UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI AGRONOMIA, ANIMALI, ALIMENTI,
RISORSE NATURALI ED AMBIENTE (DAFNAE)

CORSO DI DOTTORATO DI RICERCA IN
SCIENZE ANIMALI E AGROALIMENTARI
CURRICOLO: SCIENZE ANIMALI
CICLO XXIX

FROM MILK TO CHEESE: GENOMIC BACKGROUND, BIOLOGICAL PATHWAYS AND LATENT PHENOTYPES OF BOVINE CHEESE-RELATED TRAITS

Coordinatore Del Corso: Ch.mo Prof. Stefano Schiavon

Supervisore: Ch.mo Prof. Alessio Cecchinato

Dottorando: Christos Dadousis

FROM MILK TO CHEESE: GENOMIC BACKGROUND, BIOLOGICAL PATHWAYS AND LATENT PHENOTYPES OF BOVINE CHEESE-RELATED TRAITS

by Christos Dadousis

That's the true meaning of science, searching for ever, and your desire for finding things out shall never meet an end

---

Λάμπει μέσα μου κείνο που αγνοώ, μα ωστόσο λάμπει

(In me shines what I ignore. And yet it shines)

(Elytis)

To my family

and

In memory of my bosom friend Christos Toxidis and of my Mentor Prof. Zafeiris Ampas

CONTENTS

RIASSUNTO

Lo scopo di questa tesi di dottorato è stato lo studio del background genomico, biologico e fenotipico di caratteri legati al processo di caseificazione nella specie bovina. L'obiettivo primario è stato quello di determinare il background genomico di caratteri tecnologici del latte bovino legati al processo di caseificazione (CAPITOLI da 1 a 3). Per raggiungere questo obiettivo, l'abilità della bovina di produrre formaggio è stata ripartita in 26 fenotipi: 11 caratteri di attitudine casearia e proprietà di coagulazione, comprendenti le tradizionali proprietà di coagulazione del latte (MCP) e nuovi parametri modellizzati di consistenza della cagliata (CF_t), e 7 fenotipi di resa in formaggio (CY) e recupero dei nutrienti del latte nella cagliata (REC). Tuttavia, l'elevato numero di variabili necessarie per descrivere la produzione di formaggio bovino pone delle restrizioni nella costruzione di indici di selezione, e quindi nel prendere decisioni di selezione. Per superare il problema della elevata dimensionalità, è stata utilizzata un'analisi fattoriale (FA) per studiare la struttura latente dei 26 caratteri coinvolti nel processo di caseificazione (CAPITOLI 4 e 5).

I caratteri MCP includevano le 3 proprietà lattodinamografiche tradizionali basate su singola misurazione dello strumento (RCT: tempo di coagulazione, in min; k_{20}: tempo di rassodamento, in min; a_{30}: consistenza del coagulo (CF) 30 min dopo l'aggiunta del caglio, in mm). I fenotipi CF_t comprendevano un set di 6 parametri modellizzati sulla base di 360 dati di CF misurati per ciascun campione di latte (CF_p: CF potenziale, in mm; k_{CF}: tasso di rassodamento del coagulo, in % × min^{-1}; k_{SR}: tasso di sineresi, in % × min^{-1}; RCT_{eq}: RCT stimato dal modello; CF_{max}: massima CF, in mm; t_{max}: tempo necessario per raggiungere CF_{max}, in min), delle proteine del latte (%) e del pH. I 3 caratteri CY includevano resa a fresco (% CY_{CURD}), resa in solidi (% CY_{SOLIDS}), e acqua ritenuta nella cagliata (% CY_{WATER}), espresse come percentuale del latte trasformato. Le

4 misure di REC (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, e REC_{ENERGY}) sono state calcolate come rapporto espresso in percentuale tra il valore di nutrienti nella cagliata e il corrispettivo nel latte. L'analisi FA ha considerato tutti i 26 caratteri oggetti di studio, comprendenti produzione e qualità del latte (incluse le frazioni proteiche del latte), parametri CFt e caratteri CY-REC.

La metodologia adottata comprendeva analisi di associazione *genome-wide* (GWAS), accompagnata da analisi di arricchimento genetico e di tipo *pathway-based*. Le analisi genomiche hanno considerato un totale, 1.152 bovine di razza Bruna Italiana allevate in 85 allevamenti, genotipizzate attraverso il v.2 Illumina SNP50 Beadchip. Le analisi GWAS sono state condotte mediante analisi di regressione a singolo marcatore, fittate utilizzando il pacchetto GenABEL del software R (GRAMMAR-GC). I database Gene Ontology (GO) e Kyoto Encyclopedia of Genes and Genomes (KEGG) sono stati interrogati per le analisi di arricchimento.

Nell'analisi GWAS (CAPITOLI 1 e 2) sono stati individuati picchi nitidi sull'autosoma 6 di *Bos taurus* (BTA) tra 84-88 Mbp, con il picco più alto rilevato a 87,4 Mbp nella regione ospitante i geni della caseina e più precisamente della κ-CN (*CSN3*). Il marcatore Hapmap52348-rs29024684 (~ 87,4 Mbp), localizzato in prossimità dei geni della caseina su BTA6, ha mostrato una forte associazione con REC_{FAT} ($P = 1.91 \times 10^{-15}$) e CF_P ($P = 1.62 \times 10^{-17}$). Sullo stesso cromosoma, è stata trovata evidenza di loci per i caratteri quantitativi a 82,6 e 88,4 Mbp. Su BTA11, il marcatore ARS-BFGL-NGS-104.610 (~ 104,3 Mbp) è risultato fortemente associato con $REC_{PROTEIN}$ ($P = 6{,}07 \times 10^{-36}$). Oltre a BTA6 e 11, altri SNP situati in altri 15 cromosomi (1, 2, 9, 12, 13, 14, 15, 16, 18, 19, 20, 23, 26, 27 e 28) sono risultati significativamente associati con MCP, CFt e con i caratteri CY-REC.

L'analisi di arricchimento e *pathway-based* (CAPITOLO 3) ha rivelato 21 categorie GO e 17 categorie KEGG significativamente associate (tasso di errore controllato al 5%) con 7 tra i

caratteri fenotipici considerati (RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS} e REC_{ENERGY}) e alcune categorie sono risultate in comune tra i caratteri. Le categorie significativamente arricchite includevano vie di segnalazione del calcio, di secrezione salivare, vie metaboliche, di digestione e assorbimento dei carboidrati, di giunzioni occludenti e del fosfatidilinositolo, così come vie legate allo stato di salute della ghiandola mammaria bovina, per un totale di 150 geni situati in tutti i cromosomi tranne 9, 20 e 27.

Nella FA (CAPITOLI 4 e 5) sono stati ottenuti dieci Fs mutualmente ortogonali utilizzando una rotazione *varimax*. I 10 Fs spiegavano il 74% della variabilità originale. Tali Fs erano biologicamente riconducibili a elementi base del processo di trasformazione “dal latte al formaggio". Più precisamente, i primi 4 Fs, ordinati sulla base della varianza spiegata, sono stati in grado di definire la struttura latente della CY percentuale ($F1_{\%\,CY}$), del processo di CF nel tempo ($F2_{CFt}$), del rendimento di latte e solidi ($F3_{Yield}$) e della presenza di azoto (N) nel formaggio ($F4_{Cheese\ N}$). Inoltre, 4 Fs ($F5_{\alpha s1\text{-}\beta\text{-}CN}$, $F7_{\beta\text{-}\kappa\text{-}CN}$, $F8_{\alpha s2\text{-}CN}$, $F9_{\alpha s1\text{-}CN\text{-}P}$) erano associati alle caseine del latte (as_1-CN, as_2-CN, β-CN, κ-CN, e la forma fosforilata as_1-CN) e 1 fattore alla proteina del siero α-LA ($F10_{\alpha\text{-}LA}$). É stato inoltre ottenuto un fattore in grado di descrivere lo stato di salute della mammella bovina ($F6_{Udder\ health}$), basato principalmente sulla produzione di lattosio e di altri composti azotati e sulle cellule somatiche.

In generale, i risultati nell’analisi FA sono risultati coerenti con l’attribuzione del significato biologico dato al fattore. La maggior parte degli Fs è risultata significativamente influenzata dallo stadio di lattazione, seguito dall’ordine di parto. Sono state inoltre riscontrate correlazioni genetiche rilevanti tra i fattori (CAPITOLO 4). Nell’analisi GWAS tutti gli Fs hanno mostrato associazioni significative ($P < 5 \times 10^{-5}$), ad eccezione di $F5_{Yield}$. I picchi elevati su BTA6 (~ 87Mbp) e sulla coda di BTA11 (~ 104Mbp) erano principalmente associati a $F6_{\beta\text{-}\kappa\text{-}CN}$ e $F1_{Cheese}$

$_{N}$, rispettivamente. Inoltre, 33 termini GO e 6 categorie KEGG sono risultati arricchiti e associati con $F1_{\%\,CY}$, $F4_{Cheese\,N}$, $F8_{\alpha s2\text{-}CN}$ e $F10_{\alpha\text{-}LA}$. Le vie di segnalazione biologica descritte dai fattori erano principalmente correlate alle categorie più generali di attività ionica, neuroni e giunzioni occludenti. Poichè un numero considerevole di categorie arricchite GO e KEGG è risultato associato al fattore $F8_{\alpha s_2\text{-}CN}$, maggiore attenzione dovrebbe essere posta sulla frazioneαs_2-CN (CAPITOLO 5).

ABSTRACT

The aim of this PhD thesis was the study of the genomic, biological and phenotypic background of bovine cheese-related traits. The primary goal of this PhD thesis was to unravel the genomic background of bovine milk technological and cheese-related traits to specific chromosomic regions (CHAPTERS 1 to 3). To achieve this, the cow's ability to produce cheese was decomposed into 11 milk coagulation (MCP) and curd-firming properties (CF_t), and 7 cheese yield and milk component recoveries into the curd (REC) traits. Besides, to tackle the problem of the large number of variables required to describe the cow's ability to produce cheese, posing restrictions in the construction of selection indices, and thereby selection decisions, factor analysis (FA) was used (CHAPTERS 4 and 5).

The MCP traits were: 3 traditional single point lacto-dynamographic properties (RCT: rennet coagulation time, min; k_{20}: time to a curd firmness (CF) of 20 mm, min; a_{30}: CF 30 min after rennet addition), 6 parameters modeling 360 CF data for each milk sample (CF_P: potential asymptotic CF at infinite time, mm; k_{CF}: curd firming instant rate constant, $\%\times min^{-1}$; k_{SR}: syneresis instant rate constant, $\%\times min^{-1}$; RCT_{eq}: RCT from modeling; CF_{max}: maximum CF, mm; t_{max}: time at CF_{max}, min), milk- protein (%) and pH. The 3 CY traits were the weight (wt) of fresh curd ($\%CY_{CURD}$), curd solids ($\%CY_{SOLIDS}$), and curd moisture ($\%CY_{WATER}$) as % of wt of milk processed. The 4 REC (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, and REC_{ENERGY}) were calculated as the % ratio between the nutrient in curd and the corresponding nutrient in processed milk. For FA 26 traits related to milk yield and quality (including milk protein fractions), MCP-CF_t and CY-REC traits were analyzed.

Single marker genome-wide association analyses (GWAS) complemented by gene-set enrichment and pathway-based analyses were conducted. In total, 1,152 Italian Brown Swiss cows reared in 85 herds were genotyped with the Illumina SNP50 Beadchip v.2. Single marker regression GWAS were fitted using the GenABEL R package (GRAMMAR-GC). The Gene Ontology (**GO**) and Kyoto Encyclopedia of Genes and Genomes (**KEGG**) pathway databases were queried for the enrichment analyses.

In GWAS (CHAPTERS 1 and 2), sharp peaks were detected on *Bos taurus* autosome (BTA) 6, at 84 to 88 Mbp, with the highest peak detected at 87.4 Mbp in the region harboring the casein genes and more precisely of κ-CN (*CSN3)*. Marker Hapmap52348-rs29024684 (~87.4 Mbp), closely located to the casein genes on BTA6, was strongly associated with REC_{FAT} ($P = 1.91\times10^{-15}$) and CF_P ($P = 1.62\times10^{-17}$). Evidence of quantitative trait loci at 82.6 and 88.4 Mbp on the same chromosome was found. On BTA11, marker ARS-BFGL-NGS-104610 (~104.3 Mbp) was highly associated with $REC_{PROTEIN}$ ($P = 6.07\times10^{-36}$). Apart from BTA6 and 11, SNP located in 15 more chromosomes (1, 2, 9, 12, 13, 14, 15, 16, 18, 19, 20, 23, 26, 27 and 28) were significantly associated to the MCP-CF_t and CY-REC traits.

The gene-set enrichment and pathway-based analysis (CHAPTER 3) revealed 21 GO and 17 KEGG categories significantly associated (false discovery rate controlled at 5%) with 7 of the traits (RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS} and REC_{ENERGY}), with some being in common between traits. The significantly enriched categories included calcium signaling pathway, salivary secretion, metabolic pathways, carbohydrate digestion and absorption, the tight junction and the phosphatidylinositol pathways, as well as pathways related to the bovine mammary gland health status, and contained a total of 150 genes located in all chromosomes but 9, 20, and 27.

In FA (CHAPTERS 4 and 5), ten mutual orthogonal Fs were obtained using a *varimax* rotation. The 10 Fs explaining 74% of the original variability. Those Fs captured basic concepts of the "*milk to cheese*" process. More precisely, the first four Fs, sorted by variance explained, were able to capture the underlying structure of the CY percentage ($F1_{\%CY}$), the CF process with time ($F2_{CFt}$), the milk and solids yield ($F3_{Yield}$) and the presence of nitrogen (N) into the cheese ($F4_{Cheese\ N}$). Moreover, 4 Fs ($F5_{\alpha s_1\text{-}\beta\text{-}CN}$, $F7_{\beta\text{-}\kappa\text{-}CN}$, $F8_{\alpha s_2\text{-}CN}$, $F9_{\alpha s_1\text{-}CN\text{-}P}$) were related to the basic milk caseins (as_1-CN, as_2-CN, β-CN, κ-CN, and the phosphorylated form of as_1-CN) and 1 factor was associated with the α-LA whey protein ($F10_{\alpha\text{-}LA}$). A factor describing the udder health status of a cow ($F6_{Udder\ health}$), mainly loaded on lactose, other nitrogen compounds and SCS, was also obtained.

In general, FA results were coherent to the given name of the factor. Stage of lactation had a significant effect for the majority of the Fs, followed by parity. Moreover, considerable genetic correlations existed among the Fs (CHAPTER 4). All Fs showed significant associations ($P < 5 \times 10^{-5}$) in GWAS, but $F5_{Yield}$. High peaks on BTA6 (~87Mbp) and at the tail of BTA11 (~104Mbp) were mainly associated to $F6_{\beta\text{-}\kappa\text{-}CN}$ and $F1_{Cheese\ N}$, respectively. In addition, 33 GO terms and 6 KEGG categories were mainly enriched for $F8_{\alpha s_2\text{-}CN}$, but also for $F1_{\%CY}$, $F4_{Cheese\ N}$, and $F10_{\alpha\text{-}LA}$. Biological pathways were mainly related to the broader categories of ion activity, neurons and the tight junction. Moreover, the considerably large number of enriched GO and KEGG terms for $F8_{\alpha s_2\text{-}CN}$ suggests that, perhaps, more focus should be given on αs_2-CN (CHAPTER 5).

GENERAL INTRODUCTION

Milk and dairy products are important components of the human diet. The last decades the proportion of bovine milk destined for manufactured products, e.g. cheese, is steadily increasing in many countries worldwide (Food and Agriculture Organization of the,United Nations, 2015).

Cheese yield is the outcome of a complex process on which a variety of interrelated factors, derived from different disciplines (e.g. microbiology, physical chemistry, engineering process, etc) are involved, such as: i) the quantity (e.g. caseins and fat) as well as the quality (e.g. the fraction of the caseins to the total milk proteins, and the dimension of fat globules) of milk components; ii) other milk characteristics (e.g. milk acidity, minerals and microbial flora); iii) milk pre-treatments (e.g. milk natural creaming, heat treatment, etc.); and iv) cheese-making conditions (e.g. coagulation temperature, type and concentration of rennet, curd cutting, curd cooking, and pressing).

Milk characteristics, e.g. acidity and solid components (in particular casein and fat) are considered as the cornerstone of cheese-making, and their role in this process has been previously investigated (Walstra et al., 2014). In addition, cheese process strongly depends on milk coagulation (clotting of milk by rennet enzymes) as well as the syneresis (shrinkage of the curd with expulsion of whey). Milk coagulation after rennet (or similar coagulation agents) addition is the first step to cheese production. Therefore, milk coagulation properties (MCP) such as rennet coagulation time (RCT, min), time to curd firmness of 20 mm (k_{20}, min) and curd firmness 30 min after rennet addition (a_{30}, mm) are important factors for the description of cheese manufacture. In addition, previous analyses have shown important genetic variation of the MCP

traits [for a recent review on MCP genetics see (Bittante et al., 2012)]. The heritability of MCP is higher compared to milk yield and similar to other quality traits of milk.

Nevertheless, traditional MCP measures are problematic from a practical point of view. For example, due to late or non-coagulating samples (especially in Holstein-Friesian cows) it might be difficult to obtain the MCP values (Cecchinato et al., 2011), while those measurements show low repeatability as well (Bittante et al., 2012). In addition, a_{30} is strongly related to RCT (both phenotypically and genetically), thus providing no extra information (Ikonen et al., 2004; Bittante et al., 2012). Further, MCP traits partly describe the coagulation process (i.e. the change of the curd firmness over time – CF_t). To overcome the problems related to the classical single point estimates of MCP, such as late and non-coagulating milk samples and low repeatability, it has been proposed to model the CF as a function of time (Bittante, 2011; Bittante et al., 2013b). In this way, CF values are estimated over a longer time period through a model equation, thus providing extra information of the coagulation and CF processes (which also takes into account the phenomenon of syneresis).

Despite this, milk components and MCP-CF_t traits can only be used as indicators of the cheese-making process. On the other hand, traits like the quantity of cheese obtained from a given amount of processed milk, or the recovery of milk components into the cheese are direct measures of the cheese-making aptitude of milk and so of great economic interest. However, although there is a considerable literature on cheese-making, knowledge is mostly based on bulk milk. The importance of the percentage of cheese yield (%CY) at the individual level (i.e., based on individual milk and not on bulk milk) has been pointed out by (Othmane et al., 2002). Moreover, Banks (2007) discussed the significance of the milk constituents' recovery into the curd, as well as their loss in the whey for improved cheese quantity and quality. Previous studies have explored

the potential of individual %CY using bovine milk, albeit based on relative small numbers of individuals (Hurtaud et al., 1995; Wedholm et al., 2006). Recently, a large dataset (n = 1,264) of different measures of individual cow %CYs and milk nutrient and energy recovery in the cheese (REC) became available, using a cheese-making model approach assessed at the lab level (Cipolat-Gotet et al., 2013). Further analysis showed important genetic variation in individual %CY that does not solely depend on milk components but also heavily relies on the recovery of milk components in the curd (Bittante et al., 2013a).

Nevertheless, integration of the new knowledge, as well as of the MCP-CF_t traits, into breeding programs is hampered by high costs, intensive labour requirements, and lack of appropriate technology. At present, few potential alternatives have been suggested to overcome this problem, e.g., prediction of the aforementioned traits through infrared spectroscopy. Determination of milk components using spectral data is routinely used in the dairy industry (International Committee for Animal Recording (ICAR), 2012) and ongoing research is focused on the precision of the technology for predicting detailed milk components (Rutten et al., 2009), technological traits of milk such as milk coagulation (Cecchinato et al., 2009; Chessa et al., 2014) or different cheese measures (Ferragina et al., 2013; Bittante et al., 2014).

Genomic information offers a unique potential for a better understanding of the genetics underlying complex traits. A first step towards this direction is, for example, the application of genome wide association studies (GWAS) (McCarthy et al., 2008; Visscher et al., 2012) where thousands of DNA markers, in the form of single nucleotide polymorphisms (SNP), are scanned throughout the entire genome (The Bovine Genome Sequencing and Analysis Consortium et al., 2009) linking the phenotype of interest to specific regions on the genome (Goddard and Hayes, 2009). In addition, genomic information can be used in marker-assisted or genomic selection

breeding programs (Goddard and Hayes, 2009; de los Campos et al., 2013; Van Eenennaam et al., 2014) or within the emerging technology of genome/gene editing (Jenko et al., 2015; Proudfoot et al., 2015).

However, despite the potential of GWAS in identifying genomic regions associated with the traits of interest, limitations exist and new challenges are ahead. For instance, due to the stringent statistical thresholds used to deal with multiple testing, a considerable number of important markers may remain undetected when dealing with polygenic traits (Peng et al., 2010). Moreover, with high SNP density panels each gene might be represented by several proximal SNP, thus splitting its effect into parts that, in turn, might not be able to pass the defined GWAS threshold in a single marker regression (Ha et al., 2015). Additionally, especially in livestock species linkage disequilibrium (LD) spans a wide region in the genome. As a result, a plethora of SNP might be in LD with the causal genomic region which creates extra difficulties in detecting the causal mutation (Hayes, 2013). Besides, while GWAS may be able to locate SNPs significantly associated with the trait of interest, it does not make use of the fact that genes work together in biological pathways and are organized into networks. Further, the effect of a multi-allelic QTL may be not fully captured due to the bi-allelic nature of SNP. As a result, GWAS alone may provide a limited understanding of the complex nature of quantitative traits.

A solution to alleviate the aforementioned problems, and deepen the understanding of the genetic background of complex traits, is to move up the analysis from the SNP to the gene and gene-set levels. In a gene-set analysis, a group of related genes (such as genes in a specific pathway or gene ontology) that harbor significant SNPs previously identified in GWAS, is tested for over-representation in a specific pathway (Wang et al., 2011). Indeed, an increasing interest on pathway

analysis has been recently observed in dairy cattle, to complement GWAS analyses of quantitative traits (Gambra et al., 2013; Peñagaricano et al., 2013; Iso-Touru et al., 2016; Abdalla et al., 2016).

