J. Dairy Sci. 100:9085–9102
https://doi.org/10.3168/jds.2017-13219

Genome-wide association and pathway-based analysis using latent variables related to milk protein composition and cheesemaking traits in dairy cattle

Christos Dadousis,* Sara Pegolo,* Guilherme J. M. Rosa,† Giovanni Bittante,* and Alessio Cecchinato*[1]
*Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy
†Department of Animal Sciences and Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison 53706

ABSTRACT

The aim of this study was to perform genome-wide associations (GWAS) and gene-set enrichment analyses with protein composition and cheesemaking-related latent variables (factors; F) in a cohort of 1,011 Italian Brown Swiss cows. Factor analysis was applied to identify latent structures of 26 phenotypes related to bovine milk quantity and quality, protein fractions [α_{S1}-, α_{S2}-, β-, and κ-casein (CN), β-lactoglobulin, and α-lactalbumin (α-LA)], coagulation and curd firming at time t (CF_t) measures, and cheese properties [cheese yield (%CY) and nutrients recovery in the curd] of individual cows. Ten orthogonal F were extracted, explaining 74% of the original variability. Factor $1_{\%CY}$ underlined the %CY characteristics, $F2_{CFt}$ was related to the CF_t process parameters, $F3_{Yield}$ was considered as descriptor of milk and solids yield, whereas $F4_{Cheese\ N}$ underscored the presence of nitrogenous compounds (N) into the cheese. Four more F were related to the milk caseins ($F5_{\alpha S1\text{-}\beta\text{-}CN}$, $F7_{\beta\text{-}\kappa\text{-}CN}$, $F8_{\alpha S2\text{-}CN}$, and $F9_{\alpha S1\text{-}CN\text{-}Ph}$) and 1 F was linked to the whey protein ($F10_{\alpha\text{-}LA}$); 1 F underlined the udder health status ($F6_{Udder\ health}$). All cows were genotyped with the Illumina BovineSNP50 Bead Chip v.2 (Illumina Inc., San Diego, CA). Single marker regression GWAS were fitted. Gene-set enrichment analysis was run on GWAS results, using the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway databases, to reveal ontologies or pathways associated with the F. All F but $F3_{Yield}$ showed significance in GWAS. Signals in 10 *Bos taurus* autosomes (BTA) were detected. High peaks on BTA6 (~87 Mbp) were found for $F6_{\beta\text{-}\kappa\text{-}CN}$, $F5_{\alpha S1\text{-}\beta\text{-}CN}$, and at the tail of BTA11 (~104 Mbp) for $F4_{Cheese\ N}$. Gene-set enrichment analyses showed significant results (false discovery rate at 5%) for $F8_{\alpha S2\text{-}CN}$, $F1_{\%CY}$, $F4_{Cheese\ N}$, and $F10_{\alpha\text{-}LA}$. For $F8_{\alpha S2\text{-}CN}$, 33 Gene Ontology terms and 3 Kyoto Encyclopedia of Genes and Genomes categories were enriched, including terms related to ion transport and homeostasis, neuron function or part, and GnRH signaling pathway. Our results support the feasibility of factor analysis as a dimension reduction technique in genomic studies and evidenced a potential key role of α_{S2}-CN in milk quality and composition.

Key words: factor analysis, milk protein, cheesemaking, GWAS, gene-set enrichment

INTRODUCTION

Cheese production has a relevant economic and social importance, being the primary use of bovine milk produced in many countries worldwide. Cheese manufacturing involves a complex biological process comprising many interrelated factors, such as milk components (e.g., fat, protein, and minerals), milk acidity and microbial flora, milk coagulation properties (**MCP**) and curd firmness (**CF**) properties, among others. Recent studies revealed an important role of animal genetics in regulating bovine cheese yield (**CY**), encouraging breeding strategies for an increased CY (Bittante et al., 2013a). Moreover, specific chromosomal regions and biological pathways associated with CY, MCP and CF properties have been detected (Dadousis et al., 2016; Dadousis et al., 2017a,b).

Animal breeding programs aim at the simultaneous improvement of several traits across generations. To achieve this, a detailed recording system is required at the population level. However, the large number of traits of interest and their complex phenotypic and genetic correlation structure pose challenges to the selection decision process, as well as to data analyses and computations. From a data analysis standpoint, several dimension reduction techniques can be used, such as factor analysis (**FA**), which is commonly adopted to identify latent structures [factors (**F**)] of correlated variables. Based on the observed covariance structure, the objective of FA is to replace n measured variables with p ($p < n$) F, where the measured variables are

Received May 23, 2017.
Accepted July 16, 2017.
[1]Corresponding author: alessio.cecchinato@unipd.it

expressed as linear functions of the F, and the F capture the underlying latent concept that the original variables represent (Bollen, 2014). In dairy cattle, the potential use of F obtained from FA has been investigated for a variety of traits, including milk quality, milk technological properties [e.g., MCP and CY-related phenotypes (Macciotta et al., 2012)], type traits (Kern et al., 2014), as well as milk fatty acids (Conte et al., 2016; Mele et al., 2016). However, previous studies were focused on the sources of variation related to the F and their genetic parameters.

Genomics has long been recognized as a valuable tool in dairy cattle genetics and breeding programs, especially on the use of molecular markers in genome-wide association studies (**GWAS**) and genomic selection (Meuwissen et al., 2001). In the context of GWAS, quite often each trait is analyzed separately from each other. However, in the case of complex traits (e.g., CY), a plethora of different and possibly correlated components might be involved (Cecchinato and Bittante, 2016). Simulation studies found that integration of (correlated) phenotypes into a multivariate GWAS model might lead to an increased power for detecting causal loci compared with the classical univariate analysis (Galesloot et al., 2014). Furthermore, the replacement of the original (possibly correlated) phenotypes with a smaller set of linearly uncorrelated variables (i.e., principal components) has been also investigated. In particular, the use of traits reduction methods such as principal component analysis coupled with GWAS has been recently explored for production and functional traits in sheep and dairy cattle (Kominakis et al., 2017; Macciotta et al., 2017). Nevertheless, although principal component analysis is considered as a useful tool for data exploration, FA is preferable when the goal is to detect the structure underlying the variables (i.e., latent structure; Jolliffe, 2002).

To complement GWAS studies, it is becoming common the use of gene-set enrichment and pathway analyses. Such an approach helps to alleviate problems related to GWAS (e.g., GWAS ignores the fact that genes work together in networks in the various biological pathways), and to deepen the understanding of the biological pathways affecting quantitative traits (Gambra et al., 2013; Abdalla et al., 2016; Iso-Touru et al., 2016). Integration of F, GWAS, and pathways analyses might address some aforementioned issues and has been already used in human studies (Fanous et al., 2012), whereas its potential application in livestock breeding and genetics remains still unexplored. In addition, studies are available that performed GWAS (Schopen et al., 2011; Bijl et al., 2014; Buitenhuis et al., 2016) or GWAS plus pathway analysis (even if limited to 164 lactating cows; Gambra et al., 2013) on milk protein fractions; however, these phenotypes have been never considered in combination with milk technological traits to represent the complexity of cheesemaking process. Therefore, our objective was to conduct GWAS combined with gene ontology (**GO**) and pathway analysis using a set of latent variables obtained from 26 phenotypes related to milk yield and quality, protein composition, curd firming, and individual cheese properties in a sample of Brown Swiss cows genotyped with a 50k SNP chip.

MATERIALS AND METHODS

Animals and Sampling

A detailed description of the sampling procedure has been previously reported (Cipolat-Gotet et al., 2012). In brief, milk and blood samples from 1,264 Italian Brown Swiss cattle belonging to 85 herds were collected during evening milking. Within any given day, only 1 herd was sampled. One milk subsample per cow, immediately refrigerated after collection at 4°C without any preservative, was transported to the Milk Quality Laboratory of the Breeders Federation of Trento Province (**BFTP**; Trento, Italy) for composition analysis. All milk samples were collected within the standard milk recording schemes coordinated by technicians working at the BFTP. Additional data on the cows and herds were provided by the BFTP. In total, 29 cheese-related phenotypes were measured in the Cheese-Making Laboratory of the University of Padova and included in the analyses.

Phenotypic Data

Milk Quality and Protein Composition. Individual milk samples were analyzed for fat, protein, and lactose contents using MilkoScan FT6000 (Foss, Hillerød, Denmark). The pH analysis was carried out using a Crison Basic 25 electrode (Crison, Barcelona, Spain). Somatic cell count data were determined by a Fossomatic FC counter (Foss) and SCS were obtained through logarithmic transformation [$\log_2$(SCC/100,000) + 3; Ali et al., 1980]. Casein fractions (α_{S1}-, α_{S2}-, β-, and κ-CN) and whey proteins (β-LG and α-LA) were measured using a validated reversed-phase HPLC method (Bonfatti et al., 2008). Each fraction was expressed as the ratio to the total milk nitrogen content. Moreover, the phosphorylated form of the α_{S1}-CN was obtained by the methodology proposed by Bonfatti et al. (2011). The remaining milk N compounds were estimated as difference from the total milk nitrogen content.

Curd Firming Parameters. Six parameters related to curd firming at time t ($\mathbf{CF_t}$) and derived from the CF modeling (Bittante et al., 2013b) were included

in our analysis: rennet coagulation time ($\mathbf{RCT_{eq}}$, min), maximum curd firmness ($\mathbf{CF_{max}}$, mm) and time to reach CF_{max} ($\mathbf{t_{max}}$, min), potential asymptotical curd firmness in the absence of syneresis ($\mathbf{CF_P}$, mm), and the rate constants of curd-firming ($\mathbf{k_{CF}}$, %/min) and syneresis ($\mathbf{k_{SR}}$, %/min). Due to convergence problems, CF_P was expressed proportionally to the CF_{max}, multiplying CF_{max} by 1.34. This value is the regression coefficient resulting from the linear regression of CF_P on CF_{max} (Stocco et al., 2017). The 3 CF_t model parameters (RCT_{eq}, k_{CF}, and k_{SR}) were obtained through curvilinear regression (PROC NLIN; SAS Institute Inc., Cary, NC).

Individual CY and Curd Nutrient Recovery. Individual cow cheese phenotypes, obtained through a model cheesemaking procedure (Cipolat-Gotet et al., 2013), were included in the analysis. Individual CY, expressed as percentage of the weight of the total milk processed, comprised the weight of the curd DM ($\mathbf{\%CY_{SOLIDS}}$) and water ($\mathbf{\%CY_{WATER}}$) as well as their sum (fresh curd; $\mathbf{\%CY_{CURD}}$). Three additional traits related to the nutrients of the milk retained in the curd, calculated as the ratio (%) between the curd nutrient and the corresponding nutrient contained in the processed milk, were $\mathbf{REC_{SOLIDS}}$, $\mathbf{REC_{FAT}}$, and $\mathbf{REC_{PROTEIN}}$. Finally, the recovery of the energy within the curd ($\mathbf{REC_{ENERGY}}$), calculated as the ratio with the energy in the milk (NRC, 2001), was also obtained.

Genotyping

Genomic DNA was extracted from individual peripheral blood samples of 1,152 cows. Animals were genotyped with the Illumina BovineSNP50 v.2 BeadChip (Illumina Inc., San Diego, CA). Markers that did not fulfill the following criteria were excluded from the analysis: (1) call rate >95%, (2) minor allele frequency >0.5%, and (3) no extreme deviation from Hardy-Weinberg proportions ($P > 0.001$, Bonferroni corrected). After quality control, 1,011 cows and 37,568 SNP were retained.

Statistical Analysis

Factor Analysis. Before applying FA, 3 out of the 29 phenotypes (CF_{max}, $\%CY_{CURD}$, and REC_{SOLIDS}) were excluded to avoid severe multicollinearity problems: (1) $\%CY_{CURD}$ is the sum of $\%CY_{SOLIDS}$ and $\%CY_{WATER}$; (2) CF_{max} is proportional to CF_P; and (3) phenotypic correlation coefficients of REC_{SOLIDS} with REC_{ENERGY} were 0.93 (Bittante et al., 2013a). The following factor model was used to simultaneously analyze the remaining 26 phenotypic variables:

$$\boldsymbol{\chi} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta},$$

where $\boldsymbol{\chi}$ is a vector of the 26 phenotypes and $\boldsymbol{\xi}$ is the factor vector. The factor loadings, relating the factors to the original variables, are contained in $\boldsymbol{\Lambda}$, and $\boldsymbol{\delta}$ is the residual vector.

