was previously reported as the lead SNP on BTA14 for α_{s1}-CN-9P

379 concentration (Bijl et al., 2014), and is one of two SNPs responsible for the DGAT1 K232A
380 polymorphism.
381 The effects of DGAT1 genotypes on all studied traits (Table 2) show that the *K* allele was
382 associated with higher α_{s2}-CN-10P and α_{s2}-CN-11P concentrations (isoforms with lower
383 degrees of phosphorylation) and lower concentration of α_{s1}-CN-9P (isoform with higher
384 degrees of phosphorylation). This results in lower degrees of phosphorylation of α_{s1}- and α_{s2}-
385 CN as shown by the negative association of the *K* allele with both α_{s1}-CN PD and α_{s2}-CN PD.
386 Furthermore, *DGAT1* does not affect α_{s1}-CN-8P concentration (Bijl et al., 2014) and α_{s2}-CN-
387 12P concentration (this study) at the genome wide significance level. Bovenhuis et al. (2016)
388 showed that *DGAT1* affects α_{s2}-CN concentration in Dutch Holstein Friesian and Danish
389 Holstein Friesian. The *K* allele was associated with higher α_{s2}-CN concentration. Here, we
390 show that the increase of α_{s2}-CN concentration is due to the increase of α_{s2}-CN-10P and α_{s2}-
391 CN-11P concentrations in Dutch Holstein Friesian.
392 For the effects of DGAT1 genotypes on the yields of individual isoforms and milk production
393 traits, we detected significant effects on yields of all α_{s1}- and α_{s2}-CN phosphorylation
394 isoforms, milk, protein, phosphorus and P_{CN} as well as contents of protein, phosphorus and
395 P_{CN}. The effects of the DGAT1 genotypes on the yields of α_{s1}-CN-8P, α_{s2}-CN-11P and α_{s2}-
396 CN-12P were relatively small, and might be due to the change in the protein yield as the
397 genotype effects on the yields of α_{s1}-CN-8P, α_{s2}-CN-11P, α_{s2}-CN-12P and protein were in the
398 same direction and of similar magnitude. The highly significant effects on yields of α_{s1}-CN-
399 9P and α_{s2}-CN-10P suggest direct effects of DGAT1 on these isoforms as the direction and
400 magnitude of effects on their relative concentration in milk and yields are similar. The
401 biological relation between DGAT1, content and yield of fat, and fatty acid composition are
402 easier to comprehend as the DGAT1 enzyme is involved in biosynthesis of triacylglycerol
403 (Coleman and Lee, 2004), whereas the biological relation of DGAT1 and phosphorylation of

caseins is still unclear. The contribution of DGAT1 to the variation in specific isoforms in Dutch Holstein Friesian seems similar to the contribution of QTL2 on BTA1, suggesting a similar mode of action. This is in line with Bovenhuis et al. (2016), who showed that DGAT1 might affect the distribution of phosphorus between casein micelles and milk serum.

Additional Regions

In addition to the four QTL regions with effects on multiple casein phosphorylation traits, we also detected trait-specific QTL on BTA2, 9, 15, 18, 24 and 28. On BTA2, a QTL region located between 113.63 and 113.67 Mbp was associated uniquely with α_{s2}-CN-12P concentration. The gene closest to this QTL region is ephrin type-A receptor 4 precursor (*EPHA4*) that is located at 114.15-114.20 Mbp on BTA2. In cattle, EPHA4 is an uncharacterized protein, whereas in human, it is a kinase phosphorylating tyrosine and is involved in cell adhesion and neurogenesis (Murai et al., 2003; Poitz et al., 2015). Two QTL regions were associated uniquely with α_{s2}-CN-11P concentration. The first QTL region is located between 98.45 and 99.32 Mbp on BTA9. The lead SNP ARS-BFGL-NGS-102803 (rs109099768) is an intron variant located in the serine active site containing 1 (*SERAC1*) gene. In human, the SERAC1 protein plays an important role in mediating phospholipid exchange that is essential for both mitochondrial functioning and intracellular cholesterol trafficking (Wortmann et al., 2012). The second QTL region located between 18.55 and 19.13 Mbp on BTA28 harbors the receptor accessory protein 3 (*REEP3*) gene. A total of three unique QTL regions were associated with α_{s1}-CN PD, which were located at 54.61 Mbp on BTA15, between 35.68 and 36.09 Mbp on BTA18, and between 20.49 and 21.11 Mbp on BTA24, respectively. On BTA15, the gene closest to the QTL region is microtubule affinity-regulating kinase 1 (*MARK1*). On BTA18, the QTL region harbors the proteasome 26S subunit, non-ATPase 7 (*PSMD7*) gene (36.00-36.02 Mbp). On BTA24, the QTL region harbors the CUGBP Elav-like family member 4 (*CELF4*) gene. None of the genes mentioned

above has been associated with milk characteristics. Thus, no clear candidate genes could be identified for those trait-specific QTL.