Despite the recent technological advances, that offer detailed phenotyping required to deepen our biological knowledge of the traits, from a practical point of view animal breeding faces the following challenge: to simultaneously improve a variety of different traits in consequent generations. Selection indices usually contain a large number of phenotypes related with production, reproduction and health, etc. (Miglior et al., 2005). With the recent technological advances, this number is very likely to drastically increase in the following years (Boichard and Brochard, 2012). However, a large number of traits probably results in a complex phenotypic and genetic correlation structure, which, in turn, poses restrictions in selection decision as well as in computations.

Dealing with large amount of data, multivariate analysis is a usual candidate to reduce the data space. Within this context, factor analysis (FA) is preferable when the aim is to identify latent structures (factors; Fs) of correlated variables. This characteristic of FA has attracted the interest in animal breeding. For instance, the potential use of Fs obtained from FA has been investigated for a variety of traits, such as milk quality, milk technological properties, e.g., MCP and cheese-CY in bovine and sheep (Macciotta et al., 2012; Manca et al., 2016) as well as milk fatty acids in dairy cattle (Conte et al., 2016). In addition, Fs have been used within the framework of structural equation modelling for the analysis of carcass traits in pigs (Peñagaricano et al., 2015) and bovine mastitis (Detilleux et al., 2013).

The main idea of FA is that n observed variables, x, can be expressed as linear functions of p ($p < n$) latent variables. FA focuses in understanding relationships (the underlying latent concept that the measured variables represent) among a set of observed variables. Based on the observed

covariance structure, FA aims in capturing the underlying latent concept that the original variables represent (Bollen, 2014). Thus, data reduction is attained while at the same time only the underlying concept of interest is kept for further analysis.

Although FA has been used in studies in animal breeding those were mainly focused in the interpretation of the Fs, investigating the sources of variation related to the Fs and estimating their genetic parameters. However, the potential use of Fs in the genomic era, e.g. in genome wide associations (GWAS) and its counterpart, gene-set enrichment analysis, has been unexplored.

AIMS OF THE THESIS

The general objective of this PhD thesis was to investigate the genomic biological and phenotypic background of a cow's ability to produce cheese and to integrate phenotypic information of cheese-related traits ("*milk to cheese*" process) for practical breeding applications focusing on increased bovine cheese yield. More precisely, the five objectives/contributions of this thesis were:

The aim of the first contribution (CHAPTER 1) was to apply GWAS on traditional single-point MCP observations in connection to CF_t, in an effort to shed more light in the genomic background of cheese-making related traits. Milk acidity and protein percentage were also considered.

The objective of the second contribution (CHAPTER 2) was to conduct a GWAS analysis using individual cheese yield ($\%CY_{CURD}$, $\%CY_{SOLIDS}$ and $\%CY_{WATER}$) and milk nutrient and energy recovery into the curd measures (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, and REC_{ENERGY}) to shed light on the genetics underlying a cow's cheese-making ability.

The goal of the third contribution (CHAPTER 3) was to conduct a gene ontology and pathway analysis, to complement the obtained GWAS results from the first and second contributions, for phenotypes related to bovine MCP, CF_t, CY and REC or whey loss traits.

The objective of the fourth contribution (CHAPTER 4) was to create a new set of latent phenotypes related to milk quality and technological traits, using FA, and to assess their potential use in dairy cattle breeding by i) studying the sources of variation (i.e., dairy system, feeding, herd, parity and stage of lactation) related to the new variables and ii) estimating their genetic parameters. The phenotypes in our analysis were selected to represent major components of cheese

yield but also traits that are commonly included in selection indexes in dairy cattle breeding programs.

Integrating all the previous work, the objective of the fifth contribution (CHAPTER 5) was to conduct genome-wide associations and pathway analysis with a set of latent variables related to 26 milk yield and quality, curd firming characteristics, and individual cheese properties.

To achieve our goals, milk samples from 1,264 Italian Brown Swiss cows were collected from 85 herds located in Trento Province, north-east of Italy. Not all phenotyped animals had blood samples available. In total, 1,152 cows were genotyped using the Illumina BovineSNP50 v.2 BeadChip (Illumina Inc., San Diego, CA). The dataset contained detailed milk yield and quality traits, milk protein composition, milk coagulation and curd firmness phenotypes and cheese-making traits measured through individual model-cheese manufacture.

CHAPTER 1

Genome-wide association of coagulation properties, curd firmness modeling, protein percentage, and acidity in milk from Brown Swiss cows

C. Dadousis,* S. Biffani,† C. Cipolat-Gotet,* E.L. Nicolazzi,‡ A. Rossoni[§], E. Santus[§], G. Bittante,* and A. Cecchinato*

JOURNAL OF DAIRY SCIENCE, 2016, VOL. 99:3654–3666

http://dx.doi.org/10.3168/jds.2015-10078

*Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

† Istituto di Biologia e Biotecnologia Agraria (IBBA) - Consiglio Nazionale delle Ricerche (CNR),, and

‡ Bioinformatics core, Parco Tecnologico Padano, Via Einstein- Loc. Cascina Codazza, 26900 Lodi, Italy

[§]Italian Brown Breeders Association, Loc. Ferlina 204, Bussolengo 37012, Italy

ABSTRACT

Cheese production is increasing in many countries and a desire towards genetic selection for milk coagulation properties in dairy cattle breeding exists. However, measurements of individual cheese–making properties are hampered by high costs and labour while traditional single point milk coagulation properties (MCP) are sometimes criticized. Nevertheless, new modeling of the entire curd firmness and syneresis process (CF_t equation) offers new insight of the cheese-making process. Moreover, identification of genomic regions regulating milk cheese–making properties might enhance direct selection of individuals in breeding programs based on cheese ability rather than related milk components. Therefore, the objective of this study was to perform genome wide association studies (GWAS) to identify genomic regions linked to traditional MCP and new CF_t parameters, milk acidity (pH) and milk protein percentage (Prot%). Milk and DNA samples from 1,043 Italian Brown Swiss cows were used. Milk pH and three MCP traits (RCT, k_{20}, a_{30}) were grouped together to represent the MCP-set. Four CF_t equation parameters (RCT_{eq}, CF_P, k_{CF}, k_{SR}), two derived traits (CF_{max} and t_{max}) and Prot% were considered as the second group of traits (CF_t -set). Animals were genotyped with the Illumina SNP50 bead-chip v.2. Multi-trait animal models were used to estimate variance components. For GWAS the GRAMMAR-GC approach was used. In total, 106 significant marker-traits associations and 66 SNP were identified on 12 chromosomes (1, 6, 9, 11, 13, 15, 16, 19, 20, 23, 26 and 28). Sharp peaks were detected at 84-88Mbp on *Bos taurus* autosome (BTA) 6, with a peak at 87.4Mbp in the region harboring the casein genes. Evidence of QTL at 82.6 and 88.4Mbp on the same chromosome was found. All chromosomes but BTA6, BTA11 and BTA28 were associated to only one trait. Only BTA6 was in common between MCP and CF_t sets. The new CF_t traits reinforced the support of MCP signals and provided with additional information on genomic regions that might be involved in regulation of the coagulation

process of bovine milk.

Key words: genome-wide association study, milk coagulation, curd firmness, dairy cattle

INTRODUCTION

Milk composition (e.g. fat and protein content) as well as other milk features, e.g. its acidity (pH), are considered as the base for cheese manufacturing (Walstra et al., 2014). Moreover, cheese process strongly depends on milk coagulation (clotting of milk by rennet enzymes) as well as the syneresis (shrinkage of the curd with expulsion of whey). Milk coagulation after rennet (or similar coagulation agents) addition is the first step to cheese production. Therefore, milk coagulation properties (MCP) such as rennet coagulation time (RCT, min), time to curd firmness of 20 mm (k_{20}, min) and curd firmness 30 min after rennet addition (a_{30}, mm) are important factors for the description of cheese manufacture. In addition, previous analyses have shown important genetic variation of the MCP traits [for a recent review on MCP genetics see (Bittante et al., 2012)]. The heritability of MCP is higher compared to milk yield and similar to other quality traits of milk.

To overcome the problems related to the classical single point estimates of MCP, such as late and non-coagulating milk samples and low repeatability, it has been proposed to model the CF as a function of time (Bittante, 2011; Bittante et al., 2013b). In this way, CF values are estimated over a longer time period through a model equation, thus providing extra information of the coagulation and CF processes (which also takes into account the phenomenon of syneresis). An extra difficulty is related with the wide scale recording of MCP values that is required for application in breeding programs, yet the phenotyping of MCP is highly costly and labour demanding. For a wide application (at a population level) of MCP, infrared spectroscopy has been promising (Cipolat-Gotet et al., 2012; Cecchinato et al., 2013; Chessa et al., 2014). An alternative,

is the identification of genomic regions regulating the aforementioned traits, linking the desired traits to the genome, which in turn may enhance establishment of marker-assisted selection programs or breeding programs based on whole-genome predictions (Van Eenennaam et al., 2014). To this purpose an experimental trial using daughter design and selective genotyping identified significant associations on chromosomes 2, 18 and 24 using coagulation as a binary trait, i.e. coagulating vs. noncoagulating milk in Finnish Ayrshire cattle (Tyrisevä et al., 2008). Significant associations of the β-lactoglobulin (*LGB*), beta casein (*CSN2*) and growth hormone 1 (*GH1*) with RCT have already been reported in candidate gene studies (Bonfatti et al., 2010; Cecchinato et al., 2015b). Moreover, pH has been associated to Rho GTPase activating protein 35 (*GRLF1*), the lipase gene (*LIPE*) and *SCD-1* (acyl-CoA desaturase) (Cecchinato et al., 2015b). Cheese yield as well as milk coagulation properties have also been associated to leptin (*LEP*), leptin receptor (*LEPR*) and kappa casein (*CSN3*) (Glantz et al., 2011). Recently, the CF and syneresis traits have also been tested for genomic associations resulting in new candidate genes regulating cheese–making properties of the milk, additional to those identified by the traditional MCP measures (Cecchinato et al., 2015b). Note, however, that all the above studies were performed on a small number of preselected DNA markers and not on a whole genome scale. With genome wide association studies, where a large number of SNP (single nucleotide polymorphisms) distributed along the whole genome is used, new, previously unknown, chromosomal regions associated to traits under investigation can potentially be identified (Schopen et al., 2011).

Exploration of the genetic background and identification of genomic regions affecting milk coagulation and curd firmness might be useful for establishing gene-assisted selection programs or incorporate the new knowledge for direct genomic prediction purposes (Glantz et al.,

2012). A first attempt on GWAS for MCP traits has been recently presented using a high density SNP chip but a relatively small number of individuals (379 cows) (Gregersen et al., 2015).

The aim of the present study was to apply GWAS on Italian Brown Swiss dairy cows genotyped with a 50k SNP chip. Traits investigated were traditional single-point MCP observations in connection to curd firmness and syneresis traits, in an effort to shed more light in the genomic background of cheese-making related traits. Milk acidity and protein percentage were also considered.

MATERIALS AND METHODS

Field Data

Milk samples from 1,264 Italian Brown Swiss cows were collected from 85 herds located in Trento Province, north-east of Italy. With few exceptions, 15 cows from each herd were individually sampled once during evening milking. After collection, milk samples (without preservative) were immediately refrigerated (4°C). One random subsample was transported to the Milk Quality Laboratory of the Breeders Association of Trento Province (Trento, Italy) for composition analysis. The other subsample was transferred to the Cheese-Making Laboratory of the Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE) of the University of Padova (Legnaro, Padova, Italy) for milk MCP analysis. All samples were processed within 20h after collection. Information on cows and herds were provided by the Breeders Association of Trento Province (Italy). Phenotypic data were matched to pedigree information supplied by the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy).

Analysis of Milk Quality and MCP

Individual milk subsamples were analyzed for fat, protein, and casein contents using MilkoScan FT6000 (Foss). The pH of the subsamples was measured before MCP analysis, using a Crison Basic 25 electrode (Crison, Barcelona, Spain).

Measures of MCP were obtained using the Formagraph instrument (FRM) by Foss Electric A/S according to the procedure described in (Cipolat-Gotet et al., 2012). In brief, milk samples (10 mL) were heated to 35°C and 200 μL of a rennet solution [Hansen Standard 160, with 80 ± 5% chymosin and 20 ± 5% pepsin; 160 international milk clotting units (IMCU)/mL; Pacovis Amrein AG, Bern, Switzerland], diluted to 1.6% (wt/vol) in distilled water, was added at the beginning of analysis. Ten samples were analyzed simultaneously, one sample for each measuring unit of the coagulation meter (pendula), which records the width (mm) of the graph during testing every 15 sec. The observation period continued for 90 min after rennet addition. Rennet coagulation time (RCT) is defined as the time (min) from addition of enzyme to the beginning of coagulation, k_{20} (min) is the interval from RCT to the time at which a curd firmness of 20 mm is attained, and a_{30} (mm) is a measure of the extent of curd firmness 30 min after coagulant addition. Samples that did not coagulate within 30 min were classified as noncoagulating (Ikonen et al., 1999), although extension of analysis allowed RCT and k_{20} values to be detected for all samples.

Modeling the CF of Individual Milk Samples

Files containing 360 CF values for each milk sample, recorded every 15 sec for 90 min, were retrieved and used to estimate a set of parameters of CF at time t (CF_t) according to equations and methodology developed by Bittante (Bittante, 2011) and Bittante et al (Bittante et al., 2013b).

Estimated parameters included: rennet coagulation time (RCT_{eq}, min), potential asymptotical curd firmness (CF_P, mm), representing the maximum potential curd firmness of a given sample after infinite time in the absence of syneresis, curd-firming rate constant (k_{CF}, % x min^{-1}) which measures the relative velocity of CF, syneresis rate constant (k_{SR}, % x min^{-1}), maximum curd firmness (CF_{max}, mm) and time to CF_{max} (t_{max}, min). A graphical representation of the aforementioned parameters is reported in Figure 1.

Genotyping

Not all phenotyped animals had blood samples available. In total, 1,152 cows were genotyped using the Illumina BovineSNP50 v.2 BeadChip (Illumina Inc., San Diego, CA). SNP that fulfilled the following criteria were included in the analysis: (1) call rate >95%, (2) minor allele frequency >0.005, and (3) no extreme deviation from Hardy-Weinberg equilibrium (HWE; $p > 0.001$, Bonferroni corrected). After quality control 37,418 SNP, distributed across 29 autosomes and X-chromosome, and 1,043 animals were retained. In total, 109 animals were excluded from the analysis due to low call rate (<95%).

Genetic Parameters and Yield Deviation Estimation

Two multi-trait animal models were used to estimate variance components. The first model (MCP-set) was used to analyze the following traits: milk pH, RCT (min), k_{20} (min) and a_{30} (mm). The second model (CF_t-set) included milk-protein percentage (Prot%) together with the four traits modeling the curd-firming process over time, namely RCT_{eq} (min), CF_P (mm), k_{CF} (% x min^{-1}), k_{SR}, (% x min^{-1}) and two traits estimated from the CF equation, CF_{max} (mm) and t_{max} (min). Cows without pedigree and with a lactation longer than 400 days in milk (DIM) were discarded.

According to previous findings (Cecchinato et al., 2013), models for MCP- and CF_t -sets accounted for the following fixed effects: DIM (14 classes of days in milk defined as one class for every 30 d), parity (1 to 5 or more), and renneting meter sensor of the lactodynamograph (15 levels). The latter effect was fitted only for the a_{30} trait. Two random effects were included in both models: an uncorrelated random herd effect and an additive genetic effect. In total, 18,905 animals were included in the pedigree.

The basic model to describe the observations written in matrix notation was:

$$y = X\beta + Z_h h + Z_u u + e \quad (1)$$

where **y** is a vector of MCP-set or CF_t -set traits, **β** is the vector of fixed effects, **h**, **u** and **e** are vectors of random herd, additive genetic and residual effects, respectively. The **X** is the design matrix of fixed effects, $\mathbf{Z}_h$ and $\mathbf{Z}_u$ represent the corresponding incidence matrices linking the phenotypic records to the appropriate random effects. The following normal distributions were assumed for the random effects, $h \sim N(0, \mathbf{I} \otimes \mathbf{H_0})$, $u \sim N(0, \mathbf{A} \otimes \mathbf{G_0})$ *and* $e \sim N(0, \mathbf{I} \otimes \mathbf{R_0})$. The (co)variance structure of the random effects was as follows:

$$Var\begin{vmatrix}\mathbf{a}\\ \mathbf{h}\\ \mathbf{e}\end{vmatrix} = \begin{vmatrix}\mathbf{A} \otimes \mathbf{G_0} & 0 & 0\\ 0 & \mathbf{I} \otimes \mathbf{H_0} & 0\\ 0 & 0 & \mathbf{I} \otimes \mathbf{R_0}\end{vmatrix}$$

where **A** is the numerator relationship matrix; $\mathbf{G}_0$ and $\mathbf{H}_0$ are 4 x 4 or 7 X 7 matrices of polygenic and random herd (co)variances of the traits in models for MCP-set and CF_t -set, respectively; **I** is the identity matrix and $\mathbf{R_0}$ is a diagonal matrix of residual variances corresponding to each trait. MCP-set model was implemented using the program AIREML90 (Misztal et al., 2002). Program MTJAAM version 3.8 was used to implement the model CF_t-set (Gengler et al., 1997).

Intra-herd heritabilities (h^2_{IH}) were defined as:

$$h_{IH}^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2} \quad (2)$$

where σ_a^2 and σ_e^2 are additive genetic and residual variances, respectively.

Additive genetic correlations (r_g) were estimated as:

$$r_g = \frac{\sigma_{ai,aj}}{\sigma_{ai}.\sigma_{aj}} \quad (3)$$

where, $\sigma_{ai,aj}$ is the additive genetic covariance between trait *i* and *j*, and σ_{ai} and σ_{aj} are the additive genetic standard deviations for trait *i* and *j*, respectively.

The estimated variance and covariance components were used to estimate Yield Deviations (YD) for the 11 traits. The analyses were carried out using the BLUPF90 program (Misztal et al., 2002). The YD were estimated by adjusting daughter performance for all non genetic effects (Mrode, 2005):

$$YD = (Z_u' {R_0}^{-1} Z_u)^{-1} Z_u' {R_0}^{-1} \left(y - X\hat{\beta} - Z_h\hat{h}\right) \quad (4)$$

where YD is a vector of yield deviations and $\mathbf{R_0}$, $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, $\mathbf{Z}_u$ and $\mathbf{h}$ were as earlier defined in equation 1.

Genome-Wide Associations

The YD, obtained as previously described, were considered as response variables to perform genome-wide associations (GWAS) using the GenABEL package in R (R Core Team, 2013; GenABEL project developers, 2013) adopting the GRAMMAR-GC (Genome wide Association using Mixed Model and Regression - Genomic Control) approach (Amin et al., 2007; Svishcheva et al., 2012). The R package "qqman" was used for graphical representation of GWAS results (Turner, 2014). The GRAMMAR-GC involves a three step analysis. In the first step of the procedure, an additive polygenic model is fitted to the adjusted data (YD), using the genomic relationship matrix estimated from SNP data. The reason for including a polygenic term is to obtain residuals that are not biased by the possible population structure, which in our case might be due to the presence of closely related animals. In the second step, associations between residuals and marker genetic polymorphisms are tested in a single-SNP linear regression model. Finally, the Genomic Control (GC) approach is applied to correct for conservativeness of the GRAMMAR procedure. Genomic control is based on the estimation of the ζ deflation factor:

$$\zeta = \frac{Median\ (T_1^2, T_2^2, \ldots T_N^2)}{0.456} \qquad (5)$$

which is the median of the squares of all genome-wide computed test statistics divided by the expected median of the test statistic under the null hypothesis of no association, assuming that the number of true associations is very small compared to the number of tests that are actually performed. *P*-values of $\leq 5 \times 10^{-5}$ were considered as significant associations (Burton et al., 2007). A search for genes located within 1Mbp of the most significant SNP was followed (Ensembl Bos taurus UMD3.1).

RESULTS

Descriptive Statistics

Descriptive statistics for MCP-set and CF_t -set traits are presented in Table 1. Traditional RCT from FRM had an average of 20.19 min vs. 19.74 min of RCT_{eq} estimated modeling the CF process over time. As expected, the latter presented a lower standard deviation (4.34 vs. 6.54) than RCT from FRM because it was estimated using individual equations. The average CF time (k_{20}) was 5.72 min (± 3.67) while CF at 30 min after enzyme addition was 29.19 mm (± 10.92).

The syneresis rate constant had values much lower than the curd-firmness rate constant (average 1.39 vs. 12.45 % min^{-1}, respectively), but its coefficient of variation was high (~ 40%). The extension of the observation period to 90 min permitted to identify the maximum value of CF (CF_{max}), and the time at which this value is attained (t_{max}), after addition of rennet to milk. On average, CF_{max} was 37.05 mm, with a coefficient of variation of 19.4 %. The maximum CF was achieved, on average, 40.68 min after rennet addition (t_{max}), being almost twice the RCT value, on average.

Genetic Parameters

Table 2 shows the variance components estimates and the heritabilities for the two blocks of considered traits, namely MCP-set and CF_t -set. The three MCP properties were moderately heritable, with h^2_{IH} ranging between 0.15-0.27. For the traits of CF_t -set heritabilities were, again, moderate (ranged between 0.21-0.26) except the CF_P and k_{SR} that had low h^2_{IH} (0.06).

The genetic correlations of the traits analyzed within MCP-set are presented in Supplementary Table S1. Genetic correlations among milk pH, RCT and k_{20} were positive

(unfavourable), ranging between 0.36-0.65. On the contrary, a_{30} was negatively correlated with milk pH, RCT and k_{20} (-0.48, -0.88 and -0.93, respectively).