The Kaiser-Meyer-Olkin (**KMO**) measure of sampling adequacy was adopted to quantify the difference between partial and Pearson correlations of the 26 variables (Dziuban and Shirkey, 1974; Kaiser and Rice, 1974). The KMO is a commonly used criterion in FA to assess if the correlation between 2 variables is mediated by other variables. A high KMO value indicates the presence of a latent structure. Partial correlation coefficients were calculated using the *corpcor* package in R (Schäfer and Strimmer, 2005); furthermore, exploratory FA was applied. To identify simple structure, a *varimax* factor rotation was used. The criteria used to extract the factors were prior knowledge, biological interpretation, and percentage of original variance explained by the F. To explain the F, a threshold of factor loadings >|0.4| was considered as significant (Fanous et al., 2012). The FA was implemented using the *psych* package (Revelle, 2017) in the R environment.

GWAS. A single marker regression was fitted for GWAS using the GenABEL package in R (Aulchenko et al., 2007) and the GRAMMAR-GC (Genome-Wide Association using Mixed Model and Regression-Genomic Control) approach, with the default function gamma (Amin et al., 2007; Svishcheva et al., 2012). The GRAMMAR-GC consists of 3 steps. First, an additive polygenic model with a genomic relationship matrix is fitted; then the obtained residuals of this model are regressed on SNP to test for associations; and, finally, the genomic control corrects for conservativeness of the procedure (Svishcheva et al., 2012). The polygenic model was

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathrm{a} + \mathrm{e},$$

where $\mathbf{y}$ is a vector containing the latent variables; $\boldsymbol{\beta}$ is a vector with the fixed effects of (1) DIM of the cow (classes of 30 d each), (2) parity level of each cow (with classes 1, 2, 3, ≥4), (3) the effect of the instrument detector (considered only for the CF_t measures), and (4) herd-date effect (n = 85); $\mathbf{X}$ is an incidence matrix connecting each observation to specific levels of factors in $\boldsymbol{\beta}$. The nongenetic effects have been previously studied in the same data set (Bittante et al., 2013a; Cipolat-Gotet et al., 2013). The 2 random terms in the model were the animal (a) and the residual effects (e), which were assumed to be normally distributed as $\mathrm{a} \sim N\left(0, \mathbf{G}\sigma_g^2\right)$ and $\mathrm{e} \sim N\left(0, \mathbf{I}\sigma_e^2\right)$, where $\mathbf{G}$ and $\mathbf{I}$ are the genomic rela-

tionship matrix and an identity matrix of appropriate order, respectively, and σ_g^2 and σ_e^2 are the additive genetic and the residual variances, respectively. The **G** matrix was constructed within the GenABEL R package using identical by state coefficients. A threshold of *P*-value equal to 5×10^{-5} was adopted to declare significance (Burton et al., 2007). Manhattan plots were drawn using the *qqman* R package (Turner, 2014).

Gene-Set Enrichment and Pathway-Based Analysis. Nominal *P*-values <0.05 obtained from the GWAS were used as threshold to split the SNP into 2 groups for each factor. The SNP were assigned to genes if they were located within the gene or in a flanking region of 15 kb up- and downstream of the gene (Pickrell et al., 2010) using the *biomaRt* R package (Durinck et al., 2005, 2009). For mapping, the Ensembl *Bos taurus* UMD3.1 assembly was used as reference (Zimin et al., 2009). In the enrichment analysis, the total SNP tested in GWAS represented the background SNP, whereas the background genes were the genes associated with those SNP. The Kyoto Encyclopedia of Genes and Genomes (**KEGG**; Ogata et al., 1999) and the GO (Ashburner et al., 2000) databases were queried to assign the genes to functional categories. The KEGG database contains regulatory and metabolic pathways, signifying the knowledge on molecular interactions and reaction networks. The GO database entitles biological descriptors (GO terms) to genes based on features of the gene-encoded products. The GO database is partitioned into 3 classes, namely biological process (**BP**), molecular function (**MF**), and cellular component (**CC**). To avoid testing broad or narrow functional categories, GO and KEGG terms with <10 and >1,000 genes were excluded from the analysis. For each functional category, a Fisher's exact test was applied to test for overrepresentation of the significant genes. To account for multiple testing, a false discovery rate correction was used (controlled at 5%). The gene-set enrichment analysis was carried out using the *goseq* package in the R environment (Young et al., 2010).

RESULTS

Extraction of Factors

Summary statistics for all 29 measured phenotypes are shown in Table 1. Marginal and partial correlations among the 26 variables used to estimate the KMO are presented in Supplemental Figure S1 (https://doi.org/10.3168/jds.2017-13219). The average KMO value in our data set was 0.55. Ten F were extracted and kept for further analysis; the F explained 74% of the original variability. The F loadings with their given names (sorted by maximum variance explained) are shown in Table 2.

The first F, in order of explained variance, was primarily loaded on $\%CY_{SOLIDS}$, and then on fat and protein (%) and REC_{ENERGY}; therefore, it was considered as a F representing the quantity of cheese obtained from a given amount of milk processed (**F1$_{\%CY}$**). The second F was associated with all CF_t phenotypes, except CF_P, and also to REC_{FAT} underlying the curd firmness process (**F2$_{CFt}$**) and its importance for fat recovery in milk. In particular, positive loadings on the curd firming and syneresis rate constants (k_{CF} and k_{SR}) were detected, whereas negative relationships with the

Table 1. Summary statistics of milk (yield and quality), protein fractions, curd firming, and cheesemaking (%CY and REC) phenotypes

Trait[1]	Mean	CV (%)
Milk trait		
Milk yield (kg/d)	24.95	31
Fat yield (kg/d)	1.09	37
Protein yield (kg/d)	0.92	30
Fat (%)	4.37	20
Protein (%)	3.71	11
Lactose (%)	4.86	4
pH	6.64	1
SCS	2.87	65
Milk protein fraction (%)		
α_{S1}-CN	25.69	7
α_{S1}-CN-Ph	1.45	42
α_{S2}-CN	9.20	12
β-CN	32.26	8
κ-CN	9.44	16
β-LG	8.68	18
α-LA	2.39	21
Other N compounds	10.89	21
Curd firming		
RCT_{eq} (min)	20.96	29
CF_P (mm)	49.20	20
k_{CF} (%/min)	12.90	32
k_{SR} (%/min)	1.23	37
CF_{max} (mm)	36.91	20
t_{max} (min)	41.83	30
Cheese yield (CY, %)		
$\%CY_{CURD}$	14.95	12
$\%CY_{SOLIDS}$	7.17	13
$\%CY_{WATER}$	7.77	16
Nutrient recovery (REC, %)		
REC_{SOLIDS}	51.80	7
REC_{FAT}	89.75	4
$REC_{PROTEIN}$	78.16	3
REC_{ENERGY}	67.15	5

[1]SCS = calculated as $\log_2(SCC \times 100{,}000) + 3$. Milk protein fractions: Ph = phosphorylated form. Curd firming (CF): RCT_{eq} = estimated rennet coagulation time; CF_P = asymptotical potential value of CF; k_{CF} = curd-firming instant rate constant; k_{SR} = syneresis instant rate constant; CF_{max} = maximum curd firmness achieved within 90 min; and t_{max} = time at achievement of CF_{max}. %CY = ratios of the weight (g) of the fresh curd ($\%CY_{CURD}$), curd DM ($\%CY_{SOLIDS}$), and curd water ($\%CY_{WATER}$) versus the weight of the processed milk (g); REC = ratio of the weight (g) of the curd constituent (DM, fat, protein or energy, respectively) versus that of the same constituent in the processed milk (g).

Table 2. Rotated factor (F) pattern, F name, communality (com)[1] of variables, and cumulative variance[2] explained by the F[3]

Phenotype	$F1_{\%CY}$	$F2_{CFt}$	$F3_{Yield}$	$F4_{Cheese\ N}$	$F5_{\alpha S1\text{-}\beta\text{-}CN}$	$F6_{Udder\ health}$	$F7_{\kappa\text{-}\beta\text{-}CN}$	$F8_{\alpha S2\text{-}CN}$	$F9_{\alpha S1\text{-}CN\text{-}Ph}$	$F10_{\alpha\text{-}LA}$	com
Milk trait											
Milk yield (kg/d)	−0.19	0.08	**0.96**	0.00	0.03	0.11	0.01	0.01	0.02	0.08	0.99
Fat yield (kg/d)	0.28	0.08	**0.89**	−0.02	0.06	0.17	−0.01	−0.01	−0.01	0.07	0.92
Protein yield (kg/d)	0.00	0.00	**0.97**	−0.04	0.03	0.02	0.02	0.05	−0.02	0.03	0.96
Fat (%)	**0.90**	0.01	0.07	−0.04	0.06	0.08	−0.06	−0.03	−0.05	0.00	0.84
Protein (%)	**0.59**	−0.22	−0.11	−0.14	0.02	−0.30	0.02	0.08	−0.07	−0.17	0.56
Lactose (%)	−0.07	0.01	0.08	0.05	−0.01	**0.62**	−0.01	0.00	0.03	0.04	0.40
pH	−0.08	−0.31	0.00	0.11	−0.13	−0.02	0.03	0.04	0.17	0.15	0.18
SCS	0.06	−0.02	−0.08	0.04	−0.05	**−0.41**	0.09	0.01	0.03	−0.09	0.20
Milk protein fraction (%)											
α_{S1}-CN	0.04	0.01	0.07	−0.14	**0.94**	0.25	−0.04	−0.08	−0.14	−0.04	0.99
α_{S1}-CN-Ph	0.04	−0.02	0.00	0.09	−0.10	0.01	−0.04	−0.04	**0.98**	0.06	1.00
α_{S2}-CN	0.02	−0.08	0.04	−0.06	0.01	−0.03	−0.07	**0.98**	−0.04	0.14	1.00
β-CN	−0.10	−0.05	−0.11	0.12	**−0.70**	0.38	**−0.47**	−0.33	−0.03	−0.05	1.00
κ-CN	0.12	0.17	0.00	−0.09	0.05	−0.11	**0.96**	−0.09	−0.04	0.01	1.00
β-LG	0.05	−0.03	0.02	**−0.98**	0.12	0.01	0.07	−0.05	−0.10	0.00	0.99
α-LA	−0.06	−0.02	0.18	−0.14	−0.02	0.30	0.01	0.17	0.07	**0.90**	0.99
Other N compounds	−0.05	0.01	−0.02	**0.76**	−0.09	**−0.60**	−0.10	−0.01	−0.02	−0.21	1.00
Curd firming											
RCT_{eq} (min)	0.01	**−0.74**	−0.05	−0.06	−0.01	−0.15	0.02	0.08	0.00	−0.01	0.59
CF_P (mm)	0.38	0.03	0.01	−0.08	0.23	−0.11	0.23	−0.07	−0.11	−0.20	0.32
k_{CF} (%/min)	−0.01	**0.94**	0.00	0.00	−0.03	−0.09	0.08	0.00	0.00	−0.01	0.90
k_{SR} (%/min)	−0.06	**0.88**	0.00	−0.01	−0.05	−0.08	0.06	0.00	0.01	0.00	0.79
t_{max} (min)	0.04	**−0.90**	−0.04	−0.03	−0.01	−0.06	−0.05	0.03	0.00	−0.01	0.82
Cheese yield (%CY)											
$\%CY_{SOLIDS}$	**0.99**	0.02	0.04	0.00	0.04	−0.08	0.02	0.01	0.01	−0.03	0.99
$\%CY_{WATER}$	0.39	−0.05	−0.07	0.11	−0.08	−0.03	0.05	0.01	0.08	0.02	0.19
Nutrient recovery (REC, %)											
REC_{FAT}	0.28	**0.44**	0.19	0.01	0.11	0.00	0.21	0.10	0.08	0.12	0.39
$REC_{PROTEIN}$	0.23	−0.02	−0.02	**0.46**	−0.03	0.23	0.00	−0.14	0.03	−0.02	0.34
REC_{ENERGY}	**0.87**	0.25	0.14	0.10	0.08	−0.03	0.10	0.01	0.03	0.06	0.88
Cumulative variance (%)	0.14	0.27	0.38	0.45	0.51	0.56	0.61	0.66	0.7	0.74	

[1]Communality = the sum of the squared factor loadings per trait.

[2]Factors have been sorted based on proportion of variance explained. Values >|0.4| in bold.