CONCLUSION

We detected a total of 10 QTL regions for relative concentrations of individual α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN (α_{s1}-CN PD and α_{s2}-CN PD) on chromosomes 1, 2, 6, 9, 11, 14, 15, 18, 24 and 28. Regions associated with multiple traits were found on BTA 1, 6, 11, and 14. We showed two QTL regions on BTA1: one affects α_{s2}-CN production and the other harboring the *SLC37A1* gene affects the phosphorylation of α_{s1}-CN and α_{s2}-CN. The QTL region on BTA6 harbors the casein gene cluster and affects the production of casein. The QTL region on BTA11 harbors the *PAEP* gene encoding β-LG and affects both casein production and phosphorylation. The QTL region on BTA14 harbors the *DGAT1* gene and effects on phosphorylation of α_{s1}-CN and α_{s2}-CN are likely to be indirect, i.e. due to the effect of DGAT1 on traits like milk yield and protein content. Elucidation of the actual roles of SLC37A1, β-LG and DGAT1 in α_{s1}- and α_{s2}-CN phosphorylation in Dutch Holstein Friesian requires further investigation. Furthermore, more knowledge on the effects of the phosphorylation degrees of α_{s1}-CN and α_{s2}-CN on technological properties of milk is needed before results can be implemented in breeding.

ACKNOWLEDGMENTS

This study is part of the Dutch Milk Genomics Initiative, funded by Wageningen University (Wageningen, the Netherlands), the Dutch Dairy Association NZO (Zoetermeer, the Netherlands), Cooperative Cattle Improvement Organization CRV (Arnhem, the Netherlands), and the Dutch Technology Foundation STW (Utrecht, the Netherlands). The first author benefited from an Erasmus-Mundus fellowship and a grant by APIS-GENE, within the framework of the European Graduate School in Animal Breeding and Genetics (EGS-ABG, Paris, France).

REFERENCES

Bijl, E., H.J.F. van Valenberg, T. Huppertz, A.C.M. van Hooijdonk, and H. Bovenhuis. 2014. Phosphorylation of αS1-casein is regulated by different genes. J. Dairy Sci. 97:7240–7246. doi:10.3168/jds.2014-8061.

Bingham, E.W., and H.M. Farrell, Jr. 1974. Casein kinase from the Golgi apparatus of lactating mammary gland. J. Biol. Chem. 249:3647–3651.

Bobe, G., D.C. Beitz, A.E. Freeman, and G.L. Lindberg. 1999. Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. J. Dairy Sci. 82:2797–2804. doi:10.3168/jds.S0022-0302(99)75537-2.

Bonfatti, V., G. Di Martino, A. Cecchinato, L. Degano, and P. Carnier. 2010. Effects of β-κ-casein (CSN2-CSN3) haplotypes and β-lactoglobulin (BLG) genotypes on milk production traits and detailed protein composition of individual milk of Simmental cows. J. Dairy Sci. 93:3797–3808. doi:10.3168/jds.2009-2779.

Bovenhuis, H., M.H.P.W. Visker, N.A. Poulsen, J. Sehested, H.J.F. van Valenberg, J.A.M. van Arendonk, L.B. Larsen, and A.J. Buitenhuis. 2016. Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. J. Dairy Sci. 99:3113–3123. doi:10.3168/jds.2015-10462.

Buitenhuis, B., N.A. Poulsen, G. Gebreyesus, and L.B. Larsen. 2016. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC Genet. 17. doi:10.1186/s12863-016-0421-2.

Caroli, A.M., S. Chessa, and G.J. Erhardt. 2009. Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. J. Dairy Sci. 92:5335–5352.

Coleman, R.A., and D.P. Lee. 2004. Enzymes of triacylglycerol synthesis and their regulation. Prog. Lipid Res. 43:134–176. doi:10.1016/S0163-7827(03)00051-1.

Dabney, A., J.D. Storey, and G.R. Warnes. 2010. qvalue: Q-value estimation for false discovery rate control. R Packag. version 1.

Dalgleish, D.G., and M. Corredig. 2012. The structure of the casein micelle of milk and its changes during processing. Annu. Rev. Food Sci. Technol. 3:449–467. doi:10.1146/annurev-food-022811-101214.

Fang, Z.H., H. Bovenhuis, A. Delacroix-Buchet, G. Miranda, D. Boichard, M.H.P.W. Visker, and P. Martin. 2017a. Genetic and nongenetic factors contributing to differences in αS-casein phosphorylation isoforms and other major milk proteins. J. Dairy Sci. 100:5564–5577. doi:10.3168/jds.2016-12338.

Fang, Z.H., H. Bovenhuis, H.J.F. van Valenberg, P. Martin, T. Huppertz, and M.H.P.W. Visker. 2017b. Genetic parameters for αS1- and αS2-casein phosphorylation isoforms in Dutch Holstein Friesian. J. Dairy Sci. 101:1–11.

Fang, Z.H., M.H.P.W. Visker, G. Miranda, A. Delacroix-Buchet, H. Bovenhuis, and P. Martin. 2016. The relationships among bovine αS-casein phosphorylation isoforms suggest different phosphorylation pathways. J. Dairy Sci. 99:8168–8177. doi:10.3168/jds.2016-11250.