Genetic correlations of the CF_t -set traits are summarized in Supplementary Table S2. The percentage of milk-protein was almost uncorrelated with half of the CF_t traits (RCT_{eq}, k_{CF} and t_{max}) while moderately correlated to the rest. Also, RCT_{eq}, apart from Prot%, was uncorrelated to k_{SR} and CF_{max}. The highest positive r_g (0.87) was found between RCT_{eq}-t_{max}, while the highest negative between t_{max}-k_{CF} (-0.87). The rest of the genetic correlations among the traits were moderate, ranging between ~0.20 to 0.79, in absolute values.

Results of the Genome-Wide Associations

A total of 106 significant marker-trait associations (66 significant SNP) were detected (Table 3, Supplementary Table S3). From those 106 cases, 81 significant marker-trait associations were identified on *Bos taurus* autosome (BTA) chromosome 6 (76%), while 46 cases were unique, i.e. SNP that were found to be significantly associated to only one trait. From those 46 cases, 17 belong to CF_{max} and 10 to Prot %. On BTA6, 25 out of the 46 cases were detected (54%). In total, significant associations were spotted on 12 chromosomes (1, 6, 9, 11, 13, 15, 16, 19, 20, 23, 26 and 28). The traits CF_P and CF_{max} showed the maximum number of significant associations (34 and 17, respectively). On the contrary, milk pH was the only trait without any signal. The most frequently identified region was on BTA6 between 85.42 to 88.44Mbp. For both sets of traits analysed, marker Hapmap52348-rs29024684 (87,396,306bp) showed the strongest association.

On BTA6 signals were distributed across five regions (Table 3, Figure 2). At the first (region 6a; 51.7Mbp) k_{20} was linked. In the region 6b (64.18Mbp) t_{max} showed one significant association. The trait CF_P was associated to a genomic region (6c) between 73.6-74.6Mbp. The

broader region on BTA6 between 77.52-88.44Mbp (region 6d) showed signals for all but milk pH and the syneresis rate constant traits, with *P*-values ranging between 3.89×10^{-05} (k_{20}) and 1.62×10^{-17} (CF_P). That was also the region with the strongest signals. At the tail of BTA6 (~113.5Mbp) Prot% showed a signal (region 6e). The only trait without any significant association on BTA6 was k_{SR}.

Apart from BTA6, signals in eleven more chromosomes were detected (Table 3, Figure 3). More precisely, a region on BTA1 at ~9.5Mbp was associated to RCT_{eq}. Signals of Prot % on BTA9 (~70 and 83.6Mbp), BTA20 (~17.4Mbp) and BTA28 (38.45Mbp) were identified. On BTA28 t_{max} was also linked at around 33.8Mbp. On BTA11 the region at ~86-88Mbp was associated to both RCT_{eq} and t_{max} with a peak at 87.7Mbp. One region on BTA13 (47.88Mbp) and two regions on BTA15 (14.2 and 55.5Mbp), were associated to RCT. One SNP on BTA16 (~76.3Mbp) was associated to CF_P, while a small region on BTA19 (~2.1-2.3Mbp) was linked to a_{30}. The syneresis rate had only two signals, both on BTA 23 at, approximately, 8.8 and 10.6Mbp, while k_{CF} was associated to BTA26 (~20.4Mbp).

With the exception of the region 6d on BTA6, the rest of the signals were mainly one trait – one genomic region associations with, relatively, mild strength. As a supporting evidence of the associations identified, the quantile-quantile (Q-Q), plots visualizing the distribution of the observed test statistics derived from the GWAS analyses, were checked (Supplementary Figure S1). Extreme departures in the tail of the distributions were observed for RCT and a_{30} from the MCP-set and for all the CF_t-set with the exception of k_{SR}.

Summarizing the GWAS results for the two sets of traits, significant SNP for the three MCP properties were mainly located on BTA6, with the majority of them mapped within or in close proximity to the casein cluster (~87.14-87.39Mbp) in a broader area at approximately 84.71-

88.44Mbp. The regions at 84.89, 85.42, and 87.39Mbp on BTA6 were significant for all three MCP properties. For the CF_t traits sharp peaks were found on BTA6 (~77.5-87.4Mbp) mainly associated with CF_P and CF_{max}. The majority of common SNP between CF_P-CF_{max} are laying in the area at ~84.89-87.40Mbp. Between MCP-set and CF_t -set traits only the BTA6 was in common and, more specifically, the region between 81-88.44Mbp.

Conditional analysis

Due to the wideness of the region 6d, a conditional analysis was carried out by fixing in the GWAS model the most significant marker on BTA6 (Hapmap52348-rs29024684; 87,396,306bp), to test if multiple QTL exist on BTA6 (Table 4). The region 6c as well as a part of the region 6d between 77.5-81Mbp were not confirmed in the conditional analysis. Mild signals remained at 81.6-84.7Mbp (for k_{20}, Prot%,CF_P and k_{CF}), at 87.15-88.8Mbp (for RCT_{eq}, Prot%, CF_{max} and k_{CF}) as well as the area 6e. The region 6a was replaced with a weak signal at 51.7Mbp (for k_{20}) while another weak association at 64.18Mbp (for t_{max}) was detected replacing the region 6b. Concerning the rest of the chromosomes, on BTA1 a_{30} was linked to a region at 29.7Mbp, while milk pH was associated to BTA8 (113.25Mbp). On BTA11, although the general region of the signals was the same, the peak was detected at 86.8Mbp. On BTA13, the association of RCT (at 47.9Mbp) was replaced with a link between k_{20} and a region at 57.2Mbp. No change on the rest of the results was observed, except marginal changes of the *P*-values and absence of signal on BTA16. It should be mentioned, that all the associations in the conditional analysis were weak with the majority of them around the threshold.

DISCUSSION

Genetic Parameters

The majority of the traits analyzed in this study, showed moderate heritabilities, which is favorable for breeding purposes. In general, the CF_t-set traits tended to have slightly higher h^2_{IH} compared to the MCP-set, except CF_P and k_{SR} that showed very low values. Genetic parameters reported here might slightly differ compared to a recently published work based on the same data (Cecchinato et al., 2015b), because of different editing of the data as well as to different approaches followed for variance component estimations, but rely on the range of heritability values of MCP reviewed by Bittante et al. (2012).

Genomic regions associated with MCP and CF_t traits

It is well established that milk protein (and more specifically the caseins) as well as milk fat content are strongly related to cheese yield and that casein genetic variants are affecting MCP (reviewed by Bittante et al., 2012). Moreover, it is already known that the casein genes are located on BTA6 (reviewed in (Caroli et al., 2009)(Caroli et al., 2009) while significant associations of milk technological traits have been reported on BTA11 as well (Heck et al., 2009; Bonfatti et al., 2010; Berry et al., 2010).

BTA6

Recently, a GWAS on rheological properties of milk, including RCT, was carried out using haplotypes reporting high peaks on BTA6 (Gregersen et al., 2015). The importance of the region on BTA6 at, roughly, 84 to 88 Mbp has also been confirmed in our study, although a broader area, starting from ~82Mbp, has been identified. The region at 88Mbp has been significantly associated

to the caseins (α, β, κ) as well as to beta-lactoglobulin content, Prot%, casein index and α-lactalbumin in another GWAS study (Schopen et al., 2011). In Gregersen et al (2015) signals around 50.9 and ~60.9Mbp have been detected, linking to the log transformed maximum gel strength and log(RCT), respectively. These regions are in close proximity to the regions 6a and 6b in our study (Table 3), but they were not confirmed in the conditional analysis (Table 4).

The main region identified in our study was at 84-88Mbp on BTA6 with a peak at 87.4Mbp. This area is gene-dense including the casein cluster: casein alpha S1 (*CSN1S1*; ~87.14-87.16Mbp), casein alpha S2 (*CSN1S2*; ~87.26-87.28Mbp), casein beta (*CSN2*; ~87.18-87.19Mbp) and casein kappa (*CSN3*; ~87.39Mbp). The highest significant SNP identified in our study (Hapmap52348-rs29024684) is located ~18kbp upstream of *CSN3*. Moreover, the marker Hapmap28023-BTC-060518 is located within the histatherin gene (*HSTN*; ~87,190-87,204Mbp) which in turn is mapped between *CSN2* and *CSN1S2* on BTA6. The *HSTN* gene is involved in salivary secretion. This SNP was significant for RCT, k_{20}, a_{30}, CF_P and CF_{max} with *P*:[1.32×10^{-5} , 1.41×10^{-11}], while remained significant after adjusting for the Hapmap52348-rs29024684 (*P*=3.65×10^{-5} for the CF_{max}).

Even after adjusting for the effect of the highest significant marker on BTA6 two weak signals remained at i) 81-84Mbp (with a peak at 82.6Mbp) and ii) 87-89Mbp (peak at 88.4Mbp), indicating the presence of a different QTL. In the first area *TECRL* (trans-2,3-enoyl-CoA reductase-like; ~81.5-81.6Mbp) is located, which is involved in lipid metabolic process. In the second region the *GC* (vitamin D-binding protein; 88.69-88.74Mbp) is mapped. The protein encoded by *GC* belongs to the general albumin family, it is a vitamin D binding protein and regulates vitamin transportation to target tissues and vitamin D metabolism. Moreover, it is involved in the metabolism of lipids and lipoproteins.

BTA11

Chromosome 11 was the second most frequently identified chromosome in the GWAS analysis where significant associations with RCT_{eq} and t_{max} were detected at around 87Mbp. In previous studies, QTL on BTA11 related to α-, β- and κ-caseins as well as α-lactalbumin and beta-lactoglobulin concentrations have been reported (Heck et al., 2009; Bonfatti et al., 2010; Berry et al., 2010) suggesting progestagen-associated endometrial protein (*PAEP*; also known as *LGB*) located at 103.3Mbp as a candidate gene. Also, association between RCT and *LGB* has been shown (Bonfatti et al., 2010; Cecchinato et al., 2012b). Moreover, the tail of BTA11 (~107Mbp) has been linked with high peaks to percentages of α-, β-, κ- caseins, beta-lactoglobulin and casein index (Schopen et al., 2011). None of those signals was confirmed in our study. Therefore, our signal at 87Mbp indicates a potential new QTL. A variety of protein codings is mapped in the identified in our study region between 85.9-88Mbp. In this area Lipin 1 (*LPIN 1*; ~86.0-86.1Mbp), grainyhead-like 1 (*GRHL1*; ~87.5-87.6Mbp) and isoamyl acetate-hydrolyzing esterase 1 homolog (*IAH1*; ~87.94-87.95Mbp) are present. All of the three genes are involved in lipid metabolic process. Associations of *LPIN 1* with k_{20} and a_{30} have been previously reported (Cecchinato et al., 2012a). Close to this region (~88.58Mbp) the DNA-binding protein inhibitor ID-2 (*ID2*) is mapped, which is involved in the mammary gland epithelium development. In addition, marker ARS-BFGL-NGS-37074 (88,028,793bp) is located within the *ASAP2* (arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2; ~88.01-88.18Mbp. This marker was significant, albeit at a weak strength.

Signals on other chromosomes apart from BTA6 and BTA11

Significant associations in ten more chromosomes were detected in our study. Apart from BTA6, six more chromosomes are in common between our study and Gregersen et al. (2015), namely BTA1, 9, 13, 15, 16 and 23. However, signals were identified in different genomic regions. For example, on BTA1 a signal at 9.5Mbp was found for RCT_{eq} while in the conditional analysis a_{30} was associated at 29.7Mbp. In Gregersen et al (2015) a QTL interval at 70.7-80Mbp on BTA1 is reported. On BTA15 our analysis identified two regions related to RCT at 14.2 and 55.5Mbp which are quite far apart from the locations reported in Gregersen et al. (2015) (37-44Mbp and 61-62Mbp, respectively). On BTA13 a significant SNP (*P*-value=1.45×10^{-5}) was detected at 47,9Mbp while a signal at 58Mbp was reported in Gregersen et al. (2015). Note, however, that in the conditional analysis a signal at 57Mbp was observed for k_{20}. These differences might not be surprising since the traits analysed in the two studies (with the exception of RCT), although all related to coagulation properties of the milk, are different.

Peaks on BTA2, 18 and 24 have also been reported in a previous study (Tyrisevä et al., 2008). No signal was found on those chromosomes in our analysis. However, in Tyrisevä et al (2008): i) the traits analysed were coagulation vs. non-coagulation, ii) a selective DNA pooling using microsatellite markers was used and iii) the analysis was carried out in a Finnish Ayrshire dairy cattle population. Moreover, while the above mentioned chromosomes in Tyrisevä et al. (2008) have been also reported in Gregersen et al. (2015) the positions of the signals do not match.

A possible consideration could be why *DGAT1* (acyl-coenzyme A:diacylglycerol acyltransferase 1), located on BTA14 and related to milk fat percentage, was not identified in our study. This was not surprising since in the Italian Brown Swiss population *DGAT1* is not segregating.

GWAS results and genetic correlations among traits

In the present study we found traits with high genetic correlations ($r_g > |0.7|$) showing different genomic signals. For example, RCT and a_{30} had a genetic correlation of -0.88 (Supplementary Table S1). Both traits showed a signal on BTA6 sharing 6 significant SNP in common. However, RCT was also associated to BTA13 and BTA15, while a_{30} was linked to BTA19. Another example is between RCT_{eq} and t_{max} which are strongly genetically correlated ($r_g = 0.87$). Three significant SNP were in common (one on BTA6 and two on BTA11), but RCT_{eq} was also associated to BTA1 while t_{max} to BTA28. In addition, RCT_{eq} – CF_P, CF_P – k_{SR} and CF_{max} – k_{SR} had no common significant SNP but they are strongly genetically related (r_g of 0.7, -0.79 and -0.74, respectively). These results could be dually interpreted. Thus, perhaps, the different genomic regions identified on highly genetically correlated traits could be attributed to the non-correlated part between the traits. On the other hand, it could also be claimed that the information of genetic correlation between related traits involved in the same biological process might be an extra hint on distinguishing false positives as well as supporting evidence of true signals. In any case, an extra benefit can be gained by taking into account the genetic correlations among phenotypes describing the same process in GWAS studies, resulting in a more holistic view of the complex trait under study.

Partitioning a complex phenotype into different components

In the genome wide associations performed in this study the phenotypes used were either three single point measurements that are traditionally used in dairy industry or they were obtained after allowing an extended coagulation period (Bittante et al., 2013b). Between the two groups of phenotypes only BTA6 was in common. However, for both sets, marker Hapmap52348-

rs29024684 was identified as the most significant. The CF_t traits strengthened the evidence of the obtained signals for MCP traits on BTA6, while also pinpointed genomic regions not captured by the MCP, for instance on BTA11. There is, therefore, evidence that by partitioning the traits into different (potentially correlated) components – phenotypes more information might be extracted concerning the underlying genetic mechanism regulating the traits. The importance of "phenomics" and phenotype definition has been recently emphasized as an essential part of information for the better understanding of the genomic regulation of traits (Houle et al., 2010).

Breeding

The clear peaks on BTA6 together with the identification of Hapmap52348-rs29024684 (closely located to *CSN3*) as highly significant in multiple traits indicate that, perhaps, more emphasis should be given for selection on this marker (or directly to *CSN3*). The minor allele frequency (MAF) of the marker in this dataset was 0.23. Moreover, depending upon the trait, the proportion of phenotypic variance captured by this marker ranged between 40% (for k_{20}) to 2% (for a_{30}) (Supplementary Table S3). The same marker has been significantly associated in a recent preliminary GWAS analysis, in the same population, with the recoveries of the milk-fat into the cheese, explaining a large proportion of the phenotypic and genetic variability (Dadousis et al., 2015). This provides extra support for the significance of this genomic region for cheese related properties of milk. Therefore, our results show that selection of this marker is possible for the Italian Brown Swiss population.

With the exception of the region 6d, the rest of the associations identified were mild to weak. This can be partly attributed to the limited sample size. As direct MCP measurements is costly and time requiring, much larger sample size seems not feasible. An alternative for increasing

the power of GWAS could be to use MCP and CF_t traits predicted by FTIR spectroscopy (Ferragina et al., 2015) that could be easily applied at population level (Cecchinato et al., 2015).

CONCLUSION

Eleven traits related to milk coagulation and curd firmness were used in genome wide association analyses. Even taking into account the limitations imposed by the sample size and the number of SNPs analyzed, the results confirmed the importance of the casein cluster on BTA6 and more precisely of *CSN3*, while potential new QTL appear at 82.6 and 88.4Mbp on the same chromosome. Evidence of a QTL on BTA11 at ~87Mbp has been found. Moreover, new signals along the genome have been detected. However these genomic regions should be reproduced in future studies, favourably of a larger scale. Despite this, to the best of our knowledge, this is the largest GWAS study carried out with milk coagulation properties.

The new set of seven traits describing the whole process of milk coagulation to syneresis strengthened the support of QTL on BTA6. Moreover, the CF_t traits showed signals on genomic regions not related to the traditionally measured MCP traits. Results indicate the importance of phenotypic integration in GWAS studies for a better connection between phenome and the genome. Replication of the GWAS results from independent, and favourably of large scale, studies remains crucial. This will also help to clear the picture of the weak signals detected in our study.

ACKNOWLEDGMENTS

Authors thank the Trento Province (Italy), the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy) and the Superbrown Consortium of Bolzano and Trento for financial and technical support. The first author benefitted from financial support of the CARIPARO (Cassa di Risparmio di Padova e Rovigo) Foundation.

Table 1. Descriptive statistics of milk coagulation properties (MCP), parameters of curd firmness modeling on time t (CF_t), milk pH and protein.

Trait[1]	Mean	SD	P1	P99	CV, %
MCP					
RCT (min)	20.19	6.54	10.30	41.00	32.4
k_{20} (min)	5.72	3.67	2.00	19.42	64.2
a_{30} (mm)	29.19	10.92	0.81	50.24	37.4
CF_t					
RCT_{eq} (min)	19.74	4.34	15.00	28.50	40.5
CF_P (mm)	54.24	13.84	26.03	96.94	25.5
k_{CF} (% × min^{-1})	12.45	5.69	2.36	28.71	45.7
k_{SR} (% × min^{-1})	1.39	0.56	0.15	2.85	40.5
CF_{max} (mm)	37.05	7.17	19.40	53.67	19.4
t_{max} (min)	40.68	10.43	23.27	72.96	25.7
Milk pH	6.64	0.08	6.42	6.84	1.2
Protein, %	3.72	0.41	2.91	4.71	11.1

P1 = 1st percentile; P99 = 99th percentile; CV, %: coefficient of variation (%); RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

Table 2. Estimates of additive genetic variances (σ_a^2), and intra-herd heritabilities (h_{IH}^2) for milk coagulation properties (MCP), parameters of curd firmness modeling on time t (CF_t), milk pH and protein.

Trait[1]	σ_a^2	h_{IH}^2
MCP		
RCT (min)	$10.40^{(0.50)}$	$0.27^{(0.01)}$
k_{20} (min)	$2.34^{(0.12)}$	$0.15^{(0.01)}$
a_{30} (mm)	$40.16^{(2.01)}$	$0.17^{(0.01)}$
CF_t		
RCTeq (min)	$3.84^{(0.62)}$	$0.26^{(0.03)}$
CF_P (mm)	$9.43^{(0.20)}$	$0.06^{(0.01)}$
k_{CF} (% x min^{-1})	$5.88^{(0.98)}$	$0.24^{(0.03)}$
k_{SR} (% x min^{-1})	$0.02^{(0.00)}$	$0.06^{(0.01)}$
CF_{max} (mm)	$6.96^{(1.31)}$	$0.21^{(0.03)}$
t_{max} (min)	$14.76^{(2.56)}$	$0.22^{(0.03)}$
pH	$0.002^{(0.00)}$	$0.45^{(0.02)}$
Protein, %	$0.02^{(0.04)}$	$0.24^{(0.05)}$

[1]RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

In parenthesis the standard errors of the estimates.

Table 3. Summary results of the genome wide association analyses

BTA	# SNP (signals)	Interval (Mbp)	*P*-value (range)	Top SNP	Top SNP location (bp)	Top SNP MAF	Trait[1]
0*	1	-	5.31×10^{-06}	BTA-76907-no-rs	0	0.26	CF_P
1	1	-	4.27×10^{-05}	ARS-BFGL-NGS-41048	9,484,167	0.09	RCT_{eq}
6a	1	-	2.57×10^{-05}	BTB-01451336	51,669,513	0.12	k_{20}
6b	1	-	4.65×10^{-05}	Hapmap43353-BTA-76584	64,179,687	0.05	t_{max}
6c	4 (4)	73.640-74.607	$(1.96 \times 10^{-05}, 8.14 \times 10^{-06})$	Hapmap43042-BTA-76779	73,688,640	0.20	CF_P
6d	36 (74)	77.521-88.442	$(3.89 \times 10^{-05}, 1.62 \times 10^{-17})$	Hapmap52348-rs29024684	87,396,306	0.24	Prot%, RCT, k_{20}, a_{30}, RCT_{eq}, **CF_P**, CF_{max}, k_{CF}, t_{max}
6e	1	-	2.2×10^{-5}	Hapmap47844-BTA-115673	113,538,490	0.25	Prot%
9a	1	-	9.04×10^{-06}	Hapmap53034-rs29011422	69,696,334	0.34	Prot%
9b	1	-	1.06×10^{-05}	ARS-BFGL-NGS-88859	83,575,446	0.38	Prot%
11	6 (8)	85.936-88.029	$(3.47 \times 10^{-05}, 5.72 \times 10^{-06})$	BTA-110429-no-rs	87,670,344	0.42	**RCT_{eq}**, t_{max}
13	1	-	1.45×10^{-05}	Hapmap31215-BTA-32775	47,879,982	0.01	RCT
15a	2 (2)	14.243-14.272	$(2.81 \times 10^{-06}, 2.77 \times 10^{-06})$	ARS-BFGL-NGS-114291	14,242,668	0.01	RCT
15b	1	-	6.76×10^{-06}	ARS-BFGL-NGS-68607	55,488,319	0.01	RCT
16	1	-	2.74×10^{-05}	ARS-BFGL-NGS-17574	76,311,292	0.29	CF_P
19	2 (2)	2.094-2.271	$(2.3 \times 10^{-05}, 2.08 \times 10^{-05})$	Hapmap39832-BTA-46468	2,093,500	0.32	a_{30}
20	1	-	3.57×10^{-05}	ARS-BFGL-NGS-1751	17,412,441	0.17	Prot%
23a	1	-	2.69×10^{-05}	Hapmap38418-BTA-57213	8,819,178	0.11	k_{SR}
23b	1	-	1.75×10^{-06}	ARS-BFGL-NGS-99929	10,631,079	0.07	k_{SR}
26	1	-	4.3×10^{-05}	ARS-BFGL-NGS-23064	20,365,711	0.47	k_{CF}
28a	1	-	2.1×10^{-05}	ARS-BFGL-NGS-115508	33,729,338	0.11	t_{max}
28b	1	-	2.64×10^{-05}	Hapmap48306-BTA-36540	38,449,335	0.13	Prot%

BTA= *Bos taurus* autosome chromosome; #SNP (signals)= number of the single nucleotide polymorphisms significantly associated to the trait. In parenthesis 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)= The *P*-value of the highest significant SNP adjusted for genomic control and the range of the *P*-values when multiple SNP were significantly associated to one 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 of the top SNP; [1]RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Prot % = protein percent; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

In bold the trait with the highest *P*-value in each genomic region.