[3]SCS = calculated as $\log_2(SCC/100{,}000) + 3$. Milk protein fractions: Ph = phosphorylated form. Curd firming (CF): RCT_{eq} = estimated rennet coagulation time; CF_P = asymptotical potential value of CF; k_{CF} = curd-firming instant rate constant; k_{SR} = syneresis instant rate constant; CF_{max} = maximum curd firmness achieved within 90 min; and t_{max} = time at achievement of CF_{max}. %CY = ratios of the weight (g) of the fresh curd ($\%CY_{CURD}$), curd DM ($\%CY_{SOLIDS}$), and curd water ($\%CY_{WATER}$) versus the weight of the processed milk (g); REC = ratio of the weight (g) of the curd constituent (DM, fat, protein or energy, respectively) versus that of the same constituent in the processed milk (g). $F1_{\%CY}$ = factor related to the percentage of individual cheese yield; $F2_{CFt}$ = factor related to the curd firmness; $F3_{Yield}$ = factor related to the milk yield; $F4_{Cheese\ N}$ = factor related to the milk nitrogen that is present into the cheese curd; $F5_{\alpha S1\text{-}\beta\text{-}CN}$ = factor related to the α_{S1}- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; $F6_{Udder\ health}$ = factor related to the udder health of a cow; $F7_{\kappa\text{-}\beta\text{-}CN}$ = factor related to the κ- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; $F8_{\alpha S2\text{-}CN}$ = factor related to the milk a_{S2}-CN, expressed as relative content to the total milk nitrogen; $F9_{\alpha S1\text{-}CN\text{-}Ph}$ = factor related to the milk α_{S1}–phosphorylated CN expressed as content to the total milk nitrogen; $F10_{\alpha\text{-}LA}$ = factor related to the milk α-LA.

time required for achieving milk coagulation and maximum curd firmness (RCT_{eq} and t_{max}) were observed. The subsequent F was associated with the daily milk production phenotypes, and thus named as milk yield factor (**$F3_{Yield}$**). Factor 4 was heavily and negatively associated with β-LG, whereas positively related to other N compounds in milk. Consequently, the F was considered as representative of the nitrogen found in the cheese (**$F4_{Cheese\ N}$**), as whey proteins (β-LG is the most representative) are mainly lost in whey. The next F was primarily and positively linked to α_{S1}-CN, but also to the β-CN (negatively); therefore, the fifth F was considered as representative of these 2 casein fractions (**$F5_{\alpha S1\text{-}\beta\text{-}CN}$**). Factor 6 was associated with lactose (positively) but also had a weaker and negative relation with the SCS and the remaining milk N compounds; therefore, the sixth F reflected the cow's udder health status (**$F6_{Udder\ health}$**). The subsequent factor was primarily associated with the κ-CN (positively) but also with the β-CN (negatively), and hence considered an indicator of the κ- and β-CN (**$F7_{\kappa\text{-}\beta\text{-}CN}$**). Finally, factors 8, 9, and 10 were each heavily loaded to only 1 trait and named accordingly [**$F8_{\alpha S2\text{-}CN}$**, **$F9_{\alpha S1\text{-}CN\text{-}Ph}$** (for phosphorylated α_{S1}-CN), and **$F10_{\alpha\text{-}LA}$**, respectively].

GWAS

The GWAS results of the 10 latent variables are summarized in Table 3. More details can be found in Supplemental Table S1 (https://doi.org/10.3168/jds.2017-13219). In total, 149 SNP were found significantly associated with at least 1 latent variable. Among them, 146 SNP were located on 10 chromosomes (1, 2, 6, 9, 10, 11, 19, 20, 25, and 27), whereas 3 others had unknown positions on the genome. All latent variables showed signals except $F5_{Yield}$. Shared signals among F were found. The strongest signals were detected on BTA6 (~87.4 Mbp) and BTA11 (~104.3 Mbp). More precisely, the marker Hapmap52348-rs29024684 located at 87,396,306 bp of BTA6 was significantly associated with $F7_{\kappa\text{-}\beta\text{-}CN}$ ($P = 9.81 \times 10^{-56}$). Near to this position, at 87,201,599 bp, marker Hapmap28023-BTC-060518 was strongly associated with $F5_{\alpha S1\text{-}\beta\text{-}CN}$ ($P = 2.84 \times 10^{-47}$). Moreover, $F5_{\alpha S1\text{-}\beta\text{-}CN}$ had another strong signal at 87,245,049 bp (Hapmap24184-BTC-070077; $P = 7.00 \times 10^{-45}$). Albeit at a weaker strength, both positions were also highly significant for $F8_{\alpha S2\text{-}CN}$ ($P = 8.34 \times 10^{-11}$ at 87,201,599 bp; $P = 1.67 \times 10^{-10}$ at 87,245,049 bp). The same was observed for $F9_{\alpha S1\text{-}CN\text{-}Ph}$, with $P = 3.86 \times 10^{-11}$ at 87,201,599 bp and $P = 7.80 \times 10^{-11}$ at 87,245,049 bp. All casein F showed signals on BTA6 in the region 6e (~77.2–89.1 Mbp; Figure 1b). On BTA11, marker ARS-BFGL-NGS-104610 (104,293,559 bp) was strongly linked to $F4_{Cheese\ N}$ ($P = 9.81 \times 10^{-26}$). On BTA 6, 11, 20, and 27, signals were distributed in more than 1 chromosomic region.

On BTA6, 8 subregions were detected overall (Table 3, Figure 1). In regions 6a (~40 Mbp), 6b (~46.6 Mbp), and 6c (~68.5 Mbp), relatively weak signals were detected for the factors $F6_{Udder\ health}$, $F2_{CFt}$, and $F8_{\alpha S2\text{-}CN}$, respectively. The region 6d (~71–74.6 Mbp) was associated with both $F8_{\alpha S2\text{-}CN}$ and $F7_{\kappa\text{-}\beta\text{-}CN}$. The denser region (6e) was found between ~77 and 89 Mbp and included 71 significant SNP. In this genomic area, all factors except $F10_{\alpha\text{-}LA}$ and $F3_{Yield}$ showed associations with a peak at ~87.4 Mbp corresponding to the marker Hapmap52348-rs29024684. Especially for $F7_{\kappa\text{-}\beta\text{-}CN}$, the proportion of additive genetic variance explained by this SNP reached 74.2%. In addition, the marker Hapmap28023-BTC-060518, located at ~87.2 Mbp, explained ~53% of the additive genetic variance for $F5_{\alpha S1\text{-}\beta\text{-}CN}$. For the aforementioned markers, the effects were considerably large, around 1 standard deviation from the mean (Table 4).

Close to region 6e, at ~90.7 to 92.6 Mbp (region 6f), 8 SNP were significant for $F8_{\alpha S2\text{-}CN}$, $F7_{\kappa\text{-}\beta\text{-}CN}$ and $F5_{\alpha S1\text{-}\beta\text{-}CN}$. Moreover, $F8_{\alpha S2\text{-}CN}$ and $F7_{\kappa\text{-}\beta\text{-}CN}$ were associated with a region at ~94.2 Mbp (region 6g). A relatively weak association, close to the significance threshold, was detected at ~114.2 Mbp for $F1_{\%CY}$ (region 6h).

Five distinct genomic regions were identified on BTA11 (Table 3, Figure 2). The regions at ~4.4, ~77.5, ~87.7, and ~97.8 Mbp were associated with $F6_{Udder\ health}$, $F9_{\alpha S1\text{-}CN\text{-}Ph}$, $F2_{CFt}$, and $F4_{Cheese\ N}$, respectively. In the range ~101.3 to 106.5 Mbp (region 11e), 18 significant SNP were detected for $F4_{Cheese\ N}$ and $F2_{\alpha\text{-}LA}$, with a peak at ~104.3 Mbp.

Apart from BTA6 and BTA11, significant associations were detected on other chromosomes, albeit at a weaker strength (Table 3). Two regions were detected on BTA20, at ~7.9 and ~46.7 Mbp. The first region was associated with $F10_{\alpha\text{-}LA}$ and the second with $F9_{\alpha S1\text{-}CN\text{-}Ph}$. Moreover, on BTA27, 2 chromosomic regions were detected. Although close to each other, they were associated with different F. More precisely, $F1_{\%CY}$ was associated with 1 marker at ~42.1 Mbp, whereas 3 SNP were linked to $F10_{\alpha\text{-}LA}$ in the range ~43.4 to 43.9 Mbp. The rest of the signals were 1 trait-1 factor associations and close to the significance threshold. Factor $7_{\kappa\text{-}\beta\text{-}CN}$ was associated wth BTA1 at ~90.1 Mbp. A weak signal on BTA2 at ~122.5 Mbp was detected for $F1_{\%CY}$. A SNP at ~36.8 Mbp on BTA9 was linked to $F5_{\alpha S1\text{-}\beta\text{-}CN}$. One marker at ~10.7 Mbp on BTA10 was associated with $F8_{\alpha S2\text{-}CN}$. At the beginning of BTA19 (~1.8 Mbp), a weak signal was detected for $F1_{\%CY}$. Finally, on BTA25, 1 marker was associated with $F6_{Udder\ health}$ at ~5.4 Mbp.

Table 3. Summary results of the genome-wide association analyses[1]

BTA	No. of SNP (signals)	Interval (Mbp)	*P*-value (range)	Top SNP	Top SNP location (bp)	Top SNP MAF	Trait[2]
0[3]	5 (3)	—	(4.66×10^{-5}, 9.47×10^{-17})	BTA-76907-no-rs	0	0.26	$F5_{\alpha S1\text{-}\beta\text{-}CN}$, **$F7_{\kappa\text{-}\beta\text{-}CN}$**, $F9_{\alpha S1\text{-}CN\text{-}Ph}$
1	1	—	4.19×10^{-5}	BTB-00041036	90,156,001	0.01	$F7_{\kappa\text{-}\beta\text{-}CN}$
2	1	—	4.43×10^{-5}	ARS-BFGL-NGS-101039	122,509,616	0.34	$F1_{\%CY}$
6a	3 (3)	39,503–40,378	(7.89×10^{-6}, 1.19×10^{-6})	Hapmap26618-BTC-070864	39,597,740	0.03	$F6_{\text{Udder health}}$
6b	1	—	1.07×10^{-5}	Hapmap23226-BTA-159656	46,599,570	0.24	$F2_{CFt}$
6c	1	—	2.18×10^{-5}	ARS-BFGL-NGS-111636	68,546,212	0.05	$F8_{\alpha S2\text{-}CN}$
6d	8 (7)	71,154–74,607	(2.87×10^{-5}, 1.12×10^{-7})	Hapmap29639-BTC-041962	71,350,048	0.02	$F7_{\kappa\text{-}\beta\text{-}CN}$, $F8_{\alpha S2\text{-}CN}$
6e	140 (71)	77,186–89,104	(4.89×10^{-5}, 9.81×10^{-56})	Hapmap52348-rs29024684	87,396,306	0.19	$F1_{\%CY}$, $F2_{CFt}$, $F4_{\text{Cheese N}}$, $F5_{\alpha S1\text{-}\beta\text{-}CN}$, $F6_{\text{Udder health}}$, **$F7_{\kappa\text{-}\beta\text{-}CN}$**, $F8_{\alpha S2\text{-}CN}$, $F9_{\alpha S1\text{-}CN\text{-}Ph}$
6f	10 (8)	90,730–92,579	(3.23×10^{-5}, 1.63×10^{-9})	Hapmap43045-BTA-76998	90,730,485	0.01	$F5_{\alpha S1\text{-}\beta\text{-}CN}$, $F7_{\kappa\text{-}\beta\text{-}CN}$, **$F8_{\alpha S2\text{-}CN}$**
6g	5 (3)	94,229–94,360	(3.77×10^{-6}, 1.36×10^{-8})	BTB-01687386	94,360,125	0.02	$F7_{\kappa\text{-}\beta\text{-}CN}$, **$F8_{\alpha S2\text{-}CN}$**
6h	1	—	4.85×10^{-5}	BTB-02092741	114,223,059	0.01	$F1_{\%CY}$
9	1	—	4.24×10^{-5}	BTA-21753-no-rs	36,790,663	0.01	$F5_{\alpha S1\text{-}\beta\text{-}CN}$
10	1	—	3.22×10^{-5}	Hapmap41952-BTA-73370	10,659,761	0.38	$F8_{\alpha S2\text{-}CN}$
11a	1	—	4.72×10^{-5}	BTB-01723556	4,419,032	0.06	$F6_{\text{Udder health}}$
11b	1	—	3.40×10^{-5}	ARS-BFGL-NGS-56195	77,493,775	0.04	$F9_{\alpha S1\text{-}CN\text{-}Ph}$
11c	12 (12)	85,367–88,214	(3.32×10^{-5}, 3.82×10^{-7})	BTA-110429-no-rs	87,670,344	0.42	$F2_{CFt}$
11d	5 (5)	94,687–97,845	(4.51×10^{-5}, 2.63×10^{-7})	Hapmap56906-rs29014970	97,844,929	0.31	$F4_{\text{Cheese N}}$
11e	21 (21)	101,301–106,543	(3.04×10^{-5}, 2.08×10^{-26})	ARS-BFGL-NGS-104610	104,293,559	0.45	**$F4_{\text{Cheese N}}$**, $F10_{\alpha\text{-LA}}$
19	1	—	4.32×10^{-5}	ARS-BFGL-NGS-102974	1,822,133	0.34	$F1_{\%CY}$
20a	1	—	8.02×10^{-6}	BTA-51080-no-rs	7,881,875	0.01	$F10_{\alpha\text{-LA}}$
20b	1	—	1.84×10^{-5}	Hapmap51592-BTA-41521	46,709,345	0.36	$F9_{\alpha S1\text{-}CN\text{-}Ph}$
25	1	—	3.12×10^{-6}	Hapmap31994-BTC-065943	5,385,729	0.14	$F6_{\text{Udder health}}$
27a	1	—	2.38×10^{-5}	ARS-BFGL-NGS-87845	42,118,037	0.03	$F1_{\%CY}$
27b	3 (3)	43,436–43,902	(4.84×10^{-5}, 2.84×10^{-5})	ARS-BFGL-NGS-24170	43,459,156	0.48	$F10_{\alpha\text{-LA}}$

[1]No. of SNP (signals) = number of the SNP significantly associated with the trait. In parentheses is the total number of significant signals per each genomic region. Interval = the region on the chromosome spanned among the significant SNP (in base pairs); *P*-value (range) = *P*-value of the highest significant SNP adjusted for genomic control and the range of the *P*-values when multiple SNP were significantly associated with 1 trait. Top SNP = the highest significant SNP per trait. Top SNP location (bp) = position of the highest significant SNP on the chromosome in base pairs on UMD3.1; Top SNP MAF = minor allele frequency (MAF) of the top SNP.