Ganai, N.A., H. Bovenhuis, J.A.M. van Arendonk, and M.H.P.W. Visker. 2009. Novel polymorphisms in the bovine beta-lactoglobulin gene and their effects on beta-lactoglobulin protein concentration in milk. Anim. Genet. 40:127–133. doi:10.1111/j.1365-2052.2008.01806.x.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Thompson. 2015. ASReml user guide release 4.1 structural specification. Hemel Hempstead VSN Int. Ltd.

Hallén, E., A Wedholm, A Andrén, A Lundén, A. Wedholm, A. Andrén, and A. Lundén.

2008. Effect of beta-casein, kappa-casein and beta-lactoglobulin genotypes on concentration of milk protein variants. J. Anim. Breed. Genet. 125:119–129. doi:10.1111/j.1439-0388.2007.00706.x.

Heck, J.M.L., C. Olieman, A. Schennink, H.J.F. van Valenberg, M.H.P.W. Visker, R.C.R. Meuldijk, and A.C.M. van Hooijdonk. 2008. Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. Int. Dairy J. 18:548–555. doi:10.1016/j.idairyj.2007.11.004.

Heck, J.M.L., A. Schennink, H.J.F. Van Valenberg, H. Bovenhuis, M.H.P.W. Visker, J.A.M. Van Arendonk, and A.C.M. van Hooijdonk. 2009. Effects of milk protein variants on the protein composition of bovine milk. J. Dairy Sci. 92:1192–1202. doi:10.3168/jds.2008-1208.

Holland, J.W., and M.J. Boland. 2014. Post-translational modifications of caseins. Second Edi. H. Singh, A. Thompson, and M. Boland, ed. Elsevier Inc.

Holt, C., J.A. Carver, H. Ecroyd, and D.C. Thorn. 2013. Invited review: Caseins and the casein micelle: Their biological functions, structures, and behavior in foods. Joural Dairy Sci. 96:6127–6146. doi:10.3168/jds.2013-6831.

van Hulzen, K.J.E., R.C. Sprong, R. van der Meer, and J.A.M. van Arendonk. 2009. Genetic and nongenetic variation in concentration of selenium, calcium, potassium, zinc, magnesium, and phosphorus in milk of Dutch Holstein-Friesian cows. J. Dairy Sci. 92:5754–5759. doi:10.3168/jds.2009-2406.

Huppertz, T., I. Gazi, H. Luyten, H. Nieuwenhuijse, A. Alting, and E. Schokker. 2017. Hydration of casein micelles and caseinates: Implications for casein micelle structure. Int. Dairy J. 74:1–11. doi:10.1016/j.idairyj.2017.03.006.

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in Chinese Holstein population.. PLoS One 5:e13661. doi:10.1371/journal.pone.0013661.

Kemper, K.E., M.D. Littlejohn, T. Lopdell, B.J. Hayes, P.M. Visscher, M. Carrick, and M.E. Goddard. 2016. Leveraging genetically simpler traits to identify small-effect variants for complex phenotypes. BMC Genomics 1–9. doi:10.1186/s12864-016-3175-3.

Kontopidis, G., C. Holt, and L. Sawyer. 2002. The ligand-binding site of bovine β-lactoglobulin: Evidence for a function?. J. Mol. Biol. 318:1043–1055. doi:10.1016/S0022-2836(02)00017-7.

De Kruif, C.G., and C. Holt. 2003. Casein micelle structure, functions and interactions. 3rd ed. P.F. Fox and P.L.H. McSweeney, ed. Springer US.

De Kruif, C.G., T. Huppertz, V.S. Urban, and A. V. Petukhov. 2012. Casein micelles and their internal structure. Adv. Colloid Interface Sci. 171–172:36–52. doi:10.1016/j.cis.2012.01.002.

Liu, Y., X. Qin, X.Z. Song, H. Jiang, Y. Shen, K.J. Durbin, S. Lien, M.P. Kent, M. Sodeland, Y. Ren, L. Zhang, E. Sodergren, P. Havlak, K.C. Worley, G.M. Weinstock, and R.A. Gibbs. 2009. Bos taurus genome assembly. BMC Genomics 10:180. doi:1471-2164-10-180 [pii]\r10.1186/1471-2164-10-180.

Moore, A., A.P. Boulton, H.W. Heid, E.D. Jarasch, and R.K. Craig. 1985. Purification and tissue-specific expression of casein kinase from the lactating guinea-pig mammary gland. Eur. J. Biochem. 152:729–37.

Murai, K.K., L.N. Nguyen, F. Irie, Y. Yamaguchi, and E.B. Pasquale. 2003. Control of hippocampal dendritic spine morphology through ephrin-A3/EphA4 signaling. Nat. Neurosci. 6:153.