*Undefined chromosome and position on the genome

Table 4. Summary results of the genome wide association analyses after fixing the marker Hapmap52348-rs29024684 at 87,396,306 base pairs on BTA6.

BTA	# SNP (signals)	Interval (Mbp)	*P*-value (range)	Top SNP	Top SNP location (bp)	Top SNP MAF	Trait[1]
1	1		4.46×10^{-05}	BTB-02049040	29,696,010	0.42	a_{30}
6a	1		4.99×10^{-05}	Hapmap23226-BTA-159656	46,599,570	0.24	t_{max}
6b	2	69,483-71,421	$(3.66\times 10^{-05}, 2.67\times 10^{-05})$	ARS-BFGL-NGS-20354	69,482,838	0.08	CF_{max}
6d_1	4(7)	81,652-84,690	$(4.73\times 10^{-05}, 5.47\times 10^{-06})$	Hapmap27307-BTC-043200	82,605,943	0.20	k_{20}, Prot%, **CF_P**, k_{CF}
6d_2	5(5)	87,153-88,822	$(4.9\times 10^{-05}, 9.22\times 10^{-06})$	BTA-122637-no-rs	88,442,145	0.07	**RCT_{eq}**, Prot%, CF_{max}, k_{CF}
6e	1		1.98×10^{-05}	Hapmap47844-BTA-115673	113,538,490	0.25	Prot%
8	1		4.97×10^{-05}	ARS-BFGL-NGS-63329	113,252,230	0.01	pH
9a	1		1.51×10^{-05}	Hapmap53034-rs29011422	69,696,334	0.34	Prot%
9b	1		7.38×10^{-06}	ARS-BFGL-NGS-88859	83,575,446	0.38	Prot%
11	6(10)	85,936-88,029	$(4.43\times 10^{-05}, 4.5\times 10^{-06})$	ARS-BFGL-NGS-119913	86,779,385	0.50	**t_{max}**, RCT_{eq}
13	1		3.74×10^{-05}	BTB-00531553	57,198,685	0.05	k_{20}
15a	2	14,243-14,272	$(7.29\times 10^{-06}, 7.24\times 10^{-06})$	ARS-BFGL-NGS-114291	14,242,668	0.01	RCT
15b	1		1.08×10^{-05}	ARS-BFGL-NGS-68607	55,488,319	0.01	RCT
19	2	2,094-2,271	$(3.47\times 10^{-05}, 2.96\times 10^{-05})$	Hapmap39832-BTA-46468	2,093,500	0.32	a_{30}
20	1		3.72×10^{-05}	ARS-BFGL-NGS-1751	17,412,441	0.17	Prot%
23a	1		2.85×10^{-05}	Hapmap38418-BTA-57213	8,819,178	0.11	k_{SR}
23b	1		2.59×10^{-06}	ARS-BFGL-NGS-99929	10,631,079	0.07	k_{SR}
26	1		2.5×10^{-05}	ARS-BFGL-NGS-23064	20,365,711	0.47	k_{CF}
28a	1		2.71×10^{-05}	ARS-BFGL-NGS-115508	33,729,338	0.11	t_{max}
28b	1		3.09×10^{-05}	Hapmap48306-BTA-36540	38,449,335	0.13	Prot%

BTA= *Bos taurus* autosome chromosome; #SNP (signals)= number of the single nucleotide polymorphisms significantly associated to the trait. In parenthesis 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)= The *P*-value of the highest significant SNP adjusted for genomic control and the range of the *P*-values when multiple

SNP were significantly associated to one 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 of the top SNP;

[1]RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Prot % = protein percent; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

In bold the trait with the highest *P*-value in each genomic region.

1

Figure 1. Modeling prolonged observations of curd firmness (CF) at time t (CFt), model parameters: RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

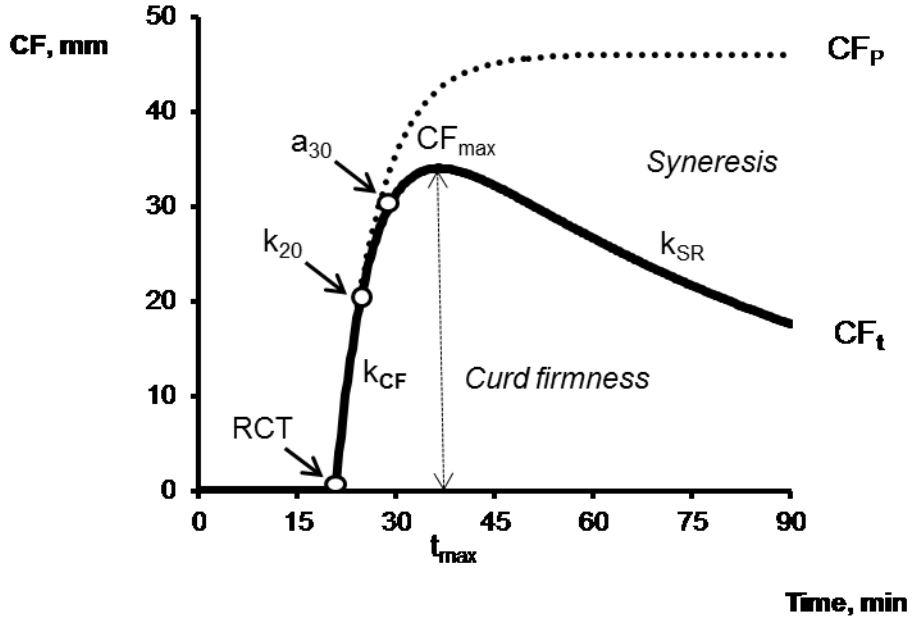


Figure 2. Manhattan plot of *P*-values for the genome wide association studies (GWAS) on *Bos taurus* autosome 6 (BTA6).

Description: Traits showed significant associations on BTA6 were RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Prot % = protein percent; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}).

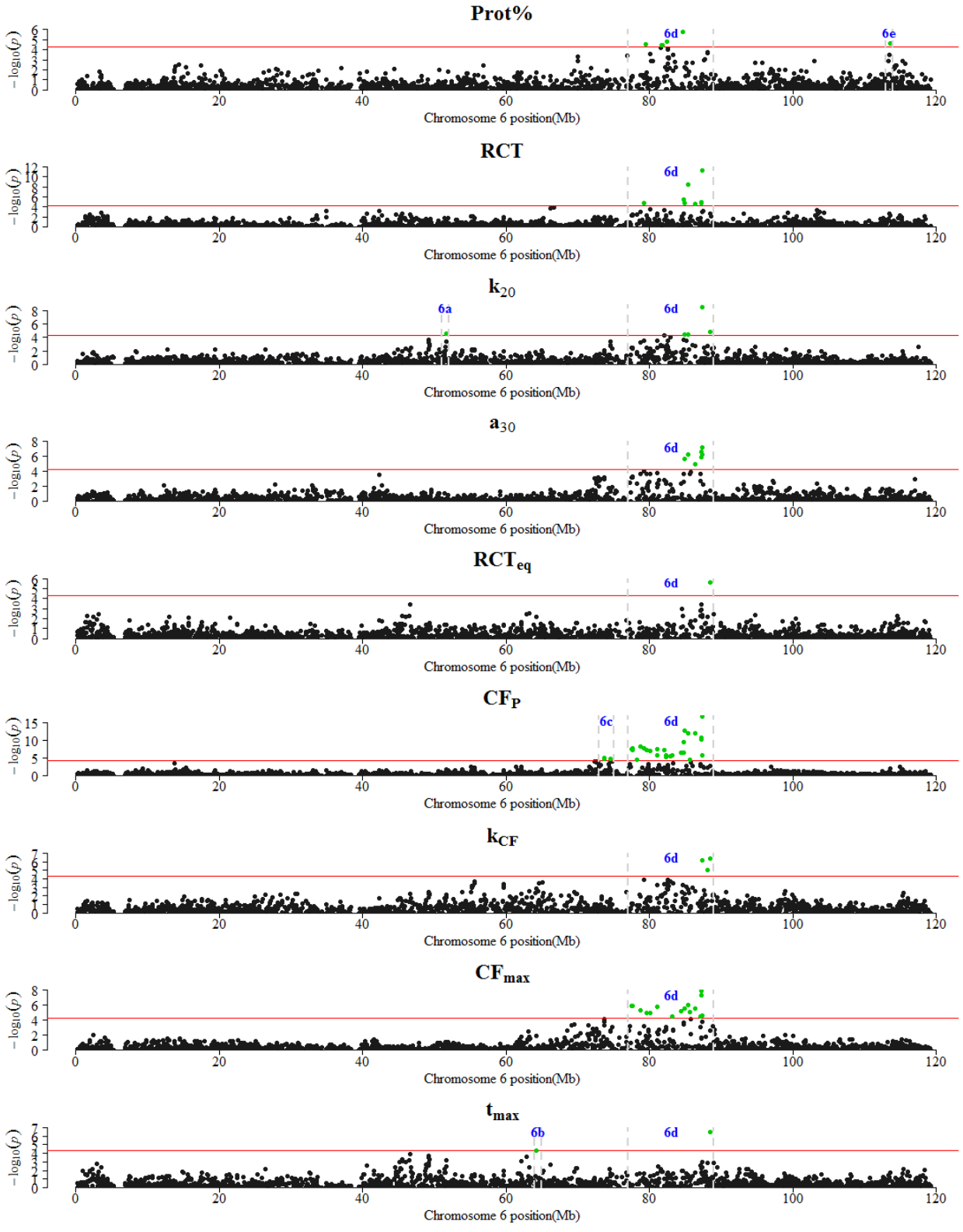
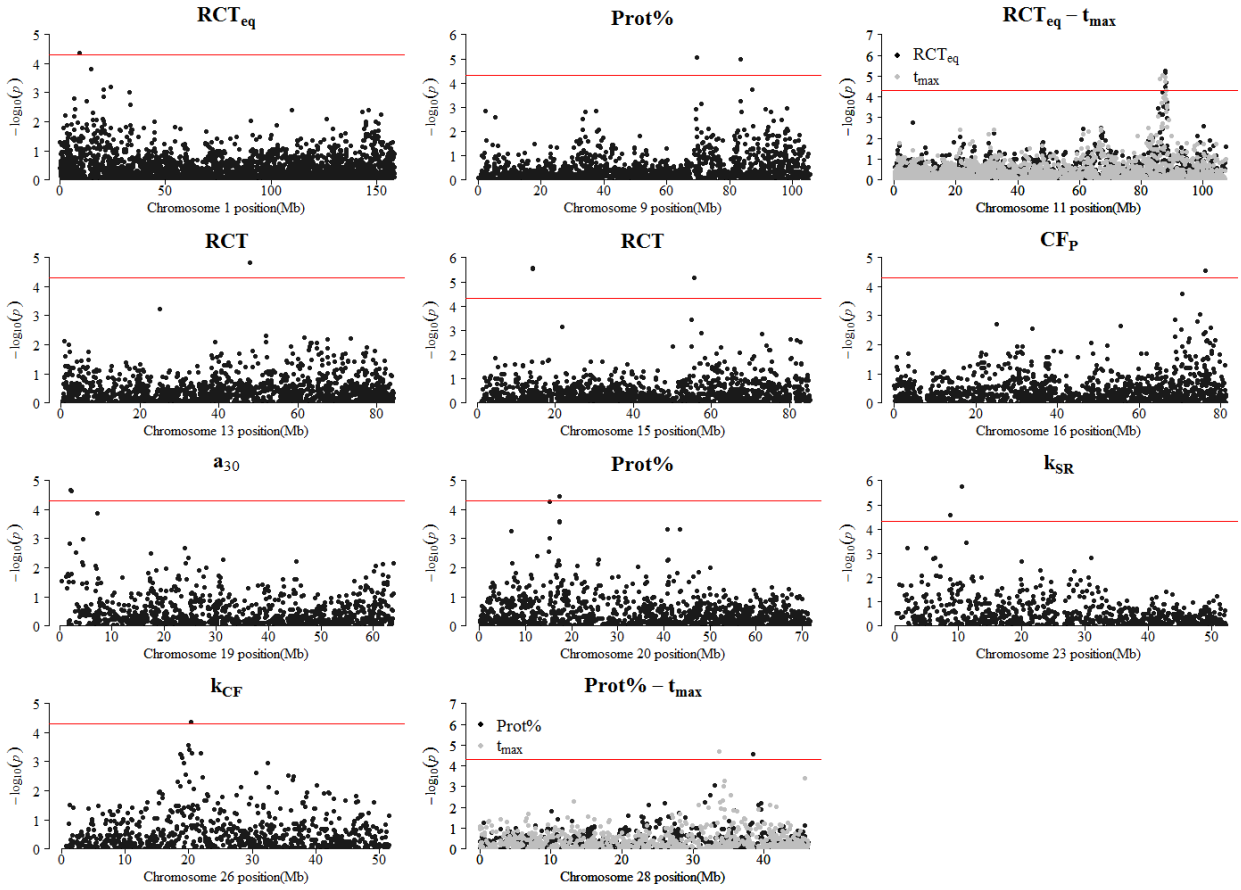

Prot%
6d
6e
RCT
6d
k20
6a
6d
a30
6d
RCTeq
6d
CFP
6c
6d
kCF
6d
CFmax
6d
tmax
6b
6d
−log10(p)
Chromosome 6 position(Mb)

Figure 3. Manhattan plot of *P*-values for the genome wide association studies (GWAS) on *Bos taurus* autosomes (BTA) 1, 9, 11, 13, 15, 16, 19, 20, 23, 26 and 28.

Description: RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Prot % = protein percent; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); t_{max} = time to maximum curd firmness (min).

The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}).



CHAPTER 2

Genome-wide association study for cheese yield and curd nutrient recovery in dairy cows

C. Dadousis,* S. Biffani,† C. Cipolat-Gotet,* E.L. Nicolazzi,‡ G. J. M. Rosa,[§] D. Gianola,[§#] A. Rossoni,[‖] E. Santus,[‖] G. Bittante,* and A. Cecchinato*

JOURNAL OF DAIRY SCIENCE, 2016 (*accepted*)

https://doi.org/10.3168/jds.2016-11586

*Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

† Istituto di Biologia e Biotecnologia Agraria (IBBA) - Consiglio Nazionale delle Ricerche (CNR), and

‡ Bioinformatics core, Parco Tecnologico Padano, Via Einstein- Loc. Cascina Codazza, 26900 Lodi, Italy

[§] Department of Animal Sciences, and

[#] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison Wisconsin, 53706 USA

[‖] Italian Brown Breeders Association, Loc. Ferlina 204, Bussolengo 37012, Italy

ABSTRACT

Cheese production and consumption are increasing in many countries worldwide. As a result, there is an increased interest in strategies for genetic selection of individuals for technological traits of milk related to cheese yield (CY) in dairy cattle breeding. However, little is known on the genetic background of a cow's cheese ability. Recently, a relatively large panel (1,264 cows) of different measures of individual cow CY and milk nutrient and energy recoveries in the cheese (REC) became available. Genetic analyses showed considerable variation for CY and for aptitude to retain high proportions of fat, protein and water in the coagulum. For the dairy industry these characteristics are of major economic importance. Nevertheless, use of this knowledge in dairy breeding is hampered by high costs, intense labor requirement, and lack of appropriate technology. However, in the era of genomics new possibilities are available for animal breeding and genetic improvement. For example, identification of genomic regions involved in cow's CY might provide potential for marker assisted selection. The objective of this study was to perform genome wide association studies (GWAS) on different CY and REC measures. Milk and DNA samples from 1,152 Italian Brown Swiss cows were used. Three CY traits expressing the weight (wt) of fresh curd (%CY_{CURD}), curd solids (%CY_{SOLIDS}), and curd moisture (%CY_{WATER}) as percentage of weight of milk processed, and 4 REC (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, and REC_{ENERGY} calculated as the % ratio between the nutrient in curd and the corresponding nutrient in processed milk) were analyzed. Animals were genotyped with the Illumina BovineSNP50 Bead Chip v.2. Single marker regressions were fitted using the GenABEL R package (GRAMMAR-GC). In total, 103 significant associations (88 SNP) were identified ($P < 5\times10^{-5}$) in 10 chromosomes (2, 6, 9, 11, 12, 14, 18, 19, 27, 28). For REC_{FAT} and $REC_{PROTEIN}$, high significance peaks were identified in Bos taurus autosomes (BTA) 6 and 11, respectively. Marker ARS-BFGL-NGS-104610 (~104.3 Mbp) was highly associated with $REC_{PROTEIN}$ ($P = 6.07\times10^{-36}$) and Hapmap52348-rs29024684 (~87.4 Mbp), closely located to the casein genes on BTA6, with REC_{FAT} ($P = 1.91\times10^{-15}$). Genomic regions identified may enhance marker assisted selection in bovine cheese breeding beyond the use of protein (casein) and fat contents, while new knowledge

will help to unravel the genomic background of a cow's ability for cheese production.

Key words: genome-wide association study, cheese yield, curd recovery, whey loss, dairy cattle.

INTRODUCTION

Milk and dairy products are important components of the human diet and the proportion of milk used for manufactured products, e.g. cheese, is steadily increasing in many countries worldwide (Food and Agriculture Organization of the,United Nations, 2015).

Milk characteristics, e.g. acidity and solid components (in particular casein and fat) are the cornerstone of cheese-making, and their role in this process has been previously investigated (Walstra et al., 2014). Moreover, milk coagulation properties (MCP) and curd firming modeling parameters (CF_t), together with the phenomenon of syneresis, are considered crucial technological features for cheese production (Bittante et al., 2012). However, milk components and MCP-CF_t traits can only be used as indicators of the cheese-making process. On the other hand, traits like the quantity of cheese obtained from a given amount of processed milk, or the recovery of milk components into the cheese are direct measures of the cheese-making aptitude of milk and so of great economic interest.

Although there is a considerable literature on cheese-making, knowledge is mostly based on bulk milk. The importance of the percentage of cheese yield (%CY) at the individual level (i.e., based on individual milk and not on bulk milk) has been pointed out by Othmane and colleagues (2002). Moreover, Banks (2007) discussed the significance of the milk constituents' recovery into the curd, as well as their loss in the whey for improved cheese quantity and quality. Previous studies have explored the potential of individual %CY using bovine milk, albeit based on relative small numbers of individuals (Hurtaud et al., 1995; Wedholm et al., 2006). Recently, a large dataset (n = 1,264) of different measures of individual cow %CYs and milk nutrient and energy recovery in the cheese (REC) became available, using a cheese-making model approach assessed at the lab level (Cipolat-Gotet et al., 2013). Further analysis has shown important genetic variation in individual %CY that does not solely depend on milk components but also heavily relies on the recovery of milk components in the curd (Bittante et al., 2013). Nevertheless, integration of the new knowledge into breeding programs is hampered by high costs, intensive labour requirements, and lack of appropriate technology.

At present, few potential alternatives have been suggested to overcome this problem, e.g., prediction of the aforementioned traits through infrared spectroscopy. Determination of milk components using spectral data is routinely used in the dairy industry (International Committee for Animal Recording (ICAR), 2012) and ongoing research is focused on the precision of the technology for predicting detailed milk components (Rutten et al., 2009), technological traits of milk such as milk coagulation (Cecchinato et al., 2009; Chessa et al., 2014) or different cheese measures (Ferragina et al., 2013; Bittante et al., 2014). On the other hand, genomic information offers a unique potential for a better understanding of the genetics underlying cheese-making properties. A first step towards this direction is, for example, the application of genome wide association studies (GWAS) (McCarthy et al., 2008; Visscher et al., 2012) where thousands of DNA markers, in the form of single nucleotide polymorphisms (SNP), are scanned throughout the entire genome (The Bovine Genome Sequencing and Analysis Consortium et al., 2009) linking the phenotype of interest to specific regions on the genome (Goddard and Hayes, 2009). In addition, genomic information can be used in marker-assisted or genomic selection breeding programs (Goddard and Hayes, 2009; de los Campos et al., 2013; Van Eenennaam et al., 2014).

Genomic regions associated with bovine milk quality traits have already been identified in a variety of studies, either using a small number of preselected DNA markers or a whole genome scan, and candidate genes have been detected. In a GWAS study using Holstein-Friesian cattle, different protein variants (especially casein variants) showed high peaks on Bos taurus autosomes (BTA) 6 and 11 (Schopen et al., 2011). Concerning MCP traits, milk coagulation has been associated to chromosomes 2, 18 and 24 in Finnish Ayrshire cattle (Tyrisevä et al., 2008), while for rennet coagulation time, β-casein (*CSN2*), β-lactoglobulin (*LGB*) and growth hormone 1 (*GH1*) have been identified as candidate genes (Bonfatti et al., 2010; Cecchinato et al., 2012; Cecchinato et al., 2015). Moreover, MCP and some cheese characteristics have also been associated to κ-casein (*CSN3*), leptin (*LEP*), and leptin receptor (*LEPR*) (Glantz et al., 2011). Recently, two GWAS studies identified other chromosomal regions associated to different MCP, CF_t and syneresis traits (Gregersen et al., 2015; Dadousis et al., 2016a). Nevertheless, these studies have identified genomic regions associated with indicators and not with direct measures of the cheese-making aptitude of milk.