[2]$F1_{\%CY}$ = factor related to the percentage of individual cheese yield; $F2_{CFt}$ = factor related to the curd firmness; $F3_{Yield}$ = factor related to the milk yield; $F4_{\text{Cheese N}}$ = factor related to the milk nitrogen that is present into the cheese curd; $F5_{\alpha S1\text{-}\beta\text{-}CN}$ = factor related to the α_{S1}- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; $F6_{\text{Udder health}}$ = factor related to the udder health of a cow; $F7_{\kappa\text{-}\beta\text{-}CN}$ = factor related to the κ- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; $F8_{\alpha S2\text{-}CN}$ = factor related to the milk a_{S2}-CN, expressed as relative content to the total milk nitrogen; $F9_{\alpha S1\text{-}CN\text{-}Ph}$ = factor related to the milk α_{S1}–phosphorylated CN expressed as content to the total milk nitrogen; $F10_{\alpha\text{-LA}}$ = factor related to the milk α-LA. The trait with the highest *P*-value in each genomic region is bolded.

[3]Undefined chromosome and position on the genome.

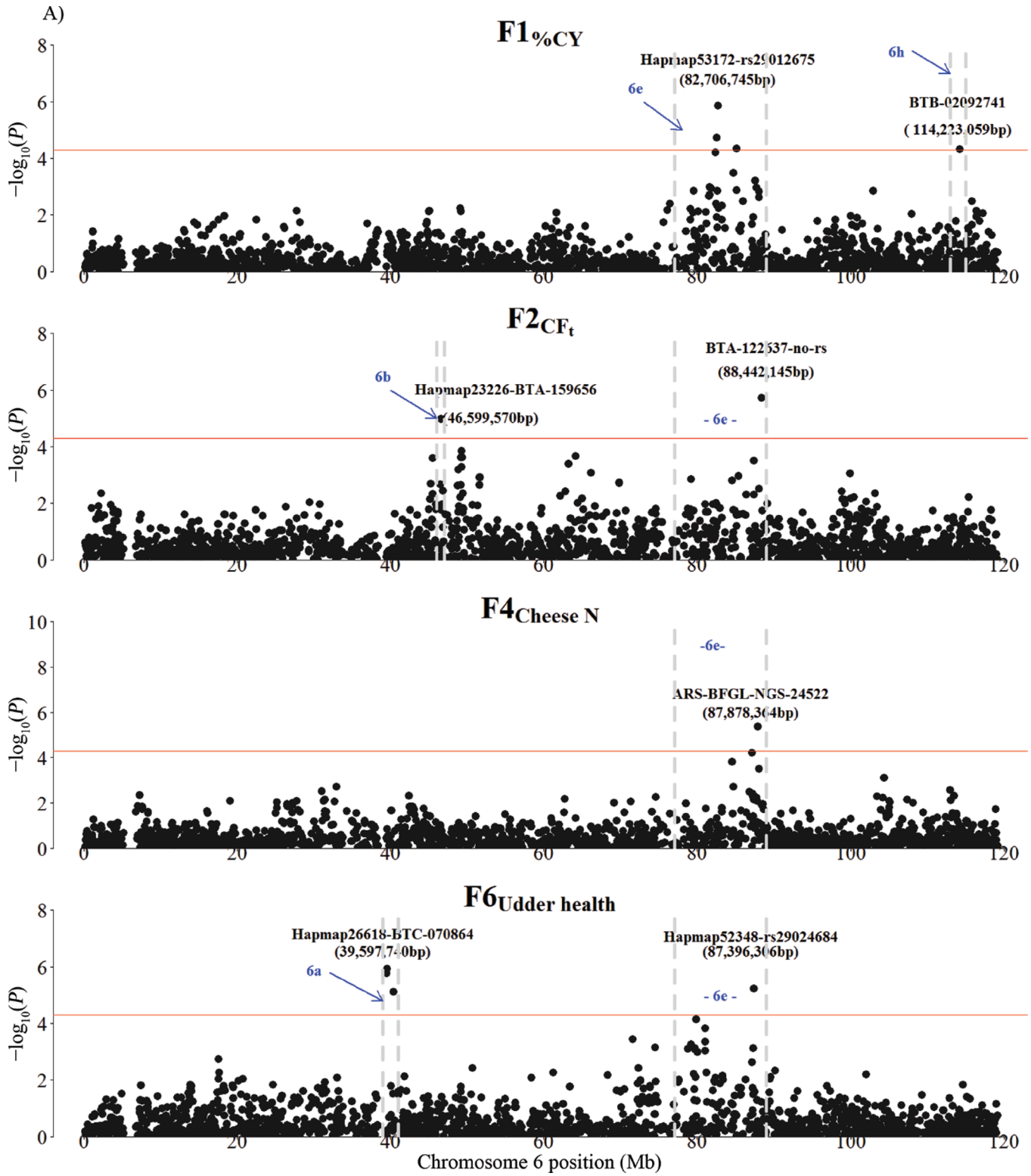

Figure 1. Manhattan plots of *P*-values from the genome-wide association study on BTA6. (A) $F1_{\%CY}$ = factor underlying the percentage of individual cheese yield; $F2_{CFt}$ = factor underlying the milk curd firmness; $F4_{Cheese\ N}$ = factor underlying the protein in the cheese; $F6_{Udder\ health}$ = factor underlying the udder health condition of a cow. (B) $F5_{\alpha S1\text{-}\beta\text{-}CN}$ = factor underlying the α_{S1}- and β-CN; $F7_{\kappa\text{-}\beta\text{-}CN}$ = factor underlying the κ- and β-CN; $F8_{\alpha S2\text{-}CN}$ = factor underlying the α_{S2}-CN; $F9_{\alpha S1\text{-}CN\text{-}Ph}$ = factor underlying the phosphorylated α_{S1}-CN. The red horizontal lines indicate a $-\log_{10}$(*P*-value) of 4.30 (corresponding to *P*-value = 5×10^{-5}). The highest significant marker on BTA6 per trait is also presented. Color version available online.

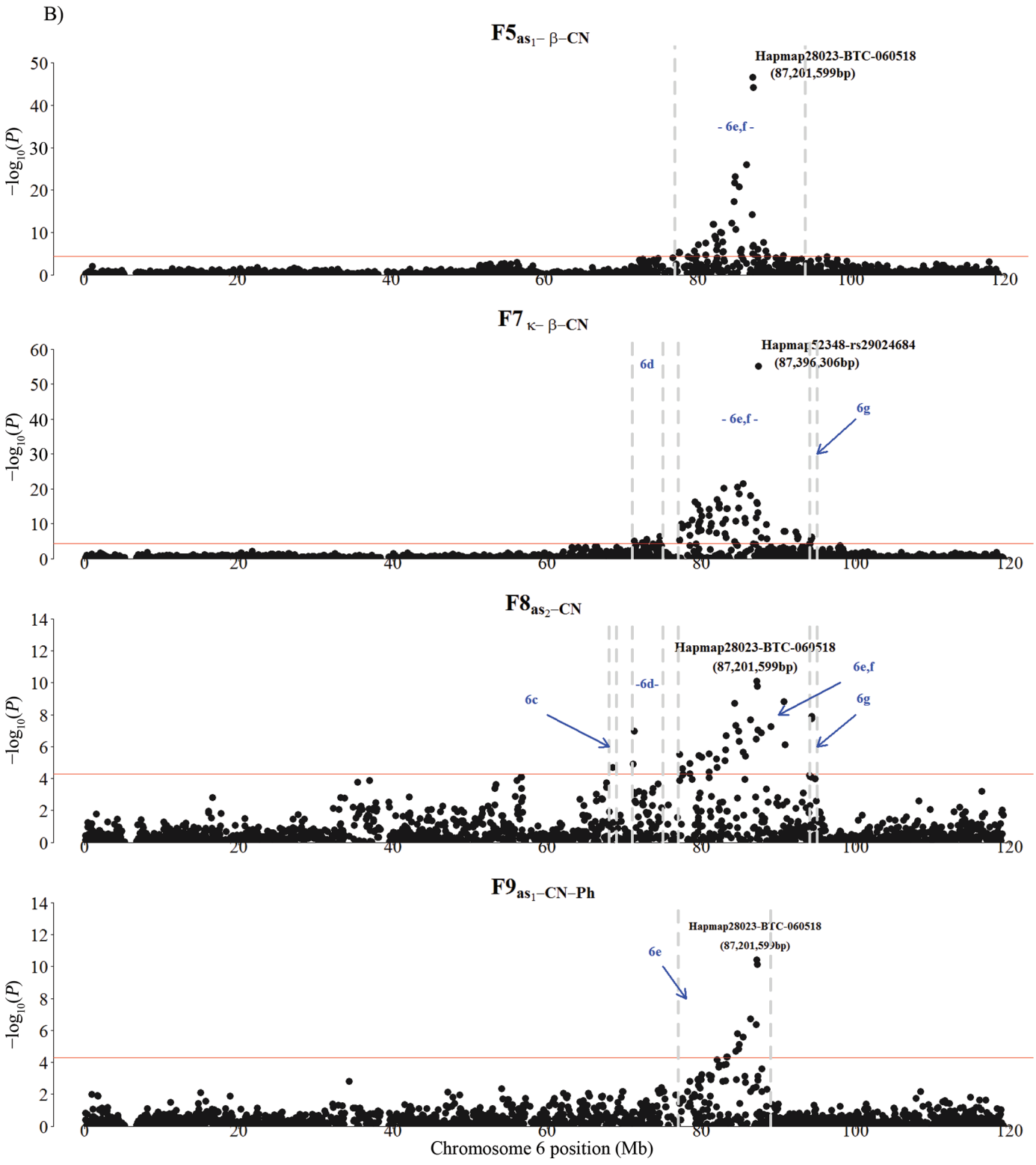

Figure 1 (Continued). Manhattan plots of *P*-values from the genome-wide association study on BTA6. (A) $F1_{\%CY}$ = factor underlying the percentage of individual cheese yield; $F2_{CFt}$ = factor underlying the milk curd firmness; $F4_{Cheese\ N}$ = factor underlying the protein in the cheese; $F6_{Udder\ health}$ = factor underlying the udder health condition of a cow. (B) $F5_{\alpha S1\text{-}\beta\text{-}CN}$ = factor underlying the α_{S1}- and β-CN; $F7_{\kappa\text{-}\beta\text{-}CN}$ = factor underlying the κ- and β-CN; $F8_{aS2\text{-}CN}$ = factor underlying the α_{S2}-CN; $F9_{\alpha S1\text{-}CN\text{-}Ph}$ = factor underlying the phosphorylated α_{S1}-CN. The red horizontal lines indicate a $-\log_{10}$(*P*-value) of 4.30 (corresponding to *P*-value = 5×10^{-5}). The highest significant marker on BTA6 per trait is also presented. Color version available online.