Pan, C.J., S.Y. Chen, H.S. Jun, S.R. Lin, B.C. Mansfield, and J.Y. Chou. 2011. SLC37A1 and SLC37A2 are phosphate-linked, glucose-6-phosphate antiporters. PLoS One 6:1–8. doi:10.1371/journal.pone.0023157.

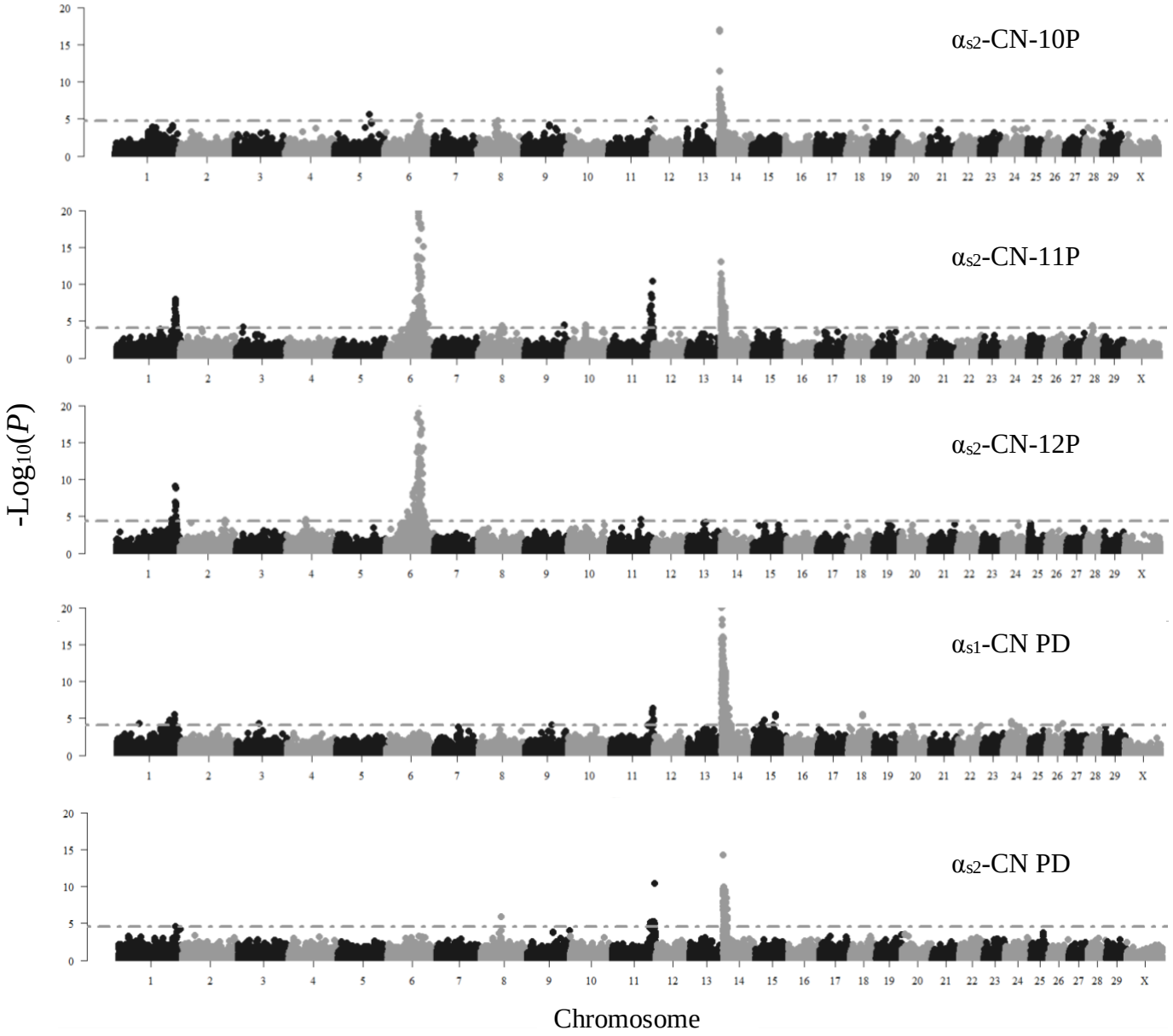
553 Poitz, D.M., G. Ende, B. Stütz, A. Augstein, J. Friedrichs, C. Brunssen, C. Werner, R.H.
554 Strasser, and S. Jellinghaus. 2015. EphrinB2/EphA4-mediated activation of endothelial
555 cells increases monocyte adhesion. Mol. Immunol. 68:648–656.
556 R Core Team. 2015. R: A language and environment for statistical computing. R Foundation
557 for Statistical Computing, Vienna, Austria. URL http//www. R-project. org.
558 Schennink, A., W.M. Stoop, M.H.P.W. Visker, J.M.L. Heck, H. Bovenhuis, J.J. Van Der
559 Poel, H.J.F. Van Valenberg, and J.A.M. Van Arendonk. 2007. DGAT1 underlies large
560 genetic variation in milk-fat composition of dairy cows. Anim. Genet. 38:467–473.
561 doi:10.1111/j.1365-2052.2007.01635.x.
562 Schopen, G.C.B., J.M.L. Heck, H. Bovenhuis, M.H.P.W. Visker, H.J.F. van Valenberg, and
563 J.A.M. van Arendonk. 2009. Genetic parameters for major milk proteins in Dutch
564 Holstein-Friesians. J. Dairy Sci. 92:1182–1191. doi:10.3168/jds.2008-1281.
565 Schopen, G.C.B., M.H.P.W. Visker, P.D. Koks, E. Mullaart, J.A.M. van Arendonk, and H.
566 Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy
567 cattle. J. Dairy Sci. 94:3148–3158. doi:10.3168/jds.2010-4030.
568 Shennan, D.B., and M. Peaker. 2000. Transport of milk constituents by the mammary gland..
569 Physiol. Rev. 80:925–51. doi:10.1152/physrev.2000.80.3.925.
570 Tagliabracci, V.S., J.L. Engel, J. Wen, S.E. Wiley, C.A. Worby, L.N. Kinch, J. Xiao, N. V
571 Grishin, and J.E. Dixon. 2012. Secreted kinase phosphorylates extracellular proteins that
572 regulate biomineralization. Science 336:1150–3. doi:10.1126/science.1217817.
573 Turner, S.D. 2014. qqman: an R package for visualizing GWAS results using QQ and
574 manhattan plots. BioRxiv 5165. doi:10.1101/005165.
575 Wedholm, A., L.B. Larsen, H. Lindmark-Månsson, A.H. Karlsson, and A. Andrén. 2006.
576 Effect of protein composition on the cheese-making properties of milk from individual
577 dairy cows. J. Dairy Sci. 89:3296–305. doi:10.3168/jds.S0022-0302(06)72366-9.
578 Wilmink, J.B.M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and
579 stage of lactation. Livest. Prod. Sci. 16:335–348. doi:10.1016/0301-6226(87)90003-0.
580 Wortmann, S.B., F.M. Vaz, T. Gardeitchik, L.E.L.M. Vissers, G.H. Renkema, J.H.M.
581 Schuurs-Hoeijmakers, W. Kulik, M. Lammens, C. Christin, and L.A.J. Kluijtmans. 2012.
582 Mutations in the phospholipid remodeling gene SERAC1 impair mitochondrial function
583 and intracellular cholesterol trafficking and cause dystonia and deafness. Nat. Genet.
584 44:797.
585