The objective of our study was to conduct a GWAS analysis using individual cheese yield ($\%CY_{CURD}$, $\%CY_{SOLIDS}$ and $\%CY_{WATER}$) and milk nutrient and energy recovery into the curd measures (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, and REC_{ENERGY}) to shed light on the genetics underlying a cow's cheese-making ability. An Italian Brown Swiss dairy cattle sample genotyped with a 50k SNP chip and with all cheese-making traits measured through individual model-cheese manufacture was used.

MATERIALS AND METHODS

Field Data

Milk and blood samples were collected from 1,264 Italian Brown Swiss cows reared in 85 herds located in Trento Province (Italy). A full description of the sampling procedure can be found in Cecchinato et al. (2013). In brief, 15 cows per herd were individually sampled once (evening milking) and all samples were processed within 20h after collection. Information on cows and herds was supplied by the Breeders Association of Trento Province.

Definition of Phenotypes

The phenotypes were obtained through a model cheese-making procedure on 1,500mL of milk for each cow. The phenotype of interest (i.e. individual cheese yield) was split into seven components forming two groups of traits. The traits analyzed were: i) three %CY traits, expressing the weight (wt) of fresh curd ($\%CY_{CURD}$), of curd dry matter ($\%CY_{SOLIDS}$), and of water retained in the curd ($\%CY_{WATER}$) as percentage of wt of milk processed, and ii) four REC traits representing the proportion of nutrients and energy of the milk retained in the curd (REC_{SOLIDS}, REC_{FAT}, $REC_{PROTEIN}$ and REC_{ENERGY} calculated as the % ratio between the nutrient in curd and the corresponding nutrient in processed milk). The energy within the curd was calculated as the difference between energy in the milk and in the whey (NRC, 2001). A detailed description of the individual model cheese-making procedure used to obtain the phenotypes analyzed in this study as well as sources of phenotypic variation can be found in Cipolat-Gotet et al. (2013).

Genotyping

In total, 1,152 cows were genotyped using the Illumina BovineSNP50 v.2 BeadChip (Illumina Inc., San Diego, CA). Markers that fulfilled the following criteria were kept in the analysis: (1) call rate >95%, (2) minor allele frequency >0.005, and (3) no extreme deviation from Hardy-Weinberg proportions (HWP; $P >$ 0.001, Bonferroni corrected). After quality control, 1,011 animals and 37,568 SNP, distributed over 29 autosomes and the X-chromosome, were retained.

Genome-Wide Association Analyses

The association analysis was conducted using the GenABEL package in R (R Core Team, 2013; GenABEL project developers, 2013). The GRAMMAR-GC (Genome wide Association using Mixed Model and Regression - Genomic Control) approach was adopted, using the default function "gamma" (Amin et al., 2007; Svishcheva et al., 2012). For graphical representation of GWAS results the R package "qqman" was used (Turner, 2014). The GRAMMAR-GC is a three step approach. Firstly, an additive polygenic model is fitted using the genomic relationship matrix. At this step, our model was:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{a} + \mathbf{e}\,, \qquad (1)$$

where **y** is a vector containing the phenotypic records; β is a vector with the effects of days in milk of each cow (defined as classes of 30 days each), parity level of the cow (with classes 1, 2, 3, ≥4) and herd-date (85 total), all considered as fixed; and **X** is an incidence matrix that associates each observation to specific levels of factors in β. These non-genetic factors have already been studied in the same dataset (Bittante et al., 2013; Cipolat-Gotet et al., 2013). The random effects in the model consisted of the animal and the residual terms, which were assumed normally distributed as $\boldsymbol{a} \sim N(0, \mathbf{G}\sigma_g^2)$ and $\boldsymbol{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where **G** is the genomic relationship matrix, **I** is an identity matrix, and σ_g^2 and σ_e^2 are the additive genomic and residual variances, respectively. The **G** matrix was constructed within the GenABEL R package, where for a given pair of individuals i and j, the identical by state coefficients ($f_{i,j}$) is calculated as:

$$f_{i,j} = \frac{1}{N}\sum_k \frac{(x_{i,k} - p_k) \times (x_{j,k} - p_k)}{p_k \times (1 - p_k)}$$

where N is the number of markers used, $x_{i,k}$ is the genotype of the i[th] individual at the k[th] SNP (coded as 0, ½ and 1), p_k is the frequency of the "+" allele and k= 1, …, N.

Then, at the second step of GRAMMAR-GC, the residuals obtained in (1) are regressed on the SNP (single marker regression) to test for associations. In the last step, the Genomic Control (GC) approach corrects for conservativeness of the GRAMMAR procedure and gives estimates of the marker effects (Svishcheva et al., 2012). A threshold of *P*-value equal to 5×10^{-5} was adopted to declare significant associations (Burton et al., 2007). The variance explained by each SNP was calculated as $2pqa^2$, where p is the frequency of one allele, q=1-p is the frequency of the second allele and a is the estimated additive genetic effect. A scan for genes around 1Mbp upstream-downstream from the significant SNP was performed using the Ensembl Bos taurus UMD3.1 database (http://www.ensembl.org/index.html).

Model (1) was also used to estimate variance components and genomic heritability of the traits based on the genomic relationship matrix. Heritability was estimated as $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$.

RESULTS

Descriptive Statistics and Genetic Parameters

Descriptive statistics of the seven phenotypes analyzed in GWAS as well as milk yield and milk quality traits (fat and protein percentage) are presented in Table 1. The $\%CY_{CURD}$ had a mean value of 14.95% (meaning that from 1.5 liters of milk ~225g of fresh cheese was obtained, on average) ranging between 10.2% and 20.5%. On average, the quantity of water retained in the curd was slightly higher than the quantity of the curd solids (7.77% vs. 7.17%, respectively). The highest level of recovery from milk to cheese was observed for fat (mean of 89.76%), followed by protein (78.16%), and energy (67.15%). The REC_{SOLIDS} showed the

lowest mean value (51.80%) because of the loss of most of the lactose and minerals with whey. For the same dataset, the average milk yield was 24.95 kg/day. The milk fat and protein percentages had average values of 4.37% and 3.71%, respectively.

Additive genomic variance and heritability estimates are presented in Table 2. The recovery of the milk protein in cheese had the highest heritability (0.46). For the rest of the traits the h^2 estimates were 0.29 for $\%CY_{SOLIDS}$ and REC_{SOLIDS} and 0.37 for $\%CY_{CURD}$, while for $\%CY_{WATER}$ and REC_{ENERGY} they were estimated at ~0.25. The lowest h^2 was found for REC_{FAT} (0.14).

Results of the Genome-Wide Associations

In total, 103 significant associations (88 SNP) were identified ($P < 5\times10^{-5}$) on 10 chromosomes (BTA 2, 6, 9, 11, 12, 14, 18, 19, 27, 28) (Table 3, Supplementary Table S1). Sharp peaks were identified on BTA6 and BTA11 for the REC_{FAT} and the $REC_{PROTEIN}$, respectively. The most frequently identified chromosome was BTA6, representing 55 out of 103 total associations (53%). This chromosome together with BTA11 accounted for 92% of the total significant associations. With the exception of BTA6, all the remaining hits were 'one trait – one chromosome' associations. The quantile-quantile (Q-Q) plots (scatter plot of the observed vs. expected values of test statistics derived from the GWAS analyses) were also inspected. The Q-Q plots showed extreme departures on the tail of the distributions for all traits, except for $\%CY_{WATER}$, providing extra evidence for true associations of the GWAS analyses (Supplementary Figure S1).

On BTA6 all traits but $\%CY_{WATER}$ and $REC_{PROTEIN}$ showed signals (Table 3, Figure 1). Three regions can be distinguished on BTA6. In the first region (6a) at ~73.7 – 74.6 Mbp the REC_{FAT} was linked. The second region (6b; at ~77.5 – 88.4 Mbp) was the broader one and all of the five traits showed signals, with REC_{FAT} showing the maximum number of associations among those traits. Finally, $\%CY_{CURD}$ was associated to one marker at ~102.9 Mbp (region 6c). The highest significant marker on BTA6 was Hapmap52348-rs29024684 (87,396,306bp). This SNP was the highest significant identified for the REC_{SOLIDS}, REC_{FAT} and REC_{ENERGY} (P-value of 3.93×10^{-7}, 1.91×10^{-15}

and 8.03×10^{-7}, for REC_{SOLIDS}, REC_{FAT} and REC_{ENERGY}, respectively).

A total of 38 SNP were significant on BTA11 at an area spanned between 91.6 – 106.7 Mbp. All of the SNP were associated to $REC_{PROTEIN}$ (Table 3, Figure 2). The area could further be partitioned into three sub-regions: between 91.6 – 94.8 (11a), 96.2 – 98.5 (11b) and 100.5 – 106.7 (11c) Mbp. In the last area 25 SNP were significant. Moreover, the most significant SNP was identified within this area, namely marker ARS-BFGL-NGS-104610 (104,293,559bp) with a *P*-value = $6.07x10^{-36}$.

Apart from BTA6 and BTA11 signals in eight more chromosomes were detected for $\%CY_{SOLIDS}$, $\%CY_{WATER}$ and REC_{ENERGY}, albeit at relatively mild strength (Table 3, Figure 3). More precisely, $\%CY_{SOLIDS}$ was linked to BTA2 (~128.8 Mbp), BTA14 (~26 Mbp), BTA19 (~1.8 Mbp) and BTA27 (~42.1 Mbp). Three markers were significantly associated to the $\%CY_{WATER}$ on BTA12 (~85.3 Mbp), BTA18 (~65.3 Mbp) and BTA28 (~38.5 Mbp). Finally, on BTA9 (~91.1 Mbp) a signal for REC_{ENERGY} was detected.

Comparing the two sets of traits (CY vs. REC), the three cheese yield traits were associated to 13 SNP in eight chromosomes (BTA 2, 6, 12, 14, 18, 19, 27 and 28) while the four nutrient and energy recoveries of milk into the cheese curd showed associations (76 SNP) on BTA6, 9 and 11. The only genomic region in common between the two groups of traits was the area 6b on BTA6, and more precisely the region ~82.5 – 88 Mbp. In this area, three significant SNP were in common between $\%CY_{CURD}$, $\%CY_{SOLIDS}$, REC_{SOLIDS} and REC_{ENERGY} (namely, Hapmap26275-BTC-043486, Hapmap50464-BTA-77021 and Hapmap53172-rs29012675), all of which located in the area around 82Mbp.

Fitting Hapmap52348-rs29024684 (~87.4 Mbp) on BTA6 and ARS-BFGL-NGS-104610 (~104.3 Mbp) on BTA11

Based on the GWAS results and the high peaks identified on BTA6 and BTA11 a further

GWAS analysis (conditional GWAS) was followed up by fitting simultaneously as covariates in model (1) the two highest significant SNP, to account for the effect of the casein variants and their quantity in milk. In this case, a total of 43 associations (26 SNP) were detected in 8 chromosomes (Table 4, Supplementary Table S2). The number of significant SNP for REC_{FAT} and $REC_{PROTEIN}$ drastically decreased (6 and 5, respectively). Moreover, $REC_{PROTEIN}$ showed new associations on BTA6 (at ~85.6 and 88 Mbp) and BTA22 (~38.8 Mbp), while associations for REC_{FAT} were still detected on BTA6 (at ~82.5, 87.2 and 88.5 Mbp). In general, results changed in two directions, namely some regions that were significant in the first analysis did not show significance in the second analysis, and vice-versa. In addition, there were changes in chromosomes as well as positions within chromosomes where significant SNP were located. For example, $\%CY_{CURD}$ was now associated to two more regions on BTA19 (ARS-BFGL-NGS-102974-1,822,133 Mbp; ARS-BFGL-NGS-24753-3,024,589 Mbp) that were not previously detected. On BTA6 the region 6a and part of the 6b region (up to ~82 Mbp) were not confirmed. Three sub-regions could be identified in the 6b area at ~82 Mbp (6b_1), ~85-86 Mbp (6b_2) and 87.2-88.6 Mbp (6b_3). The region 6c remained significant. However, an extra signal was detected at ~114 Mbp associated to the $\%CY_{SOLIDS}$. Interestingly enough, all the minor alleles of the significant markers on BTA6 had now positive effects, for all traits (Supplementary Table S2). On BTA11 only a signal within the 11c area (~103.5 Mbp) was detected, but in this case $REC_{PROTEIN}$ was associated with only 2 SNP. Also in this case, the estimated effects of the minor alleles of the significant SNP on BTA11 were now positive. Signals on BTA2, 14 and 18 were not identified while the significant associations remained on BTA27 and 28. In the conditional analysis most of the signals were mild with the exception of the regions 6b_1 and 11c.

DISCUSSION

Genome-Wide Associations for the Cheese Yield Traits

Cheese yield is the outcome of a complex process on which a variety of interrelated factors, derived from different disciplines (e.g. microbiology, physical chemistry, engineering process, etc) are involved, such as: i) the quantity (e.g. caseins and fat) as well as the quality (e.g. the fraction of the caseins to the total milk proteins, and the dimension of fat globules) of milk components; ii) other milk characteristics (e.g. milk acidity, minerals and microbial flora); iii) milk pre-treatments (e.g. milk natural creaming, heat treatment, etc.); and iv) cheese-making conditions (e.g. coagulation temperature, type and concentration of rennet, curd cutting, curd cooking, and pressing). To understand this complicated progression a plethora of studies have been carried out, and a well-documented description of the cheese process is available (Banks, 2007; Law and Tamime, 2011; Walstra et al., 2014), although our knowledge is yet incomplete. Moreover, this knowledge has been built based on bulk milk. From a breeding perspective, however, information on the individual level of the ability of a cow to produce cheese is necessary. Previous analysis indicated that there is considerable variation on the traits of interest, and breeding could help for further improvement (Bittante et al., 2013). The restriction, though, for a wide scale phenotyping on individual CY is the lack of an appropriate technology combined with high costs. Identification of chromosomic regions associated to individual CY offers a possible alternative through the application of genomic breeding programs.

Genomic heritability estimates varied between 0.14 to 0.46. These values are in the same range as those reported in Bittante et al. (2013) from an infinitesimal model. Slight differences might be attributed to the different set up of statistical model used (e.g. herd was fixed vs. random effect, frequentist vs. Bayesian analysis) and the different number of individuals included. Moreover, here we made use of the genomic relationship matrix while in Bittante et al. (2013) the analysis was carried out using pedigree information.

Previous GWAS studies have linked various milk components and milk technological

characteristics related to cheese yield to a variety of chromosomic regions (Schopen et al., 2011; Gregersen et al., 2015; Dadousis et al., 2016a). In the present study, for the first time, seven traits derived from direct cheese measures, representing an individual cow's cheese production ability and obtained from a model cheese-making process, were analyzed. All the traits showed significant associations and were linked to at least one chromosome.

BTA6

Sharp peaks were detected for REC_{FAT} on BTA6 in a broad area between ~77.5-88.5 Mbp. In this area the casein cluster is mapped (~87 Mbp). All of the traits analyzed but $\%CY_{WATER}$ and $REC_{PROTEIN}$ showed associations in this region (Table 3, Figure 3) with the highest peak spotted at ~87 Mbp (marker Hapmap52348-rs29024684; 87,396,306 bp). This marker had a negative effect of the minor allele for all the traits, more precisely -0.84, -1.34 and -0.90 for REC_{SOLIDS}, REC_{FAT} and REC_{ENERGY}, respectively. For the same traits, the proportion of phenotypic variance explained by this marker was 3.31% and 3.92% for REC_{SOLIDS} and REC_{ENERGY} while it increased to 8.45% for REC_{FAT}. Concerning the percentage of the additive genetic variance explained by this SNP it was found to be 11.5%, 58.8% and 16.3%, for REC_{SOLIDS}, REC_{FAT} and REC_{ENERGY}, respectively (Supplementary Table S1). It is interesting to note that the area 6b has recently been reported and thoroughly described in a recent GWAS study (Dadousis et al., 2016a). In that case, a trait defined as the "asymptotic value of curd firmness in the absence of syneresis" showed strong signals exactly in the same region (~77.5-88.5 Mbp. Moreover, marker Hapmap52348-rs29024684 was highly significant for a variety of traits describing milk coagulation and curd firmness (*P*-values ranged between 7.46×10^{-07} and 1.62×10^{-17}). The present analysis has confirmed the importance of the casein genes and especially the *CSN3* casein gene for an individual cow's cheese yield, since the peak is located only18kbp upstream of *CSN3*. In addition, in this area is also present the marker Hapmap28023-BTC-060518 (87.20 Mbp) that was significantly related to REC_{FAT} (*P*-value = 3.36×10^{-07}). The SNP is mapped within the histatherin

gene (*HSTN*) which is located between the *CSN2* and the *CSN1S2*. The same SNP was also reported in Dadousis et al. (2016a). Even after conditioning on the peak on BTA6 signals still remained at ~82.71, ~85.1, ~87.22 and ~88.1Mbp giving evidence for potential new quantitative trait loci (QTL) (Table 4, Supplementary Table S2).

The region around 88Mbp has also been linked to different measures of Holstein cow milk protein composition, including the casein variants *CSN1S1*, *CSN1S2*, *CSN2* and *CSN3* as well as α-lactalbumin (*α-LA*), *LGB*, the casein index and casein percentage (Schopen et al., 2011). Moreover, a similar peak at 87.2-87.4 Mbp has been recently reported in a GWAS study analyzing rheological traits of milk related to milk coagulation in Swedish Red cows (Gregersen et al., 2015). Other regions on BTA6 that have been detected in previous studies on association with milk quality and coagulation traits did not appear significant in our study for the direct measures of cheese-making traits, e.g. the area ~44Mbp in Schopen et al. (2011), the genomic region ~50-51Mbp (Gregersen et al., 2015; Dadousis et al., 2016a), the areas ~60Mbp (Gregersen et al., 2015) and ~64Mbp (Dadousis et al., 2016a). Nevertheless, the region 6d detected in the conditional analysis is in close proximity to the ~113Mbp that was reported in (Dadousis et al., 2016a) associated to milk protein (%).

By using direct cheese traits obtained through a cheese-making process it has been shown that the importance of this genomic region is related, apart from codification of protein genetic variants, with the recovery of the milk fat into the curd, and not to the recovery of protein. This was, to an extent, expected, since in healthy cows all casein variants are essential for the structure and activity of milk protein micelles (Holt et al., 2013) and participate to the formation of coagula (Caroli et al., 2009) and curd syneresis (Pearse and Mackinlay, 1989; Everard et al., 2011). However, the efficiency of capturing milk fat globules depends on the rapidity of clotting and the strength of the curd that in turn depend on casein genetic variants (Alipanah and Kalashnikova, 2007).

BTA11

Chromosome eleven was linked only to $REC_{PROTEIN}$, with all association hits being detected on the tail of BTA11 (Table 3). On this chromosome the peak was detected at 104,293,559bp (marker ARS-BFGL-NGS-104610; *P*-value=6.07×10^{-36}) while a strong signal at 103.35 Mbp remained in the conditional analysis. The effect of the minor allele of this marker was negative for the $REC_{PROTEIN}$ (-1.30). The phenotypic variance explained by this marker was 22.87%, while half of the additive genetic variance can be attributed to this SNP. The significance of this region confirms the findings of Schopen et al. (2011), in which significant associations were identified for the casein variants (α, β and κ) as well as for β-lactoglobulin content and casein index, although in that case high peaks appeared at ~107Mbp. It should be noted that $REC_{PROTEIN}$ is phenotypically very similar to the casein index. However, since part of the caseins are lost in the whey, $REC_{PROTEIN}$ is able to capture more variation compared to the casein index. The region around 86-88Mbp has been recently associated to CF_t traits (Dadousis et al., 2016a), but it was not confirmed for direct cheese-making traits in our study. Nevertheless, this batch is relatively close to marker BTA-93319-no-rs (91,639,283bp), which was significant in the present GWAS analysis ($P = 8.8\times10^{-6}$). However, this signal was diluted after conditioning on the BTA11 peak.

A closer look into the region ±1Mbp around the marker ARS-BFGL-NGS-104610 was performed. This SNP is located ~2.5kbp downstream of *SURF6* (surfeit locus protein 6). In this small area the following protein coding genes are located: i) downstream the marker: *AGPAT2* (1-acyl-sn-glycerol-3-phosphate acyltransferase beta), *FAM69B* (family with sequence similarity 69, member B), *ABO* (histo-blood group *ABO* system transferase), and ii) upstream the marker: *MED22* (mediator of RNA polymerase II transcription subunit 22), *RPFL7A* (60S ribosomal protein L7a), *SURF1* (surfeit locus protein 1), *SURF2* (surfeit locus protein 2), *STKLD1* (serine/threonine kinase-like domain containing 1), *REXO4* (RNA exonuclease 4). From those protein coding genes, *AGPAT2* is involved in fat digestion and absorption as well as in fatty acid metabolic pathways. The gene

RPFL7A, among a variety of functions, is also part of biosystem pathways regulating the metabolism of proteins. Moreover, although further apart on the chromosome, yet within the region spanning ±1Mbp around the significant marker, the progestagen-associated endometrial protein gene (*PAEP*, also known as *LGB*) can be found.

Signals on Other Chromosomes

Concerning the rest of the chromosomes, one SNP-one trait associations were observed in all the cases. Moreover, the majority of the *P*-values was around the threshold. The signals on BTA2, BTA14 and BTA18 were not retained after conditioning the peaks on BTA6 and 11. Moreover, on BTA14 the significant SNP detected is far apart from *DGAT1* (diacylglycerol O-acyltransferase 1). This should not be surprising, mainly for two reasons. Firstly, *DGAT1* has been associated to milk fat content in dairy cattle. In our study, however, the recovery of the milk fat into the curd was the focal point. The two traits are entirely different, both phenotypically and genetically, with almost zero phenotypic and genetic correlations (Bittante et al., 2013). At the beginning of the coagulation process, the fat of the milk is blocked by the caseins. This has been confirmed in our analysis by associating, with high peaks, REC_{FAT} to a region on BTA6 where the casein genes are located (see also results of (Schopen et al., 2011) on GWAS with milk protein variants). Moreover, in the Italian Brown Swiss population *DGAT1* is not segregating. Albeit at a mild strength, the same signal on BTA28 has been previously linked to milk protein percentage (Dadousis et al., 2016a).