Gene-Set Enrichment and Pathway-Based Analysis

Out of 37,568 tested SNP in GWAS, 17,006 were located in annotated genes or in the 15-kb window up- or downstream the genes. In total, 13,269 background genes were annotated in the *Bos taurus* UMD3.1 assembly. On average, 1,550 SNP per F had a nominal *P*-value <0.05. From those SNP, 529 were assigned to genes and 454 genes were mapped (average values per factor; Supplemental Table S2; https://doi.org/10.3168/jds.2017-13219).

After false discovery rate control (5%), 33 GO terms and 6 KEGG categories were associated with 4 of the 10 tested F, namely $F1_{\%CY}$, $F4_{\text{Cheese N}}$, $F8_{\alpha S2\text{-CN}}$, and $F10_{\alpha\text{-LA}}$, with the vast majority being associated with $F8_{\alpha S2\text{-CN}}$. Results of the gene-set enrichment and pathway-based analysis are outlined in Figure 3 and Supplemental Table S3 (https://doi.org/10.3168/jds.2017-13219). A total of 117 genes spanning all BTA but 21 and 29 were included into the significantly enriched GO and KEGG categories (Supplemental Table S4; https://doi.org/10.3168/jds.2017-13219). Factor $4_{\text{Cheese N}}$ was associated with the arrhythmogenic right ventricular cardiomyopathy (**ARVC**; KEGG: bta05412). The tight junction pathway (KEGG: bta04530) was enriched for both $F1_{\%CY}$ and $F10_{\alpha\text{-LA}}$. Three KEGG categories were enriched for $F8_{\alpha S2\text{-CN}}$, namely the GnRH signaling pathway (KEGG: bta04912), the vascular smooth muscle (KEGG: bta04270), and the long-term potentiation (KEGG: bta04720). Moreover, 33 GO terms were enriched for $F8_{\alpha S2\text{-CN}}$, 12 GO_BP related to cell communication and ion transport, 11 GO_CC belonging to neuron part or function, and 10 GO_MF related to ion transport.

DISCUSSION

Extraction of Factors

Using FA, we condensed 26 cheesemaking phenotypes into 10 F. Although the average KMO value was not high, it was close to the value reported in a recent and similar study on milk composition, MCP, and udder health phenotypes in dairy sheep (Manca et al., 2016). The 10 F in our study represented basic concepts of the cheesemaking process, retaining 74% of the original variability. In a similar data set, but with 11 MCP and udder health phenotypes, the total variance explained by the 4 F was 70% (Macciotta et al., 2012). The same factor scores have been previously used for estimating (co)variance components using standard quantitative genetic model (Dadousis et al., 2017c). Results were coherent to the given name of the factors. Indeed, the first 4 F, sorted by variance explained, were able to capture the underlying structure of the cheese yield (%), the curd firming process, the milk yield, and the presence of N into the cheese. Moreover, 4 F were associated with the basic milk caseins (α_{S1}-β-CN, κ-β-CN, α_{S2}-CN, and α_{S1}-CN-Ph) and 1 factor was related to a whey protein (α-LA). A factor describing the udder health status of a cow, mainly loaded on lactose, other N compounds and SCS, was also obtained.

Table 4. Top SNP[1] detected in the region 6e on *Bos taurus* autosome 6 (BTA6)[2]

Factor	Top SNP	Top SNP location (bp)	*P*-value	Top SNP effect	VG_{SNP} (%)
$F1_{\%CY}$	Hapmap53172-rs29012675	82,706,745	1.39×10^{-6}	0.45	12.0
$F5_{\alpha S1\text{-}\beta\text{-CN}}$	Hapmap28023-BTC-060518	87,201,599	2.84×10^{-47}	−0.90	52.8
$F8_{\alpha S2\text{-CN}}$	Hapmap28023-BTC-060518	87,201,599	8.34×10^{-11}	−0.35	16.0
$F9_{\alpha S1\text{-CN-Ph}}$	Hapmap28023-BTC-060518	87,201,599	3.86×10^{-11}	−0.32	24.7
$F6_{\text{Udder health}}$	Hapmap52348-rs29024684	87,396,306	5.84×10^{-6}	0.18	17.8
$F7_{\kappa\text{-}\beta\text{-CN}}$	Hapmap52348-rs29024684	87,396,306	9.81×10^{-56}	−1.01	74.2
$F4_{\text{Cheese N}}$	ARS-BFGL-NGS-24522	87,878,364	4.40×10^{-6}	0.31	5.3
$F2_{\text{CFt}}$	BTA-122637-no-rs	88,442,145	6.91×10^{-6}	−0.40	13.2

[1]Top SNP = the highest significant SNP detected in the region 6e on BTA6 for each trait.

[2]Top SNP location (bp) = position of the highest significant SNP on the chromosome in base pairs on UMD3.1; *P*-value = *P*-value of the highest significant SNP adjusted for genomic control; Top SNP effect = effect of the highest significant SNP; factor scores are standardized with zero mean and SD of 1; VG_{SNP} (%) = proportion of the additive genetic variance explained by the highest significant SNP (SNP variance was estimated as $2pqa^2$, where p is the frequency of 1 allele, $q = 1 - p$ is the frequency of the second allele, and a denotes the additive genetic effect). $F1_{\%CY}$ = factor related to the percentage of individual cheese yield; $F2_{CFt}$ = factor related to the curd firmness; $F4_{\text{Cheese N}}$ = factor related to the milk nitrogen that is present into the cheese curd; $F5_{\alpha S1\text{-}\beta\text{-CN}}$ = factor related to the α_{S1}- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; $F6_{\text{Udder health}}$ = factor related to the udder health of a cow; $F7_{\kappa\text{-}\beta\text{-CN}}$ = factor related to the κ- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; $F8_{\alpha S2\text{-CN}}$ = factor related to the milk a_{S2}-CN, expressed as relative content to the total milk nitrogen; $F9_{\alpha S1\text{-CN-Ph}}$ = factor related to the milk α_{S1}–phosphorylated CN expressed as content to the total milk nitrogen.

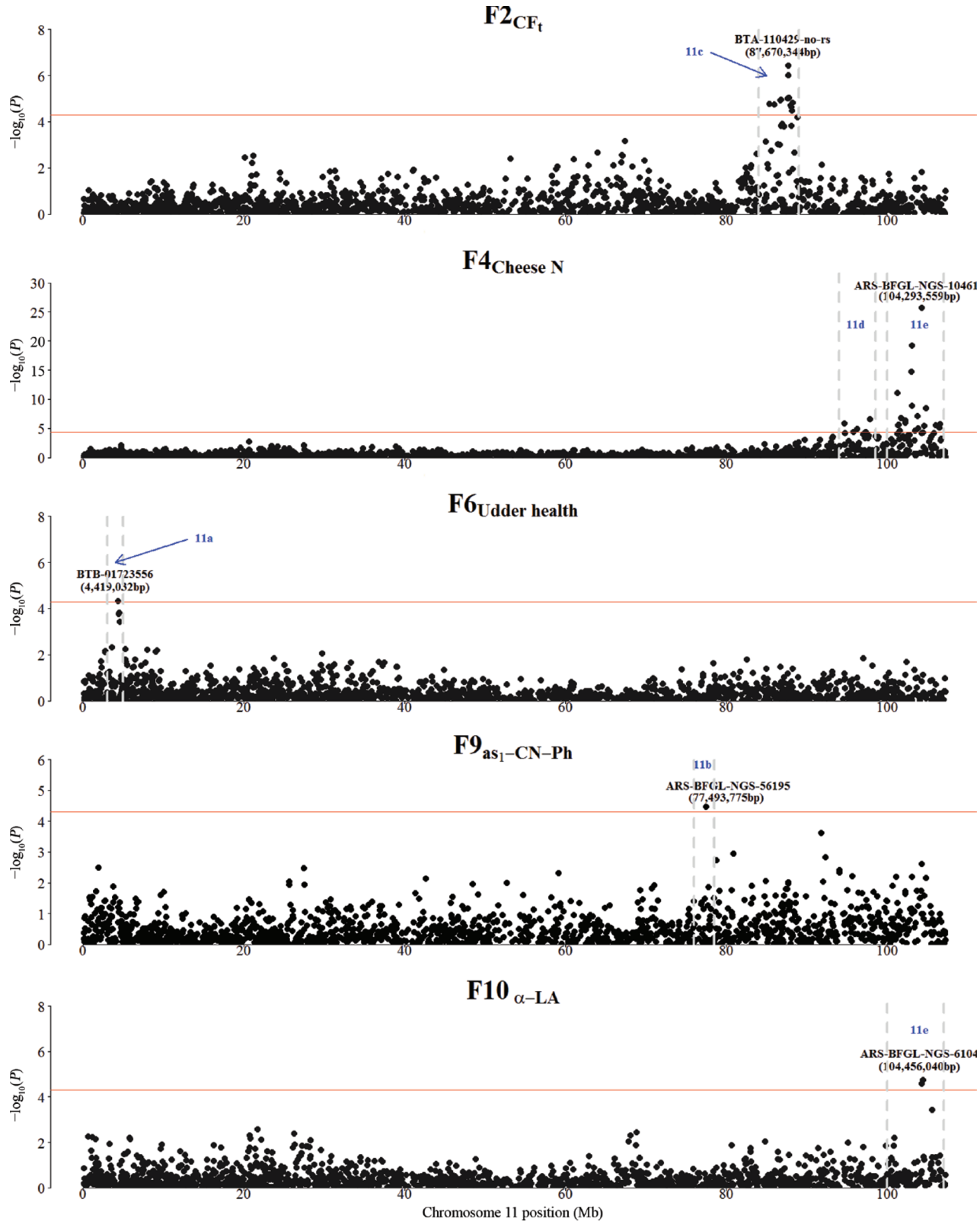

Figure 2. Manhattan plots of *P*-values from the genome-wide association study on BTA11. $F2_{CF_t}$ = factor underlying the milk curd firmness; $F4_{Cheese\ N}$ = factor underlying the protein in the cheese; $F6_{Udder\ health}$ = factor underlying the udder health condition of a cow; $F9_{\alpha S1\text{-}CN\text{-}Ph}$ = factor underlying the phosphorylated α_{S1}-CN; $F10_{\alpha\text{-}LA}$ = factor underlying the α-LA. Red horizontal lines indicate a $-\log_{10}$(*P*-value) of 4.30 (corresponding to *P*-value = 5×10^{-5}). The highest significant marker on BTA11 per trait is also presented. Color version available online.