586 **Table 1.** Mean, standard deviation (SD), intra-herd heritability estimate (h^2)[a], and proportion
587 of phenotypic variance explained by herd (h_{herd})[a] for relative concentrations of individual α_{s1}-
588 CN and α_{s2}-CN phosphorylation isoforms, and for the phosphorylation degrees (PD)[b] of α_{s1}-
589 CN and α_{s2}-CN measured on test-day morning milk samples from 1,857 Dutch Holstein
590 Friesian cows (SE in parentheses).

Trait (%wt/wt)		Mean	SD	σ_p^2	h^2	h_{herd}
α_{s1}-CN[c]		33.64	1.66	2.80	0.52 (0.11)	0.11 (0.02)
	α_{s1}-CN-8P[c]	21.26	1.13	1.32	0.48 (0.10)	0.12 (0.02)
	α_{s1}-CN-9P[c]	7.42	1.07	1.18	0.76 (0.12)	0.08 (0.02)
α_{s2}-CN		6.67	0.95	0.98	0.94 (0.12)	0.08 (0.02)
	α_{s2}-CN-10P	0.99	0.39	0.16	0.54 (0.11)	0.10 (0.02)
	α_{s2}-CN-11P	3.44	0.57	0.33	0.89 (0.12)	0.08 (0.02)
	α_{s2}-CN-12P	2.24	0.22	0.05	0.71 (0.12)	0.07 (0.02)
Phosphorylation degree						
	α_{s1}-CN PD	25.79	2.72	7.66	0.78 (0.12)	0.08 (0.02)
	α_{s2}-CN PD	34.01	4.24	18.18	0.64 (0.11)	0.09 (0.02)

591 [a,b] Adopted from Fang et al. 2017b

592 [c]α_{s1}-CN PD = α_{s1}-CN-9P / (α_{s1}-CN-8P + α_{s1}-CN-9P) × 100; α_{s2}-CN PD = α_{s2}-CN-12P / (α_{s2}-
593 CN-10P + α_{s2}-CN-11P + α_{s2}-CN-12P) × 100. P = phosphate group attached.
594
595



596 **Figure 1.** Significance [–log10(*P*)] of associations of 44,669 genome wide SNP located on 29
597 *Bos taurus* autosomes and the X chromosome with individual αs2-CN phosphorylation
598 isoforms and the phosphorylation degrees of αs1-CN (αs1-CN PD) and αs2-CN (αs2-CN PD).
599 αs1-CN PD = αs1-CN-9P / (αs1-CN-8P + αs1-CN-9P) × 100; αs2-CN PD = αs2-CN-12P / (αs2-
600 CN-10P + αs2-CN-11P + αs2-CN-12P) × 100. P = phosphate group attached. The horizontal
601 line represents a false discovery rate of 1%. The y-axes are cut off at –log10(*P*)=20.