Partitioning the Cheese Yield into Different Components

As a side note, our study highlighted the importance of partitioning a complex trait into different components for a better understanding of its genomic background. More precisely, the percentage of cheese produced by a given amount of milk was split into seven different traits. For example, phenotypically $\%CY_{CURD} = \%CY_{SOLIDS} + \%CY_{WATER}$. However, the genomic signals

of $\%CY_{CURD} \neq \%CY_{SOLIDS} + \%CY_{WATER}$. On the contrary, seven out of eight genomic regions identified for $\%CY_{SOLIDS}$ and $\%CY_{WATER}$ were not present in $\%CY_{CURD}$. Moreover, the marker Hapmap52348, located close to *CSN3*, was detected only from the recoveries of fat, solids and energy while the significance of BTA11 was depicted only with the protein recovery.

CONCLUSION

Different measures of bovine cheese ability were associated to ten chromosomic regions. Five out of the seven cheese traits used in GWAS analyses showed high peaks on BTA6 (~87.4 Mbp) in an area close to kappa casein, confirming the importance of this casein variant in cheese making. Moreover, high peaks on BTA11 (~104.3 Mbp) were linked to milk protein recovery into the curd. A number of other chromosomic regions have also been significantly associated to the cheese traits analyzed, albeit in the absence of high peaks. Highly significant markers identified, especially on BTA6 and BTA11, could further be used for genomic prediction purposes. The road for fine mapping of genomic regions associated to a cow's cheese production potential is ahead. Future research based on gene enrichment analysis might complement the GWAS results and help to deepen the understanding of the biological pathways related to the CY traits. Partition of a complex process, as is the cheese yield, into different components has been shown as a useful tool for connecting phenomics to their genomic counterpart.

ACKNOWLEDGMENTS

The authors wish to thank the Trento Province (Italy), the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy) and the Superbrown Consortium of Bolzano and Trento for financial and technical support. C. Dadousis benefitted from financial support of the CARIPARO (Cassa di Risparmio di Padova e Rovigo) Foundation (Padua, Italy).

Table 1. Descriptive statistics of individual percentage cheese yield (%CY; weight of fresh curd, curd solids, and curd water as percentage of weight of milk processed), milk nutrient and energy recovery (REC; protein, fat, solids, and energy of the curd as percentage of the protein, fat, solids, and energy of the milk processed) obtained from 1.5L of milk processed, and milk yield and quality traits (n=1,011).

Trait	Mean	SD	P1	P99	CV, %
Cheese yield, %					
$\%CY_{CURD}$	14.95	1.84	11.00	19.41	12.4
$\%CY_{SOLIDS}$	7.17	0.91	5.37	9.68	12.9
$\%CY_{WATER}$	7.77	1.26	5.04	11.11	16.3
Nutrient Recovery, %					
REC_{SOLIDS}	51.80	3.50	43.85	60.27	6.8
REC_{FAT}	89.76	3.60	78.44	95.90	4.0
$REC_{PROTEIN}$	78.16	2.44	72.41	83.44	3.1
REC_{ENERGY}	67.15	3.28	58.92	75.07	4.9
Milk traits					
Milk yield, kg/d	24.95	7.84	9.30	45.89	31.4
Fat, %	4.37	0.89	2.48	7.40	20.4
Protein, %	3.71	0.41	2.93	4.68	10.9

P1: 1st percentile; P99; 99th percentile; CV: coefficient of variation

Table 2 Additive genomic variance (σ_g^2) and heritability estimates (h^2) of individual percentage cheese yield (%CY; weight of fresh curd, curd solids, and curd water as percentage of weight of milk processed) and milk nutrient and energy recovery (REC; protein, fat, solids, and energy of the curd as percentage of the protein, fat, solids, and energy of the milk processed).

Trait	σ_g^2	h^2
Cheese yield, %		
$\%CY_{CURD}$	0.662	0.366
$\%CY_{SOLIDS}$	0.151	0.290
$\%CY_{WATER}$	0.205	0.262
Nutrient recovery, %		
REC_{SOLIDS}	2.222	0.287
REC_{FAT}	1.109	0.144
$REC_{PROTEIN}$	1.666	0.458
REC_{ENERGY}	1.810	0.241

Table 3. Summary results of the genome wide association analyses

BTA[1]	No. of SNP (signals)	Interval, Mbp	*P*-value (range)	Top SNP	Top SNP location, bp	Top SNP MAF	Trait[2]
0[3]	1	-	3.09×10^{-06}	BTA-76907-no-rs	0	0.26	REC_{FAT}
0	1	-	2.14×10^{-05}	ARS-BFGL-NGS-110734	0	0.30	REC_{FAT}
2	1	-	3.42×10^{-05}	Hapmap60596-rs29017365	128,812,635	0.34	$\%CY_{SOLIDS}$
6a	2 (2)	73.734-74.607	$(3.88 \times 10^{-05}, 2.09 \times 10^{-05})$	Hapmap60182-rs29025531	74,606,760	0.19	REC_{FAT}
6b	37 (54)	77.521-88.442	$(4.85 \times 10^{-05}, 1.91 \times 10^{-15})$	Hapmap52348-rs29024684	87,396,306	0.24	**REC_{FAT}**, REC_{ENERGY}, $\%CY_{CURD}$, $\%CY_{SOLIDS}$, REC_{SOLIDS}
6c	1	-	9.59×10^{-06}	Hapmap23975-BTC-043815	102,937,469	0.20	$\%CY_{CURD}$
9	1	-	4.74×10^{-05}	BTB-00403185	91,058,994	0.34	REC_{ENERGY}
11a	8 (8)	91.639-94.773	$(2.24 \times 10^{-05}, 1.26 \times 10^{-08})$	ARS-BFGL-NGS-42578	93,241,685	0.25	$REC_{PROTEIN}$
11b	5 (5)	96.230-98.511	$(3.31 \times 10^{-05}, 1.33 \times 10^{-07})$	ARS-BFGL-NGS-21607	97,059,246	0.08	$REC_{PROTEIN}$
11c	25 (25)	100.520-106.741	$(3.74 \times 10^{-05}, 6.07 \times 10^{-36})$	ARS-BFGL-NGS-104610	104,293,559	0.45	$REC_{PROTEIN}$
12	1	-	2.42×10^{-05}	BTB-00507211	85,272,488	0.13	$\%CY_{WATER}$
14	1	-	4.58×10^{-05}	Hapmap25446-BTC-054694	26,003,598	0.46	$\%CY_{SOLIDS}$
18	1	-	4.32×10^{-05}	Hapmap51570-BTA-17962	65,269,893	0.25	$\%CY_{WATER}$
19	1	-	3.05×10^{-05}	ARS-BFGL-NGS-102974	1,822,133	0.34	$\%CY_{SOLIDS}$
27	1	-	1.53×10^{-05}	ARS-BFGL-NGS-87845	42,118,037	0.03	$\%CY_{SOLIDS}$
28	1	-	1.66×10^{-05}	Hapmap48306-BTA-36540	38,449,335	0.13	$\%CY_{WATER}$

[1]BTA= *Bos taurus* autosome chromosome; #SNP (signals)= number of the single nucleotide polymorphisms significantly associated to the trait. In parenthesis the total number of significant signals per each genomic region; Interval: The region on the chromosome spanned among the significant SNP (in mega base pairs); *P*-value (range)= The *P*-value of the highest significant SNP adjusted for genomic control and the range of the *P*-values when multiple SNP were significantly associated to one trait; Top SNP location (bp)= position of the highest significant SNP on the chromosome in base pairs on UMD3.1 (http://www.ensembl.org/index.html); Top SNP MAF= minor allele frequency of the top SNP.

[2]%CY: weight of fresh curd, curd solids, and curd water as percentage of weight of milk processed; REC: protein, fat, solids, and energy of the curd as percentage of the protein, fat, solids, and energy of the milk processed. The trait with the highest *P*-value in each genomic region is bolded.

[3]Undefined chromosome and position on the genome

Table 4. Summary results of the genome wide association analyses after fixing the marker Hapmap52348-rs29024684 at 87,396,306bp on BTA6[1] and the marker ARS-BFGL-NGS-104610 at 104,293,559 on BTA11

BTA	No. of SNP (signals)	Interval, Mbp	*P*-value (range)	Top SNP	Top SNP location, bp	Top SNP MAF	Trait[2]
6b_1	5 (16)	82,410-82,707	$(4.28 \times 10^{-05}, 2.3 \times 10^{-08})$	Hapmap53172-rs29012675	82,706,745	0.05	**%CY$_{CURD}$**, %CY$_{SOLIDS}$, REC$_{FAT}$, REC$_{ENERGY}$, REC$_{SOLIDS}$
6b_2	2 (2)	85,178- 85,955	$(4.77 \times 10^{-05}, 4.71 \times 10^{-05})$	Hapmap60224-rs29001782	85,178,107	0.16	REC$_{PROTEIN}$, **REC$_{ENERGY}$**
6b_3	7 (10)	87,223- 88,592	$(4.04 \times 10^{-05}, 5.21 \times 10^{-06})$	ARS-BFGL-NGS-112872	88,069,548	0.14	%CY$_{CURD}$, %CY$_{SOLIDS}$, REC$_{FAT}$, **REC$_{PROTEIN}$**, REC$_{ENERGY}$, REC$_{SOLIDS}$
6c	1	-	1.33×10^{-05}	Hapmap23975-BTC-043815	102,937,469	0.20	%CY$_{CURD}$
6d	1	-	4.29×10^{-05}	BTB-02092741	114,223,059	0.01	%CY$_{SOLIDS}$
9	1	-	4.64×10^{-05}	BTB-01249963	37,442,535	0.37	%CY$_{WATER}$
11c	2 (2)	103,352-103,743	$(1.26 \times 10^{-05}, 7.92 \times 10^{-12})$	ARS-BFGL-NGS-26919	103,352,220	0.04	REC$_{PROTEIN}$
12	1	-	2.51×10^{-05}	BTB-00507211	85,272,488	0.13	%CY$_{WATER}$
19a	1	-	3.59×10^{-05}	ARS-BFGL-NGS-43028	1,706,305	0.35	%CY$_{SOLIDS}$
19b	1 (3)	-	$(3.04 \times 10^{-05}, 7.54 \times 10^{-06})$	ARS-BFGL-NGS-102974	1,822,133	0.34	%CY$_{CURD}$, **%CY$_{SOLIDS}$**, REC$_{SOLIDS}$
19c	1 (2)	-	$(3.5 \times 10^{-05}, 2.91 \times 10^{-05})$	ARS-BFGL-NGS-24753	3,024,589	0.37	%CY$_{CURD}$, **%CY$_{SOLIDS}$**
22	1	-	1.92×10^{-05}	ARS-BFGL-NGS-4481	38,782,041	0.01	REC$_{PROTEIN}$

27	1	-	2.79×10^{-05}	ARS-BFGL-NGS-87845	42,118,037	0.03	$\%CY_{SOLIDS}$
28	1	-	2.63×10^{-05}	Hapmap48306-BTA-36540	38,449,335	0.13	$\%CY_{WATER}$

[1]BTA= *Bos taurus* autosome chromosome; #SNP (signals)= number of the single nucleotide polymorphisms significantly associated to the trait. In parenthesis the total number of significant signals per each genomic region; Interval: The region on the chromosome spanned among the significant SNP (in mega base pairs); *P*-value (range)= The *P*-value of the highest significant SNP adjusted for genomic control and the range of the *P*-values when multiple SNP were significantly associated to one trait; Top SNP location (bp)= position of the highest significant SNP on the chromosome in base pairs on UMD3.1 (http://www.ensembl.org/index.html); Top SNP MAF= minor allele frequency of the top SNP.

[2]%CY: weight of fresh curd, curd solids, and curd water as percentage of weight of milk processed; REC: protein, fat, solids, and energy of the curd as percentage of the protein, fat, solids, and energy of the milk processed. The trait with the highest *P*-value in each genomic region is bolded.

Figure 1. Manhattan plot of –log(*P*-values) for the genome wide association studies (GWAS) on *Bos taurus* autosome 6 (BTA6).

Description: Traits showed significant associations on BTA6 were: individual percentage cheese yield traits (%CY; weight of fresh curd, and curd solids, as percentage of weight of milk processed) and milk nutrient and energy recovery traits (REC; fat, solids, and energy of the curd as percentage of the fat, solids, and energy of the milk processed).

The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}).

The highest significant marker on BTA6 is also presented.

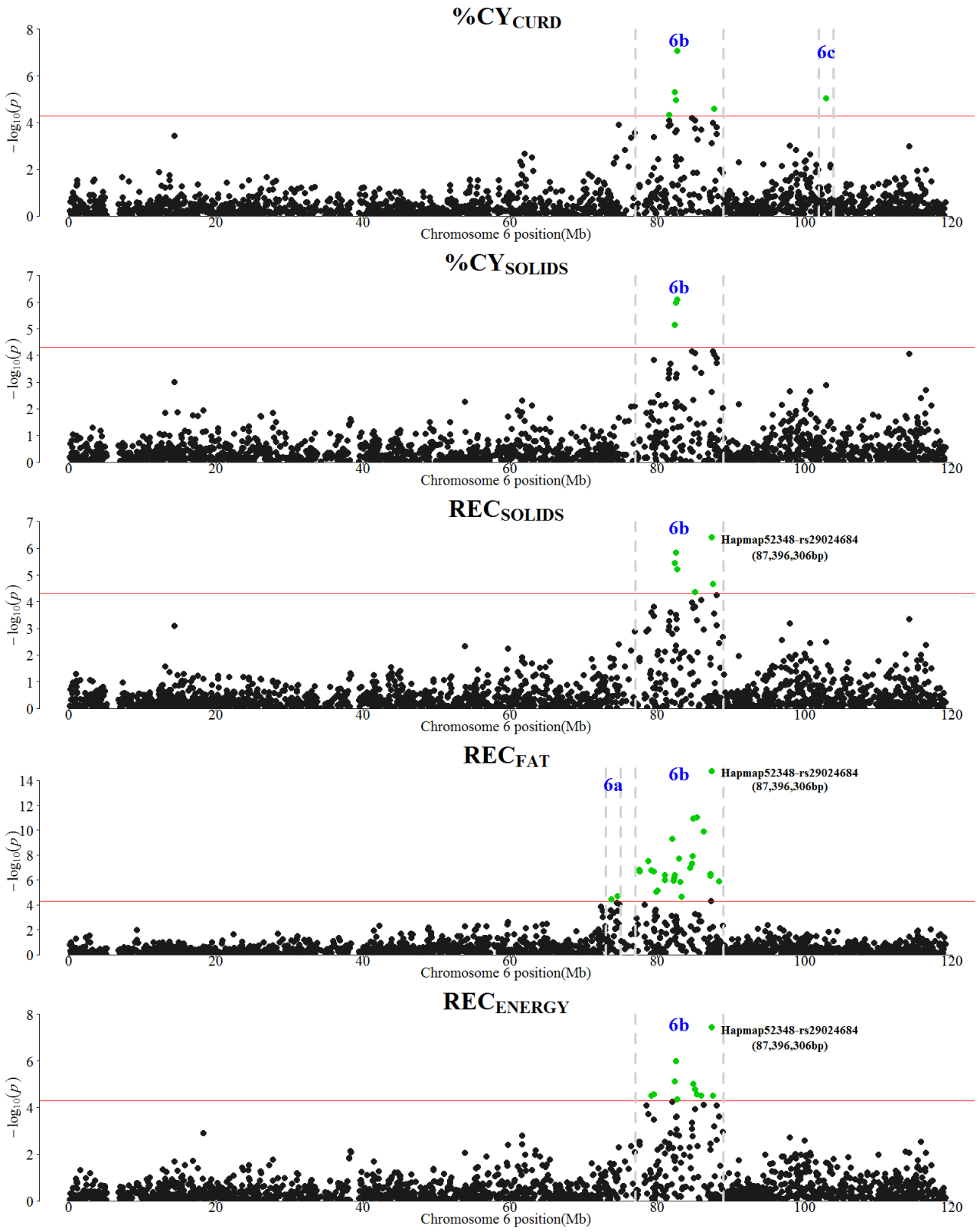

%CY_CURD
6b
6c
−log10(p)
Chromosome 6 position(Mb)
%CY_SOLIDS
6b
REC_SOLIDS
6b
Hapmap52348-rs29024684
(87,396,306bp)
REC_FAT
6a
6b
Hapmap52348-rs29024684
(87,396,306bp)
REC_ENERGY
6b
Hapmap52348-rs29024684
(87,396,306bp)

Figure 2. . Manhattan plot of –log(*P*-values) for the genome wide association studies (GWAS) on *Bos taurus* autosome 11 (BTA11).

Description: $REC_{PROTEIN}$= Protein of the curd as percentage of the protein of the milk processed. The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}). The highest significant marker on BTA11 is also presented.

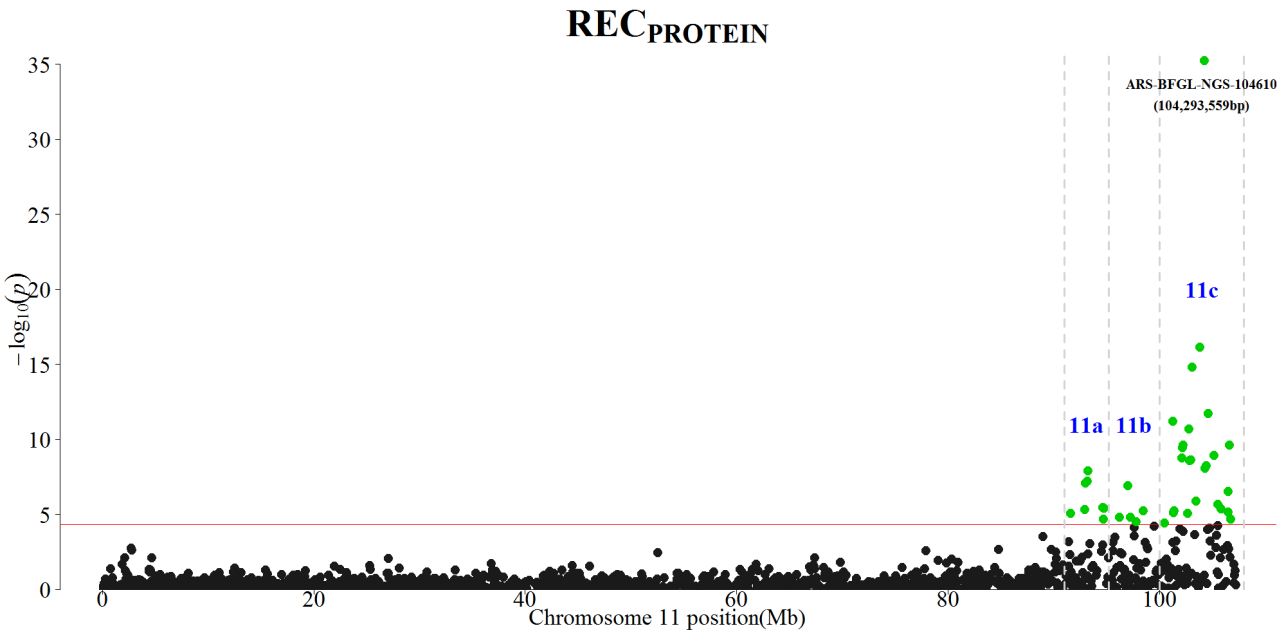
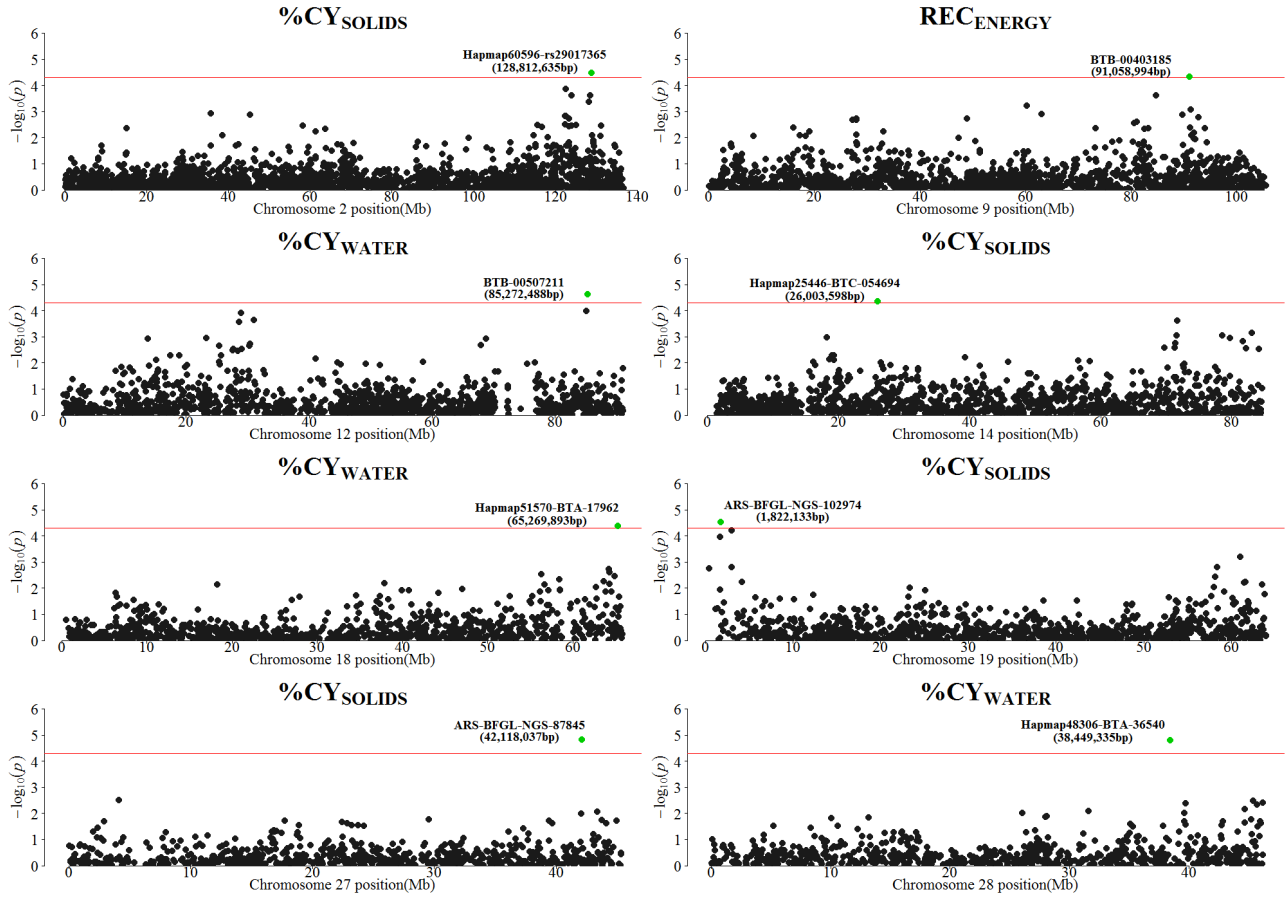


Figure 3. Manhattan plot of –log(*P*-values) for the genome wide association studies (GWAS) on *Bos taurus* autosome (BTA) 2, 9, 12, 14, 18, 19, 27 and 28.