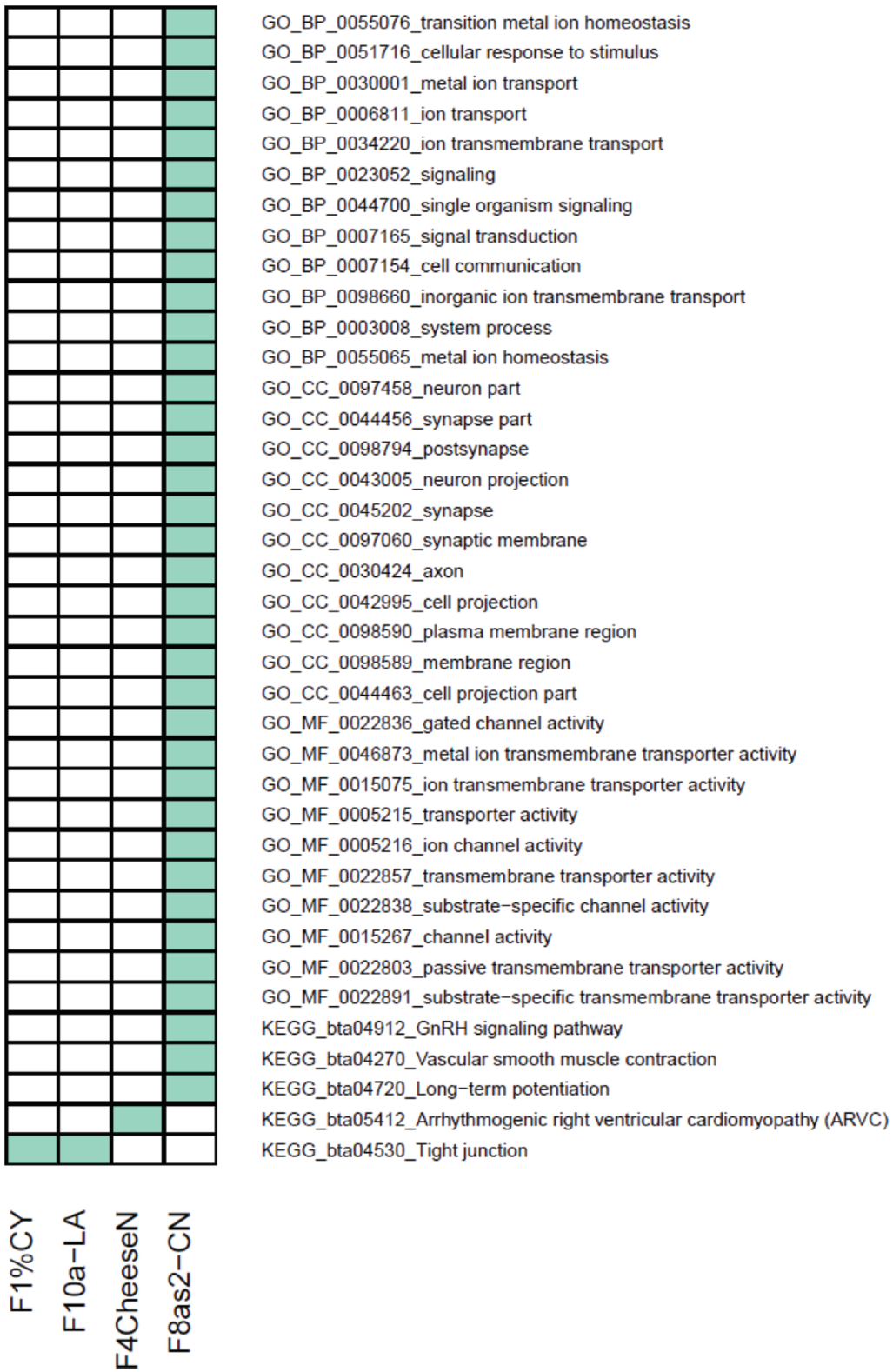

Figure 3. Gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways significantly enriched. Genes containing significant SNP ($P < 0.05$) or mapping at 15 kbp up- and downstream the significant SNP ($P < 0.05$) were used to perform the gene-set enrichment and pathway-based analyses for all the factors. $F1_{\%CY}$ = factor underlying the percentage of individual cheese yield; $F4_{Cheese\ N}$ = factor underlying the protein in the cheese; $F8_{\alpha S2\text{-}CN}$ = factor underlying the α_{S2}-CN; $F10_{\alpha\text{-}LA}$ = factor underlying the α-LA. GO_BP = GO biological process; GO_CC = GO cellular component; GO_MF = GO molecular function. Color version available online.

GWAS

Previous GWAS studies detected several chromosomal regions related to bovine milk protein components (Schopen et al., 2009; Bijl et al., 2014), MCP and CF_t characteristics (Gregersen et al., 2015; Dadousis et al., 2016), and individual CY phenotypes (Dadousis et al., 2017c). Major effects are known on BTA6 for milk technological traits and protein variants, in a region spanning between ~82 to 88 Mbp (Schopen et al., 2011; Gregersen et al., 2015; Dadousis et al., 2016), including the casein cluster, and 2 potential QTL have been suggested at ~82.6 and ~88.4 Mbp and at the tail of BTA11 (at ~87 and ~104 Mbp) (Schopen et al., 2011; Dadousis et al., 2016). The location of the casein genes on BTA6 is widely known (Caroli et al., 2009), whereas the signals on BTA11 were mainly attributed to the β-lactoglobulin gene (*BLG*). Moreover, the effect of milk protein variants on milk coagulation and cheese yield is known (Bonfatti et al., 2010; Bittante et al., 2012).

BTA6. The majority of the GWAS signals were detected in the region 6e. The strongest signal in our study was found within this area, at 87,396,306 bp (Hapmap52348-rs29024684), and it was associated with $F7_{\kappa\text{-}\beta\text{-CN}}$. Indeed, the SNP is located ~18 kbp upstream to the κ-CN gene (*CSN3*). This marker had a strong effect, explaining ~74% of the total additive genetic variance for $F7_{\kappa\text{-}\beta\text{-CN}}$ (Table 4). In previous studies, this marker was strongly linked with a trait describing the potential asymptotical curd firmness (Dadousis et al., 2016) and with the REC_{FAT} (Dadousis et al., 2017a). In our study, the same marker was also associated with $F6_{\text{Udder health}}$, albeit at a much weaker strength compared with $F7_{\kappa\text{-}\beta\text{-CN}}$. In the region between 83.4 and 88.9 Mbp, QTL associated with clinical mastitis have been detected in Nordic Holstein (Sahana et al., 2013). It is worth mentioning that SCS was a minor loading on $F6_{\text{Udder health}}$; moreover, the region at ~88.8 Mbp has been associated with SCS in US Holstein cows (Cole et al., 2011). Close to Hapmap52348-rs29024684, at ~87.2 Mbp, the Hapmap28023-BTC-060518 was associated with $F5_{\alpha\text{S1-}\beta\text{-CN}}$, $F7_{\kappa\text{-}\beta\text{-CN}}$, $F8_{\alpha\text{S2-CN}}$, and $F9_{\alpha\text{S1-CN-Ph}}$ (Table 4, Supplemental Table S1; https://doi.org/10.3168/jds.2017-13219). The highest effect of this SNP was found for $F5_{\alpha\text{S1-}\beta\text{-CN}}$, explaining ~53% of its additive genetic variability. This marker is located within the histatherin gene (*HSTN*) and there is evidence that this gene underlies QTL related to CF_t phenotypes and REC_{FAT} (Dadousis et al., 2016; Dadousis et al., 2017a). The marker BTA-122637-no-rs located at ~88.4 Mbp was associated with $F2_{CFt}$ and $F7_{\kappa\text{-}\beta\text{-CN}}$ ($P = 6.91 \times 10^{-6}$ and 2.46×10^{-10}, respectively; Supplemental Table S1; https://doi.org/10.3168/jds.2017-13219). Notably, the same marker has been previously associated with RCT_{eq}, whereas hits for CF_{max}, $k_{CF,}$ and protein percentage have also been found in the broader region ~87.2 to 88.8 Mbp (Dadousis et al., 2016). The last 2 phenotypes were the major loadings of $F2_{CFt.}$ This marker is located within the solute carrier family 4 member 4 (*SLC4A4*) gene (~88.2–88.5 Mbp), which is involved in the regulation of intracellular pH and secretion and absorption of bicarbonate. Very close to this region it is located the GC Vitamin D Binding Protein (*GC*) gene (~88.69–88.74 Mbp). This gene encodes for a protein, belonging to the albumin family, involved in the metabolism of the vitamin D, lipids, and lipoproteins. In a recent fine mapping study on BTA6, using sequencing data in Norwegian Red cattle, *GC* was suggested as a candidate gene related to milk production and clinical mastitis (Olsen et al., 2016). Factor $1_{\%CY}$ was also associated in the region 6e, with a peak at ~82.7 Mbp (Hapmap53172-rs29012675). The same marker has been previously associated with $\%CY_{SOLIDS}$, $\%CY_{CURD}$, REC_{FAT}, REC_{ENERGY}, and REC_{SOLIDS} (Dadousis et al., 2017a). Not surprisingly, $F1_{\%CY}$ was primarily loaded to $\%CY_{SOLIDS}$ as well as to REC_{ENERGY}.

In the region 6b, a relatively weak signal for $F2_{CFt}$ was detected (Hapmap23226-BTA-159656, ~46.6 Mbp); the same region was previously associated with t_{max} (Dadousis et al., 2016). The t_{max} was strongly related to $F2_{CFt}$, but it was not the heaviest loading on this F. The region 6h, at ~114.2 Mbp, was exclusively associated with $F1_{\%CY}$, albeit with a *P*-value on the significance threshold. A similar weak signal has been previously reported and related to milk protein percentage (Dadousis et al., 2016). Interestingly, this F was loaded to milk protein (%), although with the weaker relation (0.59) among the rest of the phenotypes describing the F.

BTA11. Overall, 5 of the 10 F were linked to 5 regions on BTA11. The strongest association was found between $F4_{\text{Cheese N}}$ and ARS-BFGL-NGS-104610 (104,293,559 bp). The same marker has been strongly related to $REC_{PROTEIN}$ (Dadousis et al., 2017a). Notably, $F4_{\text{Cheese N}}$ was loaded on $REC_{PROTEIN}$. Two SNP in the region 11e (~104.3–104.4 Mbp) were also associated with $F10_{\alpha\text{-LA}}$. However, there is no known QTL on BTA11 related to α-LA (Schopen et al., 2011). The region 11d associated with $F4_{\text{Cheese N}}$ is in close proximity to the region 96.2 to 98.5 Mbp, where signals have been previously detected for $REC_{PROTEIN}$ (Dadousis et al., 2017a). In both cases the same peak was observed at ~97 Mbp. Factor 2_{CFt} was linked to the region 11c, with a peak at ~87.7 Mbp; an association between the identified SNP and RCT_{eq} has been previously reported (Dadousis et al., 2016). A weak association at ~4.4 Mbp was found for $F6_{\text{Udder health}}$; although far from

this region, signals for SCS have been reported in US Holstein cows at the beginning of BTA11 at ~0.28 and ~2.8 Mbp (Cole et al., 2011).

Signals on Chromosomes Other than BTA6 and BTA11. Our study detected weaker associations in 8 additional chromosomes (Table 3). With the exception of BTA20 and BTA27, the rest of the chromosomes were linked to only 1 F. The SNP associated with $F1_{\%CY}$ on BTA19 and BTA27 have been significantly related to $\%CY_{SOLIDS}$, whereas the marker on BTA2 was ~6 Mbp downstream to the one associated with $\%CY_{SOLIDS}$ (Dadousis et al., 2017a). On BTA25, $F6_{Udder\ health}$ was associated with a SNP at ~5.4 Mbp, in close proximity to the ~5.3 Mbp region that showed significant association with SCS (Cole et al., 2011). The signal on BTA1 linked to $F7_{\kappa\text{-}\beta\text{-}CN}$ was not confirmed in the literature, as neither of these casein fractions have been associated with this region on BTA1. Moreover, individual GWAS for κ- and β-CN did not result in significant associations on BTA1 (results not shown). Further, based on the fact that only 1 SNP passed the significance threshold whereas the rest of the markers in the same region showed much lower *P*-values, one could hypothesize a spurious association. Similarly, although we found significant associations on BTA9 for $F5_{\alpha S1\text{-}\beta\text{-}CN}$, no associations have been previously reported for α_{S1}- or β-CN on this chromosome. However, GWAS analysis using the individual α_{S1}-CN content detected the same marker with a similar *P*-value (results not shown). On BTA10, 2 genomic regions are known to be related to α_{S2}-CN at ~51.4 and ~91.8 Mbp (Schopen et al., 2011). In our analysis, $F8_{\alpha S2\text{-}CN}$ was associated with a region at ~10.7 Mbp. No QTL is known at this position affecting the α_{S2}-CN. Moreover, no association on BTA20 and BTA27b have been previously found for $F10_{\alpha\text{-}LA}$.

Gene-Set Enrichment and Pathway Analysis

Four F ($F1_{\%CY}$, $F4_{Cheese\ N}$, $F8_{\alpha S2\text{-}CN}$, and $F10_{\alpha\text{-}LA}$) out of 10 tested were associated with biological pathways and ontologies in the KEGG and GO databases (Figure 3, Supplemental Table S4; https://doi.org/10.3168/jds.2017-13219). The majority of the significantly enriched terms were associated with $F8_{\alpha S2\text{-}CN}$, in which only a_{S2}-CN was loaded. This casein constitutes up to 10% of the bovine casein fraction (Ibeagha-Awemu et al., 2007). To confirm our results on $F8_{\alpha S2\text{-}CN}$, we re-ran the GWAS and gene-set enrichment analysis on the measured α_{S2}-CN content as well as on the rest of the caseins. Gene-set enrichment results for α_{S2}-CN were generally overlapping and, in particular, GO terms related to ion transport and neuron part or function were shared (results not shown). Moreover, no GO or KEGG category was enriched for the measured values of the other caseins, consistent with the F results.

Overall, some of the identified GO and KEGG categories have been previously detected in gene-set enrichment studies using the individual CF_t, CY, and REC phenotypes (Dadousis et al., 2016), milk yield traits (Iso-Touru et al., 2016), or gene expression studies of the mammary gland in mice (Ramanathan et al., 2008; Wei et al., 2013) and humans (Maningat et al., 2009).