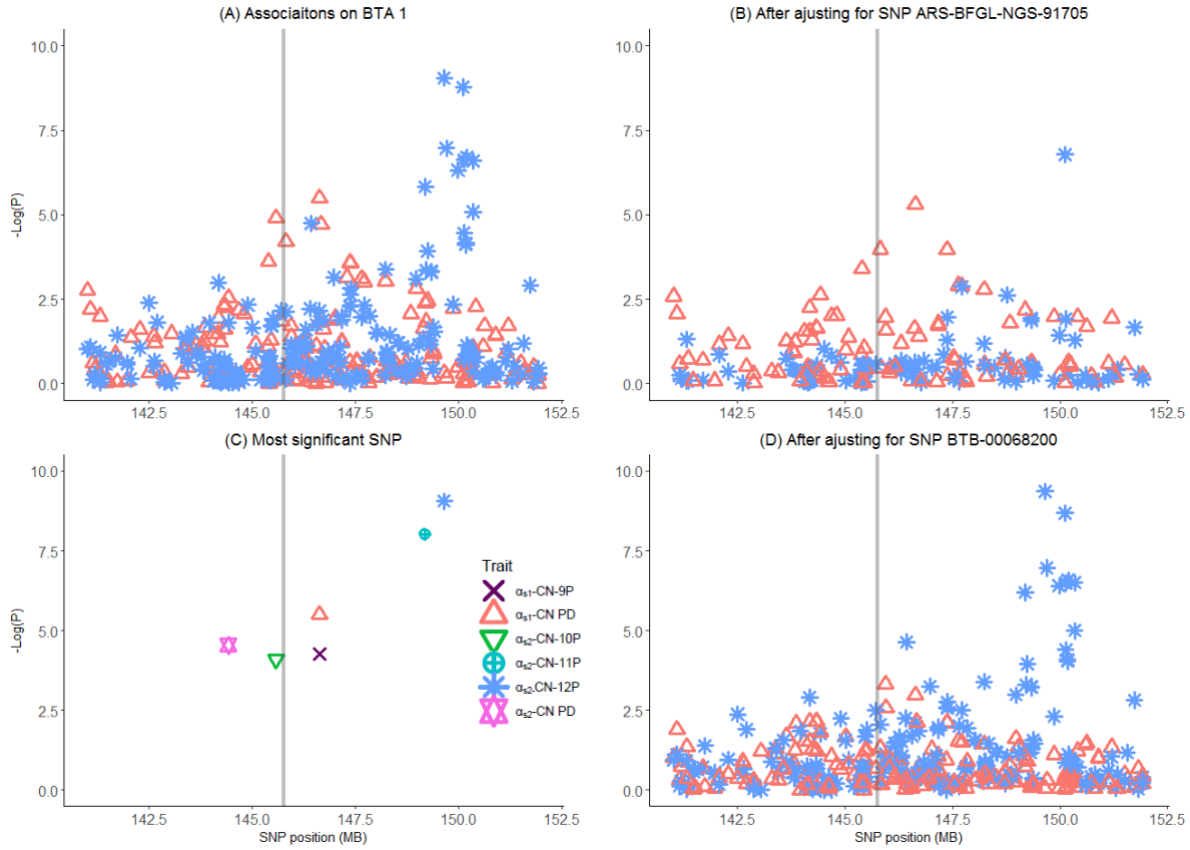


Figure 2. Significance [–$\log_{10}(P)$] of associations of SNP between 141 and 152.5 Mbp on BTA1 with (A) α_{s2}-CN-12P and the phosphorylation degree of α_{s1}-CN (α_{s1}-CN PD), (B) after including SNP ARS-BFGL-NGS-91705 genotypes as a fixed effect, (C) showing only the most significant SNP for individual α_{s1}-CN and α_{s2}-CN phosphorylation isoforms, and the phosphorylation degrees of α_{s1}-CN (α_{s1}-CN PD) and α_{s2}-CN (α_{s2}-CN PD), and (D) after including SNP BTB-00068200 genotypes as a fixed effect. α_{s1}-CN PD = α_{s1}-CN-9P / (α_{s1}-CN-8P + α_{s1}-CN-9P) × 100; α_{s2}-CN PD = α_{s2}-CN-12P / (α_{s2}-CN-10P + α_{s2}-CN-11P + α_{s2}-CN-12P) × 100. P = phosphate group attached. The shaded region corresponds to the location of the *SLC37A1* gene.

Table 2. Effects of SNP ARS-BFGL-NGS-91705 (BTA1, QTL1), SNP BTB-00068200 (BTA1, QTL2), SNP ULGR_BTC-053514 (BTA6), β-LG (BTA11) and diacylglycerol acyltransferase 1 (DGAT1, BTA14) genotypes on relative concentrations and yields of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, the phosphorylation degrees (PD)[a] of α_{s1}- and α_{s2}-CN, protein and phosphorus (P) contents, and milk, protein and phosphorus yields measured on test-day morning milk samples from 1,857 Dutch Holstein Friesian cows (SE in parentheses).