Description: Traits showed significant associations were: individual percentage cheese yield traits (%CY; weight of curd solids and water as percentage of weight of milk processed) and energy of the curd as percentage of the energy of the milk processed (REC_{ENERGY}).

The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}).

The highest significant marker on each chromosome is also presented.



CHAPTER 3

Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield and curd nutrient recovery in dairy cattle

C. Dadousis,* S. Pegolo,* G. J. M. Rosa, †‡ D. Gianola, †‡ G. Bittante,* and A. Cecchinato*[1]

JOURNAL OF DAIRY SCIENCE, 2016 (*accepted*)

*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, Madison Wisconsin, 53706 USA

ABSTRACT

It is becoming common to complement genome wide association studies (GWAS) with gene-set enrichment analysis to deepen the understanding of the biological pathways affecting quantitative traits. Our objective was to conduct a gene ontology and pathway-based analysis to identify possible biological mechanisms involved in the regulation of bovine milk technological traits: coagulation properties (MCP), curd firmness modeling (CF_t), individual cheese yield (CY) and milk nutrient recovery into the curd (REC) or whey loss traits. Results from two previous GWAS studies using 1,011 cows genotyped for 50k single nucleotide polymorphisms (SNP) were used. Overall, the phenotypes analyzed comprised: 3 traditional MCP measures (RCT: Rennet coagulation time defined as the time (min) from addition of enzyme to the beginning of coagulation; k_{20}: the interval (min) from RCT to the time at which a curd firmness of 20 mm is attained; a_{30}: a measure of the extent of curd firmness (mm) 30 min after coagulant addition), 6 CF_t traits [RCT_{eq}: RCT estimated through the CF equation (min); CF_P: potential asymptotic curd firmness (mm); k_{CF}: curd-firming rate constant (% x min^{-1}); k_{SR}: syneresis rate constant (% x min^{-1}); CF_{max}: maximum curd firmness (mm) and t_{max}: time to CF_{max} (min)], 3 individual cheese yield-related traits expressing the weight (wt) of fresh curd ($\%CY_{CURD}$), curd solids ($\%CY_{SOLIDS}$), and curd moisture ($\%CY_{WATER}$) as percentage of wt of milk processed and 4 milk nutrient and energy recoveries (REC) in the curd (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, and REC_{ENERGY} calculated as the % ratio between the nutrient in curd and the corresponding nutrient in processed milk), milk pH and protein percentage. Each trait was analysed separately. In total, 13,269 annotated genes were used in the analysis. The Gene Ontology (**GO**) and Kyoto Encyclopedia of Genes and Genomes (**KEGG**) pathway databases were queried for enrichment analyses. Overall, 21 GO and 17 KEGG categories were significantly associated (false discovery rate at 5%) with 7 traits (RCT, RCT_{eq},

k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS} and REC_{ENERGY}), with some being in common between traits. The significantly enriched categories included calcium signaling pathway, salivary secretion, metabolic pathways, carbohydrate digestion and absorption, the tight junction and the phosphatidylinositol pathways, as well as pathways related to the bovine mammary gland health status, and contained a total of 150 genes spanning all chromosomes but 9, 20, and 27. This study provided new insights into the regulation of bovine milk coagulation and cheese ability that were not captured by the GWAS.

Key words: milk coagulation properties, cow cheese ability, genome-wide association, gene-set enrichment, pathway-based analysis

INTRODUCTION

Cheese manufacture is the main final target of dairy cattle milk production in many countries worldwide. Recently, exploitable additive genetic variation has been reported for different measures of individual bovine cheese yield (Bittante et al., 2013). Moreover, milk coagulation properties (MCP) and curd firmness traits (CF) are used as indicators of cheese production. Although considerable additive genetic variation exists for a variety of direct or indirect cheese traits, high measurement costs and logistics place restrictions on the selection of cows for cheese productivity in breeding programs. A potential strategy is to identify and use genomic regions affecting the cow's ability to produce cheese that could enhance genomic breeding programs. Genome-wide association studies (GWAS) are widely used for this purpose and were proved to be effective in identifying genomic regions associated with the traits of interest. However, due to the stringent statistical thresholds used to deal with multiple testing, a considerable number of important markers may remain undetected when dealing with polygenic

traits (Peng et al., 2010). Moreover, with high SNP density panels each gene might be represented by several proximal SNP, thus splitting its effect into parts that, in turn, might not be able to pass the defined GWAS threshold in a single marker regression (Ha et al., 2015). Additionally, especially in livestock species linkage disequilibrium (LD) spans a wide region in the genome. As a result, a plethora of SNP might be in LD with the causal genomic region which creates extra difficulties in detecting the causal mutation (Hayes, 2013). Besides, while GWAS may be able to locate SNPs significantly associated with the trait of interest, it does not make use of the fact that genes work together in biological pathways and are organized into networks. Further, the effect of a multi-allelic QTL may not be fully captured due to the bi-allelic nature of SNP. As a result, GWAS alone may provide a limited understanding of the complex nature of quantitative traits.

A solution to tackle the aforementioned problems, and deepen the understanding of the genetic background of complex traits, is to move up the analysis from the SNP to the gene and gene-set levels. In a gene-set analysis, a group of related genes (such as genes in a specific pathway or gene ontology) that harbor significant SNPs previously identified in GWAS, is tested for over-representation in a specific pathway (Wang et al., 2011). Indeed, an increasing interest on pathway analysis has been recently observed in dairy cattle, to complement GWAS analyses of quantitative traits (Gambra et al., 2013; Peñagaricano et al., 2013; Iso-Touru et al., 2016; Abdalla et al., 2016).

Thus, the objective of this study was to conduct a gene ontology and pathway analysis to complement previously obtained GWAS results for phenotypes related to bovine milk coagulation properties (MCP), curd firmness modeling (CF_t), individual cheese yield (CY) and milk nutrient recovery into the curd (REC) or whey loss traits.

MATERIALS AND METHODS

Data

Phenotypes. Results of two recent GWAS analyses were used, consisting of 11 MCP and CF_t traits (Dadousis et al., 2016a) as well as 7 individual CY traits (Dadousis et al., 2016b). In brief, the milk MCP-CF_t dataset contained the milk pH, milk protein percentage, 3 traditional MCP obtained from Formagraph [RCT: Rennet coagulation time defined as the time (min) from addition of enzyme to the beginning of coagulation; k_{20}: the interval (min) from RCT to the time at which a curd firmness of 20 mm is attained; a_{30}: a measure of the extent of curd firmness (mm) 30 min after coagulant addition], 4 CF_t equation parameters [RCT_{eq}: RCT estimated through the CF_t equation (min); CF_P: potential asymptotical curd firmness (mm); k_{CF}: curd-firming rate constant (% x min^{-1}); k_{SR}: syneresis rate constant (% x min^{-1})] and 2 derived traits [CF_{max}: maximum curd firmness (mm) and t_{max}: time to CF_{max} (tmin)]. The second GWAS dataset included three individual CY traits expressing the weight (wt) of fresh curd (%CY_{CURD}), curd solids (%CY_{SOLIDS}), and curd moisture (%CY_{WATER}) as percentage of wt of milk processed, and four milk nutrient and energy recoveries (REC) into the curd (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, and REC_{ENERGY}), calculated as the % ratio between the nutrient in curd and the corresponding nutrient/energy in the processed milk. Details about the genotyping and the GWAS analyses are reported in (Dadousis et al., 2016a; Dadousis et al., 2016b).

Genotypic data. Briefly, 1,152 cows were genotyped with the Illumina BovineSNP50 Bead Chip v.2. After quality control [call rate > 95%, minor allele frequency > 0.05 and extreme deviation from Hardy-Weinberg proportions (P > 0.001, Bonferroni corrected)], 1,011 animals and 37,568 SNP, located on 29 autosomes and in the X-chromosome were retained. Slight

differences in the number of individuals and SNP between the two GWAS analyses are attributed to phenotypic editing.

Gene-set Enrichment and Pathway-based Analysis

The gene-set enrichment analysis workflow is represented in Figure 1. In brief, for each trait nominal *P*-values <0.05 from the GWAS analyses were used to identify significant SNP. Using the *biomaRt* R package (Durinck et al., 2005; Durinck et al., 2009), the SNP were assigned to genes if they were within the genomic sequence of the gene or within a flanking region of 15 kb up- and downstream of the gene, to include SNP located in regulatory regions. The size of the flanking region was based on the finding that most SNPs that affect the expression of a gene are located within 15 kb of the gene (Pickrell et al., 2010). The Ensembl *Bos taurus* UMD3.1 database was used as reference (Zimin et al., 2009). The background SNP represent all the SNP tested in the GWAS analyses, while the background genes were the genes associated to those SNP. For the assignment of the genes to functional categories, the Gene Ontology (**GO**) (Ashburner et al., 2000) and Kyoto Encyclopedia of Genes and Genomes (**KEGG**) pathway (Ogata et al., 1999) databases were used. The GO database designates biological descriptors (GO terms) to genes based on attributes of their encoded products and it is further partitioned into three components: biological process, molecular function, and cellular component. The KEGG pathway database contains metabolic and regulatory pathways, representing the actual knowledge on molecular interactions and reaction networks. To avoid testing narrow or broad functional categories, only GO and KEGG categories with more than 10 and less than 1,000 genes were tested. Finally, a Fisher's exact test was performed to test for overrepresentation of the significant genes for each gene-set, i.e., pathway/functional category. False discovery rate (**FDR**) correction (controlled at 5%) was

applied to account for multiple testing. The gene enrichment analysis was performed with the *goseq* R package (Young et al., 2010).

RESULTS AND DISCUSSSION

In total, 17,006 SNP (out of the 37,568 tested) were located in annotated genes or in the 15kb window (up-stream or down-stream from a gene). The total number of background genes annotated in the *Bos taurus* UMD3.1 assembly was 13,269. Each trait was analyzed separately. On average, 1,700 SNP, ranging between 1,301 for $REC_{PROTEIN}$ to 1,899 for RCT_{eq}, were significantly associated to each trait in the GWAS analysis. For each trait, 585 significant SNP were assigned to 500 genes, on average (Figure 1, Table 1). The minimum number of mapped genes was found for $REC_{PROTEIN}$ (n = 399) while the maximum for RCT_{eq} (n = 574).

Enrichment Pathway Analysis

After FDR correction, 21 GO and 17 KEGG categories were found associated with 7 of the tested traits, namely RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS} and REC_{ENERGY}. Some of the categories were in common between traits. A total of 150 significant genes spanning all *Bos taurus* chromosomes (BTA) but 9, 20 and 27 were included into the significantly enriched GO and KEGG categories. Table 2 summarizes all the significant pathways/ontologies, some of which were shared among the aforementioned traits. More precisely, the calcium signaling pathway (KEGG:bta04020) was associated to both RCT and REC_{FAT}; the arrhythmogenic right ventricular cardiomyopathy (ARVC) pathway (KEGG:bta05412) was enriched for both $\%CY_{SOLIDS}$ and REC_{SOLIDS}; the leucocyte transendothelial migration pathway (KEGG:bta04670) was in common between REC_{SOLIDS} and $REC_{ENERGY;}$ and the synapse part cellular component (GO:0044456) was

shared between RCT_{eq} and k_{CF}. Moreover, 6 GO biological process categories related to female sex characteristics and the ovulation cycle appeared significant for both RCT_{eq} and REC_{FAT} (Table 2). Not surprisingly, different pathways/functional categories were enriched for RCT and RCTeq, reflecting the differences in their additive genetic variance found in the GWAS analysis.

Phosphatidylinositol signaling pathway. The phosphatidylinositol signaling pathway (KEGG:bta04070) was significantly enriched for the RCT trait. In milk, phospholipids are mainly present on the surface of the milk fat globules (MFG) and are responsible for the stabilization of the milk fat against coalescence (Rombaut et al., 2007; Walstra et al., 2014). Due to their technological and nutritional properties, previous studies focused on determining the phospholipid content of various dairy products (Rombaut et al., 2007). Recently, the evolution of the phospholipids during the quark cheese process from buttermilk was also examined (Ferreiro et al., 2016). Phosphatidylinositol represents a small fraction of the phospholipid components of milk. Among the significant genes included in the phosphatidylinositol pathway, three phospholipase C beta (*PLCB*) isoforms were present: *PLCB1*, *PLCB3* and *PLCB4*. Phospholipases are responsible for the hydrolyses of phospholipids of the MFG membrane, thereby affecting the stability of the cream emulsion. In addition, the phospholipase treatment of milk was found to reduce fat losses in whey and cooking water and to increase cheese yield by improving fat and moisture retention in the cheese curd in Mozzarella cheese (Lilbæk et al., 2006). Interestingly, an association between a SNP on *PLCB1* (rs41624761) and k_{CF} has been previously reported in a candidate gene analysis (Cecchinato et al., 2015). Moreover, studies related to the effect of MFG size on milk technological properties reported significant relationship of MFG size with MCP (Bland et al., 2015), cheese ripening and structure, as well as stability of dairy products (Lopez et al., 2011). Indeed, the

biological explanation of the connection between phosphatidylinositol pathway and MCP can be found in the tight association between MFG size and phospholipids contents, with higher amounts of phospholipids in small versus large globules likely affecting, in turn, the technological properties of milk.

The salivary secretion pathway (KEGG:bta04970), which was also enriched for the RCT trait, shared 6 genes with the phosphatidylinositol signaling pathway including *PLCB1*, *PLCB3* and *PLCB4.* Interestingly, histatherin (*HSTN*) was also present in the list of significant genes for this pathway (Supplementary Table S1). Histatherin is a ruminant-specific gene that encodes for a host-defense related protein in the cow's oral cavity and milk, which may also be involved in the response to mastitis (Ju, Z., 2014). The *HSTN* has been also proposed as a candidate gene related to MCP and CY traits in our previous GWAS analyses (Dadousis et al., 2016a; Dadousis et al., 2016b) and more precisely it was associated to the three MCP, CF_P, CF_{max} and REC_{FAT}.

Calcium signaling-related pathway. The calcium signaling pathway (KEGG:bta04020) was significantly enriched for both RCT and REC_{FAT}. It is widely known that calcium (Ca) is one of the major components of the casein micelles. During the cheese process, after rennet addition, the casein reacts with Ca ions and precipitates. This phenomenon consists the basis of milk clotting (Walstra et al., 2014). Moreover, low content of the total and micellar Ca has been associated to non-coagulating milk (Gustavsson et al., 2014; Malacarne et al., 2014). Interestingly, transcriptomic analysis of mammary gland in mice evidenced that the calcium ion binding ontology was significantly over-represented among the differentially expressed genes associated with enhanced maternal performance phenotype (Ramanathan et al., 2008). Further analysis

showed a positive correlation between the calcium signaling pathway and the lactation performance in mice (Wei et al., 2013).

The ARVC pathway (KEGG:bta05412) was enriched for both $\%CY_{SOLIDS}$ and REC_{SOLIDS} with 9 significantly enriched genes being in common. Moreover, this pathway shared 6 genes with the cardiac muscle contraction pathway (KEGG:bta04260) (Supplementary Table S1) which was also enriched for the $\%CY_{SOLIDS}$. Notably, genes encoding for several subunits of the voltage-dependent calcium channel complex were included in these pathways (e.g. calcium voltage-gated channel subunit alpha-1D and calcium voltage-gated channel auxiliary subunit alpha -2/delta-3].

Bovine reproduction-related ontologies. The rennet coagulation time obtained from an extended CF testing time (RCT_{eq}) and the REC_{FAT} were associated with the GO terms related to female characteristics such as the ovulation cycle (GO:0042698) and female gonad development (GO:0008585). All the significant genes were shared among these biological processes (Supplementary Table S1). In a similar gene-set enrichment and pathway analysis, the broader GO categories of reproduction (GO:0000003) and reproductive process (GO:002214) were associated with milk yield, milk fat and protein yields, and fertility index in the Nordic Red cattle (Iso-Touru et al., 2016). Indeed, a close link between the duration of estrus and multiple ovulation rate and milk production in dairy cattle was previously reported (Wiltbank et al., 2006).

When specifically looking at the significant genes involved in these pathways/processes (Supplementary Table S1), the alpha casein S1 (*CSN1S1*) and the luteinizing hormone receptor (*LHCGR*) were included. Interestingly, in a candidate gene approach an association between an SNP on *CSN1S1* (rs109817504) and CF_P has been previously reported (Cecchinato et al., 2015). Moreover, *CSN1S1* genetic variants were shown to affect MCP of buffalo and goat milk (Caravaca

et al., 2011; Devold et al., 2011; Bonfatti et al., 2012). Considering *LHCGR*, this gene has been significantly associated with milk composition and in particular with milk fat and total solid percentages (Molee et al., 2015). Notably, a positive impact of fat on milk coagulation properties was highlighted by Bland et al. (2015).

Additionally, ontologies related to the nervous tissue and more specifically to neuron parts (e.g. dendrite) and functions (e.g. synapse) were enriched for the RCT_{eq}. A possible interpretation can rely on the fact that during pregnancy and lactation many factors and signals (including the neuroendocrine signal of prolactin) act to adapt the pattern of neuronal responses to the lactating state (Grattan, 2002; Akers, 2016). Interestingly, in 3 of these GO terms (GO:0044456, GO:0097458 and GO:0045202), the vacuolar protein sorting-associated protein 35 gene (*VPS35*) was included. This gene has been recently proposed as a candidate gene strongly related with milk coagulation in Swedish red cows (Duchemin et al., 2016).

Mammary gland- and mastitis – related pathways and ontologies. Ontologies related to potassium channels (GO:0008076, GO:0034705) were significantly enriched for RCT_{eq}. The role of the voltage-gated potassium channels is to transfer ions across the cell membrane (Yellen, 2002). In milk, the concentrations of Na^+, K^+ and Cl^- are the most important ions for electrical conductivity (EC). It is well established that milk EC can be also used as an indicator of mastitis (Norberg, 2005; Viguier et al., 2009). Interestingly, the tight junction pathway (KEGG:bta04530) was significantly enriched for the REC_{ENERGY}. Tight junctions of the mammary epithelium control the movement of lactose and K^+ to the extracellular fluid, while Na^+ and Cl^- are moving into the milk. Tight junctions are known to be related with milk mammary gland development and milk secretion (Nguyen and Neville, 1998; Ramanathan et al., 2008; Stelwagen and Singh, 2014). More

precisely, increased milk secretion is connected to a decrease in the tight junction permeability. After intramammary infection, destruction of tight junctions and of the ion-pumping system causes an increase in the concentration of Na^+ and Cl^- in the milk and consequently increases the milk EC (Norberg, 2005). It has been reported that the technological properties of milk (such as MCP and CF_t) are unfavorably influenced by mastitis indicators (Bittante et al., 2012; Bobbo et al., 2016). Indeed, other pathways related to the mastitis were significantly enriched. In particular, REC_{SOLIDS} and REC_{ENERGY} were associated to the leucocyte transendothelial migration (KEGG:bta04670). Leucocytes are typically present in the milk and consist the majority of the somatic cell count (SCC). Their concentration in milk increases after bacterial infections and thus they are widely used as an indicator of mastitis (Dosogne et al., 2003). The leucocyte transendothelial migration pathway has been previously linked to milk and fat yield in dairy cattle (Edwards et al., 2015). Immune response-related categories, e.g. the immune system process (GO:0002376), have been recently found to be related with fat yield, milk yield and fertility (Iso-Touru et al., 2016). The tight junction and the leucocyte transendothelial migration pathways shared three significant genes, namely junctional adhesion molecule 2 (*JAM2*), actinin alpha 1 (*ACTN1*) and catenin alpha 3 (*CTNNA3*) (Supplementary Table S1). Interestingly, *JAM2* is located on BTA1 at ~10.1Mb and a weak signal for RCT_{eq} at ~9.5Mb has been detected in our GWAS analysis (Dadousis et al., 2016a).

In addition to this, the cell adhesion molecules (CAMs) pathway (KEGG:bta04514) was enriched for REC_{ENERGY}. The CAMs pathway is involved in a wide range of biologic processes, including immune response and neuronal cell adhesion. In the study of Ramanathan et al. (2008) differentially expressed genes belonging to this pathway were also enriched and related with the mammary development and milk secretion in mice. Moreover, the gap junction pathway

(KEGG:bta04540), related to cell communication, was enriched for RCT. Functional analyses evidenced that the broader pathway of cell-cell signaling was significantly associated with MFG global gene expression during lactation in human (Maningat et al., 2009).