Pathways and Ontologies Related to Milk Yield and Mastitis. It has been established that caseins, apart from their importance in milk and in the cheese process (Walstra et al., 2006), also have bioactive and antimicrobial properties (Zucht et al., 1995; Silva and Malcata, 2005; López-Expósito et al., 2006). Moreover, α_{S2}-CN was found particularly responsive to mastitis infection (Smolenski et al., 2014), suggesting that it might be a biologically relevant host-defense protein. Also, an antimicrobial role of α-LA has been suggested (Pellegrini et al., 1999). For milk secretion rate, tight junctions play an important role, with a decrease in their permeability to be associated with an increased milk secretion rate (Nguyen and Neville, 1998). Mastitis, milk stasis, and high doses of oxytocin are known parameters that influence the permeability of the tight junctions.

In our gene-set enrichment analysis, a group of GO ontologies enriched for $F8_{\alpha S2\text{-}CN}$ was related to ion transport activity. Some of these terms have been previously connected with milk production in mice. More precisely, the GO_BP:0006811 (ion transport), GO_MF:0005216 (ion channel activity), GO_MF: 0022838 (substrate-specific channel activity), and GO_MF: 0015267 (channel activity) were upregulated in mice with increased milk yield (Wei et al., 2013). Moreover, it is known that in the cheese process the caseins react with calcium ions. Calcium is a major component of the casein micelles. Indeed, the α_{S2}-CN is known to be rather sensitive to Ca^{2+} (Walstra et al., 2006). Further, it is well established that in milk the most important ions for electrical conductivity (**EC**) are the concentrations of Na^+, K^+, and Cl^-. Milk EC can be considered as an indicator of mastitis (Norberg, 2005; Viguier et al., 2009). While Na^+ and Cl^- are moving into the milk, tight junctions of the mammary epithelium control the movement of lactose and K^+ to the extracellular fluid. Destruction of tight junctions and of the ion-pumping system, after IMI, causes an increase in the concentration of Na^+ and Cl^- in the milk, resulting in an increase of the milk EC (Norberg, 2005). In our results, the tight junction pathway (KEGG_bta04530) category was associated with $F1_{\%CY}$ and $F10_{\alpha\text{-}LA}$. It has been reported that milk with high SCC has lower casein con-

tent (Haenlein et al., 1973). Pathways related to mammary gland and mastitis, including the tight junction, have been previously associated with the REC_{ENERGY} (Dadousis et al., 2016), a trait that was strongly related to $F1_{\%CY}$ in the FA.

Enriched Pathways and Ontologies Related to Reproduction. Seven GO_CC categories relative to neuron functions were enriched for $F8_{\alpha S2\text{-}CN}$. A possible explanation can be the fact that during the pregnancy and lactation periods, a variety of factors and signals (including the prolactin neuroendocrine signal) are involved to assist neuronal responses to the lactating state (Akers, 2002; Grattan, 2002). Interestingly, in a recent gene enrichment and pathway study the individual CF_t phenotypes, the categories of neuron part (GO: 0097458), synapse part (GO: 0044456), neuron projection (GO: 0043005), and the synapse (GO: 0045202) were enriched for RCT_{eq} (Dadousis et al., 2016). Moreover, associations of the synapse part (GO:0044456) and the postsynapse (GO:0098794) with the k_{CF} were detected in Dadousis et al. (2016); the cellular response to stimulus category (GO_BP:0051716) was also significantly enriched for $F8_{\alpha S2\text{-}CN}$. The closely related gene ontology of response to stimulus (GO:0050896) has been previously associated with the milk fat globule transcriptome during lactation in humans (Maningat et al., 2009). Moreover, in dairy cattle, this term was significantly enriched for milk yield, fat and protein yield, and fertility (Iso-Touru et al., 2016). Additionally, the GnRH signaling pathway (KEGG_bta04912) was enriched for $F8_{\alpha S2\text{-}CN}$. The GnRH is synthesized and released in the hypothalamus from the GnRH neurons and strongly related to reproduction in mammals (Schneider et al., 2006). Interestingly, GO categories related to female gonad development and ovulation cycle were previously linked to RCT_{eq} (Dadousis et al., 2016). Moreover, GO terms of reproduction (GO:0000003) and reproductive process (GO:002214) have been associated with milk yield, fat and protein yield, and fertility index in the Nordic Red cattle (Iso-Touru et al., 2016). Indeed, a close relationship is known between the duration of estrus and multiple ovulation rate and milk production in dairy cattle. More precisely, high production is associated with shorter estrus duration and double ovulation rate (Wiltbank et al., 2006).

Other Enriched Pathways and Ontologies. In our study, ARVC was enriched for $F4_{Cheese\ N}$. The ARVC is an inherited heart disease (Elmaghawry et al., 2013) and, in a recent gene-set enrichment analysis, was linked to bovine leucosis (Abdalla et al., 2016). The same KEGG category has been recently associated with $\%CY_{SOLIDS}$ and REC_{SOLIDS} (Dadousis et al., 2016). Notably, in a transcriptome study of the swine mammary gland, ARVC was associated with the mammary gland functionality of pregnant sows (Zhao et al., 2013).

Moreover, for $F8_{\alpha S2\text{-}CN}$, GO terms related to cell communication and signaling (e.g., GO_BP:0023052, GO_BP:0007154) were enriched in our study. These categories have been shown to have a role in human milk fat globule transcriptome, which is characterized by high expression of milk protein genes (Maningat et al., 2009).

Our analysis has shown that FA can be considered as an appropriate and useful tool in genomic studies. From a practical point of view in breeding programs, F could replace the measured phenotypes in a selection index.

CONCLUSIONS

To our knowledge, this is the first analysis using latent variables in GWAS and gene-set enrichment pathway analysis in dairy cattle. Genomic regions identified were coherent with the expected signals based on the factor loadings and their interpretations. Results of gene-set enrichment analysis were also in line with previous findings based on the individual measured phenotypes, and revealed that the associated genes were mainly involved in pathways related to reproduction and mammary gland functionality. The considerably large number of enriched GO and KEGG terms for $F8_{\alpha S2\text{-}CN}$ suggests that, perhaps, α_{S2}-CN might have a relevant biological role in the regulation of processes affecting milk quality and composition. We concluded that FA can be successfully implemented in genomic studies in dairy cattle, allowing a reduction on data dimensionality without a substantial loss of information.

ACKNOWLEDGMENTS

The authors thank the Trento Province (Italy), the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy), and the Superbrown Consortium of Bolzano and Trento (Italy) for financial and technical support. C. Dadousis benefitted from financial support of the CARIPARO (Cassa di Risparmio di Padova e Rovigo) Foundation (Padua, Italy). The authors also wish to acknowledge Claudio Cipolat-Gotet (Department of Agronomy, Food, Natural Resources, Animals and Environment–DAFNAE, University of Padova) and Valentina Bonfatti (Department of Comparative Biomedicine and Food Science, University of Padova) for their cooperation in assessing the cheese-making traits and milk protein fractions respectively. The authors also thank F. Peñagaricano (Department of Animal Sciences, University of Florida) for his help in setting up the statistical analysis.

REFERENCES

Abdalla, E. A., F. Peñagaricano, T. M. Byrem, K. A. Weigel, and G. J. M. Rosa. 2016. Genome-wide association mapping and pathway analysis of leukosis incidence in a US Holstein cattle population. Anim. Genet. 47:395–407. https://doi.org/10.1111/age.12438.

Akers, R. M. 2002. Lactation and the Mammary Gland. Wiley, Hoboken, NJ.

Ali, A. K. A., G. E. Shook, F. R. Gabler, and J. Peters. 1980. An optimum transformation for somatic cell concentration in milk. J. Dairy Sci. 63:487–490. https://doi.org/10.3168/jds.S0022-0302(80)82959-6.

Amin, N., C. M. van Duijn, and Y. S. Aulchenko. 2007. A genomic background based method for association analysis in related individuals. PLoS One 2:e1274. https://doi.org/10.1371/journal.pone.0001274.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25:25–29. https://doi.org/10.1038/75556.

Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn. 2007. GenABEL: An R library for genome-wide association analysis. Bioinformatics 23:1294–1296. https://doi.org/10.1093/bioinformatics/btm108.

Bijl, E., H. van Valenberg, T. Huppertz, A. van Hooijdonk, and H. Bovenhuis. 2014. Phosphorylation of α_{S1}-casein is regulated by different genes. J. Dairy Sci. 97:7240–7246. https://doi.org/10.3168/jds.2014-8061.

Bittante, G., C. Cipolat-Gotet, and A. Cecchinato. 2013a. Genetic parameters of different measures of cheese yield and milk nutrient recovery from an individual model cheese-manufacturing process. J. Dairy Sci. 96:7966–7979. https://doi.org/10.3168/jds.2012-6517.

Bittante, G., B. Contiero, and A. Cecchinato. 2013b. Prolonged observation and modelling of milk coagulation, curd firming, and syneresis. Int. Dairy J. 29:115–123. https://doi.org/10.1016/j.idairyj.2012.10.007.

Bittante, G., M. Penasa, and A. Cecchinato. 2012. Invited review: Genetics and modeling of milk coagulation properties. J. Dairy Sci. 95:6843–6870. https://doi.org/10.3168/jds.2012-5507.

Bollen, K. A. 2014. Structural Equations with Latent Variables. Wiley, Hoboken, NJ.

Bonfatti, V., A. Cecchinato, L. Gallo, A. Blasco, and P. Carnier. 2011. Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. J. Dairy Sci. 94:5183–5193. https://doi.org/10.3168/jds.2011-4297.

Bonfatti, V., G. Di Martino, A. Cecchinato, D. Vicario, 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. https://doi.org/10.3168/jds.2009-2778.

Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier. 2008. Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. J. Chromatogr. A 1195:101–106. https://doi.org/10.1016/j.chroma.2008.04.075.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen. 2016. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC Genet. 17:114. https://doi.org/10.1186/s12863-016-0421-2.

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. https://doi.org/10.3168/jds.2009-2461.

Cecchinato, A., and G. Bittante. 2016. Genetic and environmental relationships of different measures of individual cheese yield and curd nutrients recovery with coagulation properties of bovine milk. J. Dairy Sci. 99:1975–1989. https://doi.org/10.3168/jds.2015-9629.

Cipolat-Gotet, C., A. Cecchinato, M. De Marchi, and G. Bittante. 2013. Factors affecting variation of different measures of cheese yield and milk nutrient recovery from an individual model cheese-manufacturing process. J. Dairy Sci. 96:7952–7965. https://doi.org/10.3168/jds.2012-6516.

Cipolat-Gotet, C., A. Cecchinato, M. De Marchi, M. Penasa, and G. Bittante. 2012. Comparison between mechanical and near-infrared methods for assessing coagulation properties of bovine milk. J. Dairy Sci. 95:6806–6819. https://doi.org/10.3168/jds.2012-5551.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, B. A. Crooker, C. P. Van Tassell, J. Yang, S. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics 12:408. https://doi.org/10.1186/1471-2164-12-408.

Conte, G., A. Serra, P. Cremonesi, S. Chessa, B. Castiglioni, A. Cappucci, E. Bulleri, and M. Mele. 2016. Investigating mutual relationship among milk fatty acids by multivariate factor analysis in dairy cows. Livest. Sci. 188:124–132. https://doi.org/10.1016/j.livsci.2016.04.018.

Dadousis, C., S. Biffani, C. Cipolat-Gotet, E. L. Nicolazzi, G. J. M. Rosa, D. Gianola, A. Rossoni, E. Santus, G. Bittante, and A. Cecchinato. 2017a. Genome-wide association study for cheese yield and curd nutrient recovery in dairy cows. J. Dairy Sci. 100:1259–1271. https://doi.org/10.3168/jds.2016-11586.

Dadousis, C., S. Biffani, C. Cipolat-Gotet, E. L. Nicolazzi, A. Rossoni, E. Santus, G. Bittante, and A. Cecchinato. 2016. Genome-wide association of coagulation properties, curd firmness modeling, protein percentage, and acidity in milk from Brown Swiss cows. J. Dairy Sci. 99:3654–3666. https://doi.org/10.3168/jds.2015-10078.

Dadousis, C., C. Cipolat-Gotet, G. Bittante, and A. Cecchinato. 2017c. Inferring genetic parameters on latent variables underlying milk yield and quality, protein composition, curd firmness and cheese-making traits in dairy cattle. Animal https://doi.org/10.1017/S1751731117001616.

Dadousis, C., S. Pegolo, G. J. M. Rosa, D. Gianola, G. Bittante, and A. Cecchinato. 2017b. Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield, and curd nutrient recovery in dairy cattle. J. Dairy Sci. 100:1223–1231. https://doi.org/10.3168/jds.2016-11587.