	BTA1			**BTA1**			**BTA6**			**BTA11**			**BTA14**		
	ARS-BFGL-NGS-91705[b]			**BTB-00068200[c]**			**ULGR_BTC-053514**			**β-LG**			**DGAT1**		
Trait	**AA**	**GG**	**-Log(*P*)**	**AA**	**GG**	**-Log(*P*)**	**AA**	**GG**	**-Log(P)**	**AA**	**BB**	**-Log(*P*)**	**AA**	**KK**	**-Log(*P*)**
(% wt/wt)	n=821	n=131		n=50	n=1121		n=637	n=945		n=539	n=262		n=628	n=276	
α_{s1}-CN-8P	0.00 (0.05)	0.05 (0.10)	0.06^{NS}	-0.19 (0.15)	0.02 (0.06)	0.39^{NS}	-0.42 (0.11)	0.49 (0.06)	23.80^{***}	-0.32 (0.06)	0.41 (0.07)	17.20^{***}	-0.06 (0.06)	0.01 (0.07)	0.27^{NS}
α_{s1}-CN-9P	-0.03 (0.05)	0.06 (0.10)	0.19^{NS}	0.06 (0.14)	-0.20 (0.06)	2.94^{**}	-0.12 (0.11)	0.33 (0.05)	9.39^{***}	0.07 (0.05)	-0.12 (0.07)	1.14^{NS}	0.53 (0.05)	-0.44 (0.06)	43.16^{***}
α_{s2}-CN-10P	0.02 (0.02)	-0.06 (0.04)	0.87^{NS}	-0.03 (0.05)	0.09 (0.02)	4.13^{***}	0.03 (0.04)	-0.08 (0.02)	4.42^{***}	-0.06 (0.02)	0.09 (0.03)	5.33^{***}	-0.12 (0.02)	0.13 (0.02)	19.47^{***}
α_{s2}-CN-11P	0.08 (0.03)	-0.17 (0.05)	5.71^{***}	-0.14 (0.08)	0.11 (0.03)	4.59^{***}	0.20 (0.05)	-0.33 (0.03)	38.49^{***}	-0.12 (0.03)	0.16 (0.04)	9.87^{***}	-0.23 (0.03)	0.14 (0.03)	23.97^{***}
α_{s2}-CN-12P	0.05 (0.01)	-0.07 (0.02)	8.51^{***}	-0.03 (0.03)	-0.00 (0.01)	0.16^{NS}	0.13 (0.02)	-0.14 (0.01)	46.04^{***}	-0.00 (0.01)	-0.00 (0.02)	0.07^{NS}	0.01 (0.01)	-0.04 (0.01)	1.50^{*}
α_{s1}-CN PD[a]	-0.11 (0.12)	0.20 (0.22)	0.39^{NS}	0.40 (0.33)	-0.55 (0.13)	4.57^{***}	-0.10 (0.26)	0.46 (0.13)	3.25^{***}	0.42 (0.13)	-0.60 (0.16)	6.62^{***}	1.32 (0.11)	-1.12 (0.15)	21.46^{***}
α_{s2}-CN PD[a]	0.05 (0.13)	0.23 (0.24)	0.21^{NS}	0.12 (0.36)	-0.62 (0.14)	4.44^{***}	-0.04 (0.28)	0.39 (0.13)	2.22^{**}	0.49 (0.13)	-0.89 (0.17)	10.73^{***}	0.94 (0.13)	-0.88 (0.16)	23.45^{***}
Yield (g)	n=760	n=119		n=45	n=1045		n=81	n=868		n=492	n=245		n=578	n=250	
α_{s1}-CN-8P	0.21 (0.93)	-4.53 (1.76)	1.55^{*}	-0.12 (2.69)	1.66 (1.01)	0.60^{NS}	-1.93 (2.05)	0.72 (1.00)	0.29^{NS}	-0.89 (0.98)	1.21 (1.25)	0.48^{NS}	0.18 (0.96)	-3.76 (1.24)	2.21^{**}
α_{s1}-CN-9P	-0.05 (0.36)	-1.25 (0.68)	0.65^{NS}	0.19 (1.04)	-0.17 (0.40)	0.05^{NS}	-0.47 (0.79)	0.96 (0.38)	1.46^{*}	0.60 (0.38)	-0.80 (0.48)	1.4^{*}	2.68 (0.36)	-3.17 (0.46)	28.95^{***}
α_{s2}-CN-10P	0.10 (0.11)	-0.59 (0.20)	2.31^{**}	-0.32 (0.31)	0.42 (0.12)	3.34^{***}	0.14 (0.23)	-0.45 (0.11)	4.28^{***}	-0.26 (0.11)	0.35 (0.15)	2.99^{**}	-0.60 (0.11)	0.40 (0.14)	10.66^{***}
α_{s2}-CN-11P	0.42 (0.21)	-1.64 (0.40)	5.23^{***}	-0.46 (0.61)	0.74 (0.23)	2.52^{**}	0.88 (0.45)	-1.78 (0.22)	16.86^{***}	-0.55 (0.23)	0.55 (0.29)	2.57^{**}	-1.03 (0.22)	-0.08 (0.28)	5.05^{***}
α_{s2}-CN-12P	0.27 (0.11)	-0.83 (0.21)	5.78^{***}	-0.10 (0.32)	0.18 (0.12)	0.55^{NS}	0.60 (0.24)	-0.81 (0.12)	13.71^{***}	-0.03 (0.12)	-0.14 (0.15)	0.23^{NS}	0.06 (0.11)	-0.60 (0.15)	4.24^{***}
Production and P	n=760	n=119		n=45	n=1045		n=85	n=945		n=492	n=245		n=578	n=250	
Protein content (%)	-0.01 (0.01)	-0.03 (0.03)	0.29^{NS}	-0.03 (0.04)	0.03 (0.02)	1.12^{NS}	0.02 (0.03)	-0.09 (0.01)	8.25^{***}	-0.02 (0.01)	0.02 (0.02)	0.75^{NS}	-0.15 (0.01)	0.11 (0.02)	44.41^{***}
P content (mg/kg)	4.92 (4.58)	-9.29 (8.60)	0.62^{NS}	-18.35 (12.61)	42.63 (4.96)	17.40^{***}	1.20 (9.99)	-8.90 (4.88)	0.77^{NS}	-3.93 (4.90)	2.90 (6.26)	0.22^{NS}	-47.90 (4.50)	38.66 (5.81)	41.94^{***}
P_{CN} content[d] (mM)	-0.02 (0.04)	-0.61 (0.07)	0.18^{NS}	-0.16 (0.11)	0.08 (0.04)	1.44^{*}	0.04 (0.08)	-0.22 (0.04)	7.40^{***}	-0.16 (0.04)	0.15 (0.05)	6.49^{***}	-0.34 (0.04)	0.24 (0.05)	29.65^{***}
Milk yield (kg)	0.06 (0.13)	-0.49 (0.24)	1.05^{NS}	0.19 (0.36)	0.11 (0.14)	0.16^{NS}	-0.07 (0.28)	0.13 (0.14)	0.19^{NS}	0.15 (0.13)	-0.19 (0.17)	0.65^{NS}	0.65 (0.13)	-0.88 (0.17)	15.48^{***}
Protein yield (kg)	0.00 (0.00)	-0.02 (0.01)	1.91^{*}	0.00 (0.01)	0.01 (0.00)	0.51^{NS}	0.00 (0.01)	-0.01 (0.00)	0.57^{NS}	0.00 (0.00)	-0.00 (0.01)	0.30^{NS}	0.00 (0.00)	-0.02 (0.01)	2.65^{**}
P yield (mg)	121 (132)	-712 (248)	2.30^{**}	47 (379)	650 (143)	4.56^{***}	-23 (286)	-13 (140)	0.00^{NS}	104 (140)	-130 (177)	0.30^{NS}	-6 (137)	-448 (178)	1.46^{*}
P_{CN} yield[e] (g)	0.20 (0.99)	-5.18 (1.00)	1.80^{*}	-0.12 (2.85)	1.77 (1.07)	0.61^{NS}	-0.21 (2.18)	-1.96 (1.06)	0.78^{NS}	-0.80 (1.05)	0.57 (1.32)	0.21^{NS}	0.36 (1.02)	-4.44 (1.32)	2.83^{**}