Durinck, S., Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, and W. Huber. 2005. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. Bioinformatics 21:3439–3440. https://doi.org/10.1093/bioinformatics/bti525.

Durinck, S., P. T. Spellman, E. Birney, and W. Huber. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4:1184–1191. https://doi.org/10.1038/nprot.2009.97.

Dziuban, C. D., and E. C. Shirkey. 1974. When is a correlation matrix appropriate for factor analysis? Some decision rules. Psychol. Bull. 81:358–361. /https://doi.org/10.1037/h0036316.

Elmaghawry, M., M. Alhashemi, A. Zorzi, and M. H. Yacoub. 2013. A global perspective of arrhythmogenic right ventricular cardiomyopathy. Glob. Cardiol. Sci. Pract. 2012:81–92. https://doi.org/10.5339/gcsp.2012.26.

Fanous, A. H., B. Zhou, S. H. Aggen, S. E. Bergen, R. L. Amdur, J. Duan, A. R. Sanders, J. Shi, B. J. Mowry, A. Olincy, F. Amin, C. R. Cloninger, J. M. Silverman, N. G. Buccola, W. F. Byerley, D. W. Black, R. Freedman, F. Dudbridge, P. A. Holmans, S. Ripke, P. V Gejman, K. S. Kendler, D. F. Levinson, and Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. 2012. Genome-wide association study of clinical dimensions of schizophrenia: polygenic effect on disorganized symptoms. Am. J. Psychiatry 169:1309–1317. https://doi.org/10.1176/appi.ajp.2012.12020218.

Galesloot, T. E., K. van Steen, L. A. L. M. Kiemeney, L. L. Janss, and S. H. Vermeulen. 2014. A Comparison of multivariate genome-wide association methods. PLoS One 9:e95923. https://doi.org/10.1371/journal.pone.0095923.

Gambra, R., F. Peñagaricano, J. Kropp, K. Khateeb, K. A. Weigel, J. Lucey, and H. Khatib. 2013. Genomic architecture of bovine κ-casein and β-lactoglobulin. J. Dairy Sci. 96:5333–5343. https://doi.org/10.3168/jds.2012-6324.

Grattan, D. R. 2002. Behavioural significance of prolactin signalling in the central nervous system during pregnancy and lactation. Reproduction 123:497–506.

Gregersen, V. R., F. Gustavsson, M. Glantz, O. F. Christensen, H. Stålhammar, A. Andrén, H. Lindmark-Månsson, N. A. Poulsen, L. B. Larsen, M. Paulsson, and C. Bendixen. 2015. Bovine chromosomal regions affecting rheological traits in rennet-induced skim milk gels. J. Dairy Sci. 98:1261–1272. https://doi.org/10.3168/jds.2014-8136.

Haenlein, G. F., L. H. Schultz, J. P. Zikakis, and R. M. Weinberg. 1973. Composition of proteins in milk with varying leucocyte contents. J. Dairy Sci. 56:1017–1024. https://doi.org/10.3168/jds.S0022-0302(73)85299-3.

Ibeagha-Awemu, E. M., E.-M. Prinzenberg, O. C. Jann, G. Lühken, A. E. Ibeagha, X. Zhao, and G. Erhardt. 2007. Molecular characterization of bovine CSN1S2*B and extensive distribution of zebu-specific milk protein alleles in European cattle. J. Dairy Sci. 90:3522–3529. https://doi.org/10.3168/jds.2006-679.

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

Jolliffe, I. T. 2002. Principal Component Analysis. Springer, New York, NY.

Kaiser, H. F., and J. Rice. 1974. Little Jiffy, Mark IV. Educ. Psychol. Meas. 34:111–117.

Kern, E. L., J. A. Cobuci, C. N. Costa, and C. M. M. Pimentel. 2014. Factor analysis of linear type traits and their relation with longevity in brazilian holstein cattle. Asian-australas. J. Anim. Sci. 27:784–790. https://doi.org/10.5713/ajas.2013.13817.

Kominakis, A., A. L. Hager-Theodorides, E. Zoidis, A. Saridaki, G. Antonakos, and G. Tsiamis. 2017. Combined GWAS and "guilt by association"-based prioritization analysis identifies functional candidate genes for body size in sheep. Genet. Sel. Evol. 49:41. https://doi.org/10.1186/s12711-017-0316-3.

López-Expósito, I., J. Á. Gómez-Ruiz, L. Amigo, and I. Recio. 2006. Identification of antibacterial peptides from ovine αs2-casein. Int. Dairy J. 16:1072–1080. https://doi.org/10.1016/j.idairyj.2005.10.006.

Macciotta, N. P. P., S. Biffani, U. Bernabucci, N. Lacetera, A. Vitali, P. Ajmone-Marsan, and A. Nardone. 2017. Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. J. Dairy Sci. 100:4683–4697. https://doi.org/10.3168/jds.2016-12249.

Macciotta, N. P. P., A. Cecchinato, M. Mele, and G. Bittante. 2012. Use of multivariate factor analysis to define new indicator variables for milk composition and coagulation properties in Brown Swiss cows. J. Dairy Sci. 95:7346–7354. https://doi.org/10.3168/jds.2012-5546.

Manca, M. G., J. Serdino, G. Gaspa, P. Urgeghe, I. Ibba, M. Contu, P. Fresi, and N. P. P. Macciotta. 2016. Derivation of multivariate indices of milk composition, coagulation properties, and individual cheese yield in dairy sheep. J. Dairy Sci. 99:4547–4557. https://doi.org/10.3168/jds.2015-10589.

Maningat, P. D., P. Sen, M. Rijnkels, A. L. Sunehag, D. L. Hadsell, M. Bray, and M. W. Haymond. 2009. Gene expression in the human mammary epithelium during lactation: The milk fat globule transcriptome. Physiol. Genomics 37:12–22. https://doi.org/10.1152/physiolgenomics.90341.2008.

Mele, M., N. P. P. Macciotta, A. Cecchinato, G. Conte, S. Schiavon, and G. Bittante. 2016. Multivariate factor analysis of detailed milk fatty acid profile: Effects of dairy system, feeding, herd, parity, and stage of lactation. J. Dairy Sci. 99:9820–9833. https://doi.org/10.3168/jds.2016-11451.

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection indices in holstein cattle of various countries. J. Dairy Sci. 88:1255–1263. https://doi.org/10.3168/jds.S0022-0302(05)72792-2.

NRC. 2001. Nutrient Requirements of Dairy Cattle. 7th rev. ed. Natl. Acad. Press, Washington, DC.

Nguyen, D. A., and M. C. Neville. 1998. Tight junction regulation in the mammary gland. J. Mammary Gland Biol. Neoplasia 3:233–246.

Norberg, E. 2005. Electrical conductivity of milk as a phenotypic and genetic indicator of bovine mastitis: A review. Livest. Prod. Sci. 96:129–139. https://doi.org/10.1016/j.livprodsci.2004.12.014.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27:29–34.

Olsen, H. G., T. M. Knutsen, A. M. Lewandowska-Sabat, H. Grove, T. Nome, M. Svendsen, M. Arnyasi, M. Sodeland, K. K. Sundsaasen, S. R. Dahl, B. Heringstad, H. H. Hansen, I. Olsaker, M. P. Kent, and S. Lien. 2016. Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production. Genet. Sel. Evol. 48:79. https://doi.org/10.1186/s12711-016-0257-2.

Pellegrini, A., U. Thomas, N. Bramaz, P. Hunziker, and R. von Fellenberg. 1999. Isolation and identification of three bactericidal domains in the bovine alpha-lactalbumin molecule. Biochim. Biophys. Acta 1426:439–448.

Pickrell, J. K., J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J.-B. Veyrieras, M. Stephens, Y. Gilad, and J. K. Pritchard. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464:768–772. https://doi.org/10.1038/nature08872.

Ramanathan, P., I. C. Martin, M. Gardiner-Garden, P. C. Thomson, R. M. Taylor, C. J. Ormandy, C. Moran, and P. Williamson. 2008. Transcriptome analysis identifies pathways associated with enhanced maternal performance in QSi5 mice. BMC Genomics 9:197 https://doi.org/10.1186/1471-2164-9-197.

Revelle, W. 2017. psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, IL.

Sahana, G., B. Guldbrandtsen, B. Thomsen, and M. S. Lund. 2013. Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. Anim. Genet. 44:620–626. https://doi.org/10.1111/age.12053.

Schäfer, J., and K. Strimmer. 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat. Appl. Genet. Mol. Biol. 4:32. https://doi.org/10.2202/1544-6115.1175.

Schneider, F., W. Tomek, and C. Gründker. 2006. Gonadotropin-releasing hormone (GnRH) and its natural analogues: A review. Theriogenology 66:691–709. https://doi.org/10.1016/j.theriogenology.2006.03.025.

Schopen, G. C. B., J. M. L. Heck, H. Bovenhuis, M. H. P. W. Visker, H. J. F. van Valenberg, and J. A. M. van Arendonk. 2009. Genetic parameters for major milk proteins in Dutch Holstein-Friesians. J. Dairy Sci. 92:1182–1191. https://doi.org/10.3168/jds.2008-1281.

Schopen, G. C. B., M. H. P. W. Visker, P. D. Koks, E. Mullaart, J. A. M. van Arendonk, and H. Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy cattle. J. Dairy Sci. 94:3148–3158. https://doi.org/10.3168/jds.2010-4030.

Silva, S. V., and F. X. Malcata. 2005. Caseins as source of bioactive peptides. Int. Dairy J. 15:1–15. https://doi.org/10.1016/j.idairyj.2004.04.009.

Smolenski, G. A., M. K. Broadhurst, K. Stelwagen, B. J. Haigh, and T. T. Wheeler. 2014. Host defence related responses in bovine milk during an experimentally induced *Streptococcus uberis* infection. Proteome Sci. 12:19. https://doi.org/10.1186/1477-5956-12-19.

Stocco, G., C. Cipolat-Gotet, T. Bobbo, A. Cecchinato, G. Bittante, E. Pärna, A. Cecchinato, G. Bittante, C. Bendixen, A. J. Buitenhuis, L. B. Larsen, and F. Werkmeister. 2017. Breed of cow and herd productivity affect milk composition and modeling of coagulation, curd firming, and syneresis. J. Dairy Sci. 100:129–145. https://doi.org/10.3168/jds.2016-11662.

Svishcheva, G. R., T. I. Axenovich, N. M. Belonogova, C. M. van Duijn, and Y. S. Aulchenko. 2012. Rapid variance components-based method for whole-genome association analysis. Nat. Genet. 44:1166–1170. https://doi.org/10.1038/ng.2410.

The Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678. https://doi.org/10.1038/nature05911.

Turner, S. D. 2014. qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. bioRxiv https://doi.org/10.1101/005165.

Viguier, C., S. Arora, N. Gilmartin, K. Welbeck, and R. O'Kennedy. 2009. Mastitis detection: Current trends and future perspectives. Trends Biotechnol. 27:486–493. https://doi.org/10.1016/j.tibtech.2009.05.004.

Walstra, P., T. J. Geurts, and J. T. M. Wouters. 2006. Dairy Science and Technology. CRC/Taylor & Francis, Boca Raton, FL.

Wei, J., P. Ramanathan, I. C. Martin, C. Moran, R. M. Taylor, and P. Williamson. 2013. Identification of gene sets and pathways associated with lactation performance in mice. Physiol. Genomics 45:171–181. https://doi.org/10.1152/physiolgenomics.00139.2011.

Wiltbank, M., H. Lopez, R. Sartori, S. Sangsritavong, and A. Gümen. 2006. Changes in reproductive physiology of lactating dairy cows due to elevated steroid metabolism. Theriogenology 65:17–29. https://doi.org/10.1016/j.theriogenology.2005.10.003.

Young, M. D., M. J. Wakefield, G. K. Smyth, and A. Oshlack. 2010. Gene ontology analysis for RNA-seq: Accounting for selection bias. Genome Biol. 11:R14. https://doi.org/10.1186/gb-2010-11-2-r14.

Zhao, W., K. Shahzad, M. Jiang, D. E. Graugnard, S. L. Rodriguez-Zas, J. Luo, J. J. Loor, and W. L. Hurley. 2013. Bioinformatics and gene network analyses of the swine mammary gland transcriptome during late gestation. Bioinform. Biol. Insights 7:193–216. https://doi.org/10.4137/BBI.S12205.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10:R42. https://doi.org/10.1186/gb-2009-10-4-r42.

Zucht, H. D., M. Raida, K. Adermann, H. J. Mägert, and W. G. Forssmann. 1995. Casocidin-I: A casein-alpha s2 derived peptide exhibits antibacterial activity. FEBS Lett. 372:185–188.