[a] α_{s1}-CN PD = α_{s1}-CN-9P / (α_{s1}-CN-8P + α_{s1}-CN-9P) × 100; α_{s2}-CN PD = α_{s2}-CN-12P / (α_{s2}-CN-10P + α_{s2}-CN-11P + α_{s2}-CN-12P) × 100. P = phosphate group attached.

[b]ARS-BFGL-NGS-91705 (rs43282015) is the lead SNP of α_{s2}-CN-12P concentration (%wt/wt) in the QTL region between 147.5 and 152.1 Mbp on BTA1.

[c]BTB-00068200 (rs43281569) is the lead SNP of α_{s2}-CN-10P concentration (%wt/wt) in the QTL region between 144.4 and 147.3 Mbp on BTA1.

[d]P_{CN} content= $\Sigma[\frac{\text{concentration of individual casein fraction in milk (g/L; \% wt/wt} \times \text{protein percentage} \times 10)}{\text{molecular weight (Da) of respective individual casein fraction}} \times$ number of phosphate groups attached to the respective casein fraction].

[e]P_{CN} yield= $\{\Sigma[\frac{\text{concentration of individual casein fraction in milk (g/L; \% wt/wt} \times \text{protein percentage} \times 10)}{\text{molecular weight (Da) of respective individual casein fraction}} \times$ number of phosphate groups attached to the respective casein fraction]$\}$ × milk yield.

NS = $P \geqq 0.05$, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$