Interestingly, the ARVC pathway detected in our study has also been associated with the mammary gland functionality in pregnant sows in a study focusing in sow's mammary transcriptome in late gestation (Zhao et al., 2013).

Metabolism-related pathways. The broad category of "metabolic pathways" (KEGG:bta01100) was associated to REC_{SOLIDS}. Among the genes included in this specific pathway, three polypeptide N-acetylgalactosaminyltransferase (*GALNT*) isoforms were significant in our study (*GALNT1, GALNT13* and *GALNT18*) (Supplementary Table S1). Although no association was detected in our previous GWAS analyses on BTA24, the *GALNT1* gene located on BTA24 has been reported as a candidate gene in another recent GWAS study related to MCP (Gregersen et al., 2015). This gene encodes for the GalNAc-T enzyme that is involved to κ-casein glycosylation (Holland et al., 2005). Higher content of glycosylated kappa casein has been linked to improved milk coagulation (Poulsen et al., 2016). In the recent gene-set enrichment study of (Iso-Touru et al., 2016) the metabolic process ontology (GO:0008152) was significantly enriched for milk, fat and protein yields and fertility in Nordic Red cattle. Moreover, in our study the carbohydrate digestion and absorption pathway (KEGG:bta04973) was enriched for REC_{SOLIDS}. The central carbohydrate of the milk is lactose. Although a strong influence of lactose on MCP has been recently reported (Bland et al., 2015), our knowledge on the exact mechanism is still limited. However, lactose also is related to SCC and mastitis. More precisely, a decrease of lactose is observed during mastitis (Kitchen, 1981). Not surprisingly, two genes (phosphoinositide-3-

kinase, regulatory subunits 3 and 5; *PIK3R3* and *PIK3R5*) were in common between the carbohydrate digestion and absorption, and the leukocyte transendothelial migration pathways (Supplementary Table S1).

Cancer-related pathways. Among the significantly enriched pathways detected for REC_{SOLIDS}, pathways in cancer (KEGG:bta05200), endometrial cancer (KEGG:bta05213) and melanoma (KEGG: bta05218) were present with some significant genes being in common, including also phosphoinositide-3-kinase regulatory subunit 3 (*PIK3R3*), *PIK3R5,* and AKT serine/threonine kinase 3(*AKT3*) (Supplementary Table S1). These genes have a role in the PIK-Akt signaling pathway which was associated with mammary development and breast cancer (Wickenden and Watson, 2010). Moreover, in a gene expression study in human mammary gland investigating the MFG transcriptome, one of the most significant networks associated with the top expressed genes was the cancer pathway (Maningat et al., 2009). Moreover, the aforementioned cancer-related KEGG pathways were associated with genetic variants in mammary development, prolactin signaling and involution pathways which were linked to bovine milk production traits (Raven et al., 2014). However, as highlighted by those authors, the significance of this information is challenged by the fact that KEGG database includes a large compendium of cancer-related gene sets.

Apparently, aside from pathways which could be strictly associated with milk technological properties (e.g. phosphatidylinositol signaling), pathways not directly related to the traits of our interest were also detected (e.g. pathways related to cancer). However, as publicly available ontologies and pathways in cattle are still limited (compared to human) and/or not all are well-described, some of our results may be misleading, especially when the detected genes are

involved in various biological processes (Fan et al., 2015). It is likely that, when more complete gene sets become available, more competitive pathways might be detected and the power to identify genomic regions influencing these traits might increase. In this respect, transcriptomic methods (e.g. RNA-seq) may represent a useful tool to complement the present analysis and validate the achieved biological information.

Finally, it is worth noting that our gene-set enrichment analysis was conducted using a panel of SNP obtained from a single marker regression GWAS, which relies on a simplified theory of the genomic background of traits, without considering for instance the joint effect of SNP. Hence, other approaches, e.g. GWAS exploring SNP by SNP interactions, might provide a better basis for biological pathway analysis.

CONCLUSION

To our knowledge, this is the first pathway-based association analysis related to milk technological traits. In animal breeding, studies are generally focused on single SNP-based associations with the traits of interest. The present pathway-based analysis provided new insights with respect to the previously conducted GWAS analyses, confirming that complex traits (i.e., milk technological properties) may be affected by the joint additive effect of several genes which cluster in specific biological pathways. In particular, calcium and phosphatidylinositol signaling, overall metabolism, carbohydrate digestion and absorption as well as pathways related to the bovine mammary gland health status were significantly enriched.

The highlighted pathways and gene ontologies detected associated with technological traits may be useful in further studies on fine mapping of genes and development of marker-assisted breeding programs. However, further validation and replication of the most promising described

pathways is needed to explore their role in relation to bovine milk coagulation and cheese-making characteristics.

ACKNOWLEDGMENTS

The authors wish to thank the Trento Province (Italy), the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy) and the Superbrown Consortium of Bolzano and Trento for financial and technical support. C. Dadousis benefitted from financial support of the CARIPARO (Cassa di Risparmio di Padova e Rovigo) Foundation (Padua, Italy). Authors wish to thank also Prof. F. Peñagaricano (Department of Animal Sciences, University of Florida) for his help in setting up the analysis.

Table 1 Number of significant[1] SNP identified from GWAS and genes mapped by trait.

Trait[2]	No. of significant SNP	No. of significant SNP assigned to genes	No. of significant mapped genes[3]
Milk composition			
pH	1,848	624	552
Protein, %	1,808	641	563
Traditional MCP			
RCT	1,739	551	487
k_{20}	1,789	614	539
a_{30}	1,724	572	496
Curd firming			
RCT_{eq}	1,899	639	574
CF_P	1,423	486	422
CF_{max}	1,724	599	536
T_{max}	1,786	598	531
k_{CF}	1,822	603	545
k_{SR}	1,872	625	562
Cheese yields, %			
$\%CY_{CURD}$	1,817	621	538
$\%CY_{SOLIDS}$	1,796	590	533
$\%CY_{WATER}$	1,826	605	525
Recoveries, %			
REC_{SOLIDS}	1,797	581	527
REC_{FAT}	1,496	503	447
$REC_{PROTEIN}$	1,301	444	399
REC_{ENERGY}	1,800	627	548
Background[4]	37,568	17,006	13,269

[1] *P*-value<0.05

[2]pH= milk pH; Protein, % = milk protein (%); RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min); %CY = weight of fresh curd, curd solids, and curd water as percentage of weight of milk processed; REC = protein, fat, solids, and energy of the curd as percentage of the protein, fat, solids, and energy of the milk processed.

[3]Ensembl *Bos taurus*UMD3.1 (http://www.ensembl.org/index.html); window: 15kb

[4]Background represents the total number of SNP used in the GWAS analyses, the number of SNP linked to genes and the genes mapped to those SNP.

Table 2 Gene ontology (GO) terms and Kyoto encyclopedia of genes and genomes (KEGG) pathways significantly enriched using genes associated with RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS} and REC_{ENERGY}.

Trait[1]	Category[2]	Term	No. genes in the term	No. sig genes[3]	FDR[4]
RCT	KEGG	bta04070:Phosphatidylinositol signaling	53	10	$1.87x10^{-5}$
		bta04730:Long-term depression	44	8	$1.69x10^{-4}$
		bta04540:Gap junction	57	9	$2.07x10^{-4}$
		bta04270:Vascular smooth muscle contraction	74	10	$3.47x10^{-4}$
		bta04020:Calcium signaling pathway	118	13	$3.79x10^{-4}$
		bta04970:Salivary secretion	55	8	$8.18x10^{-4}$
RCTeq	GO_BP	GO:0048511~Rhythmic process	46	9	$1.87x10^{-4}$
		GO:0008585~Female gonad development	14	5	$2.15x10^{-4}$
		GO:0022602~Ovulation cycle process	14	5	$2.15x10^{-4}$
		GO:0042698~Ovulation cycle	14	5	$2.15x10^{-4}$
		GO:0046545~Development of primary female sexual characteristics	14	5	$2.15x10^{-4}$
		GO:0046660~Female sex differentiation	14	5	$2.15x10^{-4}$
	GO_CC	GO:0030425~Dendrite	39	11	$5.01x10^{-7}$
		GO:0044456~Synapse part	76	15	$7.12x10^{-7}$
		GO:0097458~Neuron part	162	22	$1.84x10^{-6}$
		GO:0045202~Synapse	102	16	$7.30x10^{-6}$
		GO:0036477~Somatodendritic compartment	62	11	$6.15x10^{-5}$
		GO:0043005~Neuron projection	110	15	$7.58x10^{-5}$
		GO:0008076~Voltage-gated potassium channel complex	12	5	$9.15x10^{-5}$
		GO:0034705~Potassium channel complex	13	5	$1.44x10^{-4}$

		GO:0008021~synaptic vesicle	22	6	2.63×10^{-4}
		GO:0098793~Presynapse	22	6	2.63×10^{-4}
k_{CF}	GO_CC	GO:0044456~Synapse part	76	13	1.16×10^{-5}
		GO:0098794~Postsynapse	48	10	2.01×10^{-5}
$\%CY_{SOLIDS}$	KEGG	bta05412:Arrhythmogenic right ventricular cardiomyopathy (ARVC)	50	10	2.42×10^{-5}
		bta4260:Cardiac muscle contraction	41	8	1.90×10^{-4}
REC_{FAT}	GO_BP	GO:0008585~Female gonad development	14	6	3.37×10^{-6}
		GO:0022602~ Ovulation cycle process	14	6	3.37×10^{-6}
		GO:0042698~Ovulation cycle	14	6	3.37×10^{-6}
		GO:0046545~Development of primary female sexual characteristics	14	6	3.37×10^{-6}
		GO:0046660~Female sex differentiation	14	6	3.37×10^{-6}
		GO:0001541~Ovarian follicle development	10	5	9.30×10^{-6}
		GO:0008406~Gonad development	21	6	4.98×10^{-5}
		GO:0045137~Development of primary sexual characteristics	21	6	4.98×10^{-5}
		GO:0007548~Sex differentiation	24	6	1.13×10^{-4}
		GO:0048511~Rhythmic process	46	8	1.32×10^{-4}
	KEGG	bta04020:Calcium signaling pathway	118	13	1.68×10^{-4}
REC_{SOLIDS}	KEGG	bta05412:Arrhythmogenic right ventricular cardiomyopathy (ARCV)	50	9	1.34×10^{-4}
		bta05218:Melanoma	44	8	2.94×10^{-4}
		bta05200:Pathways in cancer	195	18	7.73×10^{-4}
		bta01100:Metabolic pathways	635	42	8.10×10^{-4}

		bta04260:Cardiac muscle contraction	41	7	1.04x10^{-3}
		bta04670:Leucocyte transendothelial migration	66	9	1.13x10^{-3}
		bta05213:Endometrial cancer	31	6	1.20x10^{-3}
		bta04930:Type II diabetes mellitus	32	6	1.43x10^{-3}
		bta04973:Carbohydrate digestion and absorption	24	5	2.20x10^{-3}
REC_{ENERGY}	KEGG	bta04670:Leukocyte transendothelial migration	66	10	3.49x10^{-4}
		bta04514:Cell adhesion molecules (CAMS)	79	11	3.81x10^{-4}
		bta04530:Tight junction	83	11	5.87x10^{-4}

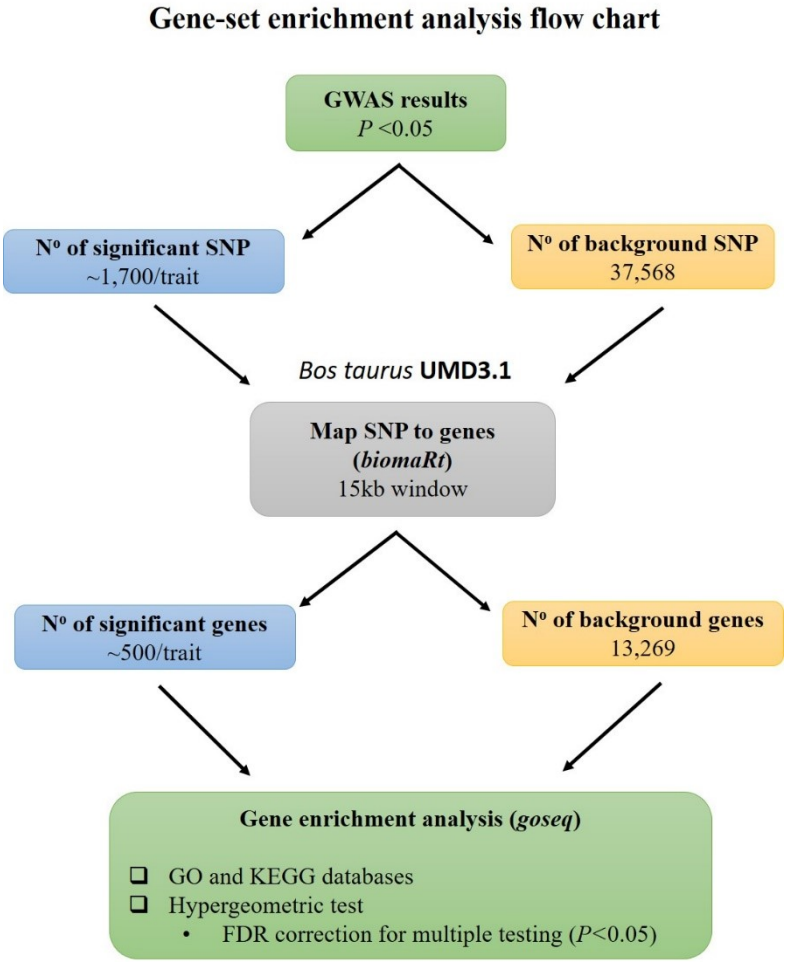
[1]RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; k_{CF} = curd-firming rate constant (% x min^{-1}); $\%CY_{SOLIDS}$ = weight of curd solids as percentage of weight of milk processed; REC = fat, solids, and energy of the curd as percentage of the fat, solids, and energy of the milk processed.

[2]KEGG: KEGG pathway; GO_BP.GO biological process; GO_CC: GO cellular component

[3]Significant genes after mapping the significant SNP to genes using Ensembl *Bos Taurus* UMD3.1 as reference (http://www.ensembl.org/index.html)

[4]False discovery rate (FDR) correction for multiple testing (*P*-value <0.05).

Figure 2. Flow chart for the gene-set enrichment analysis.



CHAPTER 4

Inferring individual cow effects, dairy system and genetics on latent variables underlying milk yield and quality, protein composition and cheese-making traits in dairy cattle

C. Dadousis,* C. Cipolat-Gotet,* G. Bittante,* and A. Cecchinato*[1]

JOURNAL OF DAIRY SCIENCE, 2016 (*submitted*)

*Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

ABSTRACT

The aim of this study was to investigate the potential use of latent variables (factors; Fs) in dairy cattle breeding with a focus in cheese related phenotypes. In total, 26 traits related to milk yield and quality (including protein fractions), coagulation and curd firmness (CF_t) indicators, and cheese-making traits [cheese yields (%CY) and nutrient recoveries in the curd (REC)] were analyzed through multivariate factor analysis (MFA) using a *varimax* rotation. All phenotypes were measured in 1,264 Brown Swiss cows. Ten mutual orthogonal Fs were obtained explaining 74% of the original variability. Those Fs captured basic concepts of the *cheese-making* process. More precisely, the first four Fs, sorted by variance explained, were able to capture the underlying structure of the CY percentage (F1: %CY), the CF process with time (F2: CF_t), the milk and solids yield (F3: Yield) and the presence of nitrogen (N) into the cheese (F4: Cheese N). Moreover, 4 Fs (F5: as_1-β-CN, F7: κ-β-CN, F8: as_2-CN and F9: as_1-CN-Ph) were related to the basic milk caseins and 1 factor was associated with the α-LA whey protein (F10: α-LA). A factor describing the udder health status of a cow (F6: Udder health), mainly loaded on lactose, other nitrogen compounds in the milk and SCS, was also obtained. To assess the practical use of the Fs into breeding, we inferred the effects of some potential source of variations (e.g. stage of lactation and parity) including feeding and management systems. Moreover, genetic parameters of the new latent variables were estimated using single and bi-variate animal models under a Bayesian framework.

Stage of lactation had a significant effect for the majority of the Fs, followed by parity. The patterns of the factor scores within the stage of lactation and parity were coherent to the given name of the factor. Differences among dairy farm systems for F3: Yield and F1: %CY existed. Heritability estimates (within-herd) varied between 0.11-0.72 (F3: Yield and F7: κ-β-CN, respectively). Although the Fs were phenotypically uncorrelated, considerable additive genetic

correlations existed among them, with highest values observed between F10: α-LA and F6: Udder health (-0.67) as well as between F9: as_1-CN-Ph and F3: Yield (-0.60). Our work demonstrates the usefulness of latent variables in reducing a large number of variables to a few latent factors with biological meaning and representing groups of traits that describe a complex process like cheese-making. Multivariate factor analysis therefore, could be a valuable tool for studying the influence of different production environments and individual animal factors on cheese-making related phenotypes, and for developing breeding strategies to improve cow's cheese-related traits.

Key words: multivariate factor analysis, milk protein, coagulation, curd firmness, individual cheese yield

INTRODUCTION

A considerable amount of dairy cattle milk is destined for cheese production, and an interest in breeding focusing on cow's cheese-making ability exists. However, cheese produced at the individual cow level is the result of a complex process with plenty of elements involved, from milk quality characteristics (e.g. percentage of protein and fat in the milk), milk coagulation properties (MCP), curd firmness (CF) and cheese-making characteristics, such as the quantity of cheese obtained from a given amount of milk (CY).

Previous analysis have shown considerable genetic variation of MCP (Bittante et al., 2012; Gustavsson et al., 2014), milk protein composition (Bonfatti et al., 2011) and CY traits (Bittante et al., 2013a; Cecchinato et al., 2015). However, breeding goals usually include a considerable number of traits focused not only on production, but also on health status, longevity, reproduction of cows, etc. Thus, although detailed phenotyping is required to understand the biological and genetic background of the traits of interest, inclusion of a large number of traits into a selection

index creates difficulties in interpretation and computations. To overcome the problem of data dimension, and for a better understanding of complex phenomena, multivariate factor analysis (MFA) has been investigated for a variety of traits, such as milk composition, MCP and CY traits in dairy cattle, sheep and goat (Macciotta et al., 2012; Manca et al., 2016) and for milk fatty acids in dairy cows (Conte et al., 2016; Mele et al., under review). Moreover, MFA has been used within the framework of structural equation modelling, for e.g., analyzing carcass traits in swine (Peñagaricano et al., 2015) and studying bovine mastitis (Detilleux et al., 2013).

Factor analysis belongs to the general framework of multivariate analysis. The main idea of MFA is that n observed variables, x, can be expressed as linear functions of p ($p < n$) latent variables (factors; Fs): this statistical approach focuses in understanding relationships (the underlying latent concept that the measured variables represent) among a set of observed variables. Thus, Fs can be considered as variables that are not measurable, but they can be extracted from a set of measured-indicator variables (Bollen, 2014) by making use of their covariance structure. In such a way, only the underlying concept of interest is kept for further analysis while at the same time data reduction is achieved.

Our objective was to create a new set of latent phenotypes related to milk quality, technological properties, and cheese-making traits using MFA, and to assess their potential use in dairy cattle breeding by i) studying some individual sources of variation (i.e., stage of lactation and parity) and herd as well as the effects of feeding and management systems on the new latent variables and ii) inferring their genetic parameters.

MATERIALS AND METHODS

Animals, Herds and Dairy Farming Systems

Milk samples from 1,264 Italian Brown Swiss cows reared in 85 herds located in Trento Province (Italy) were collected. A full description of the sampling procedure can be found in (Cipolat-Gotet et al., 2012). In brief, ~15 cows/herd were individually sampled once (evening milking). Samples were processed within 20 h after collection. Information on cows and herds was supplied by the Breeders Association of Trento Province. Pedigree information was provided by the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy). We considered cows with phenotypic records available for the investigated traits and all known ancestors. Each sampled cow had known ancestors for at least four generations and the pedigree file included 8,845 animals.

The farming systems have been previously analyzed and reported in Sturaro et al. (2009 and 2013). In brief, two main categories of farming systems were distinguished: i) traditional farms and ii) modern dairy systems. The traditional systems represent small and old barns where feeding is heavily based on forage and cows are tied. Modern farms, with loose cows and milking parlor, were further distinguished in 3 sub-categories depending on the feeding system: a) modern dairy system but without use of total mixed ratio (TMR), b) modern dairy farms with silage-based TMR and c) modern dairy farms without silage using water to moisturize TMR.

Cheese-making Phenotypes

Analysis of Milk and Milk Protein Fractions. Individual milk subsamples were analyzed for protein, fat, lactose and casein contents using MilkoScan FT6000 (Foss, Hillerød, Denmark). The pH of the subsamples was measured using a Crison Basic 25 electrode (Crison, Barcelona,

Spain). Somatic cell count measures were obtained by Fossomatic FC counter (Foss, Hillerød, Denmark) and were logarithmic transformed [(SCS = $\log_2$SCC/100,000) – 3] (Ali and Shook, 1980).

Protein fractions [αs_1-, αs_2-, β- and κ- casein (CN), β-lactoglobulin (β-LG) and α-lactalbumin (α-LA)] were measured by the RP-HPLC method (Bonfatti et al., 2008) and were expressed as proportions to the total milk nitrogen (N) content. Further, the phosphorylated form of the αs_1-CN was obtained as proposed by Bonfatti et al., (2011). The remained N milk compounds were also included in the analysis and were calculated by subtracting the sum of the protein fractions from the total N milk content.

Milk Coagulation Properties and Modeling the CF_t. Measures of milk coagulation properties were obtained using the Formagraph instrument by Foss Electric A/S according to the procedure described in (Cipolat-Gotet et al., 2012).

Files containing 360 CF values for each milk sample, recorded every 15 sec for 90 min, were retrieved and used to estimate a set of parameters of CF at time t (CF_t) according to equations and methodology developed by (Bittante et al., 2013b). Estimated parameters included: rennet coagulation time (RCT_{eq}, min), potential asymptotical curd firmness (CF_P, mm), representing the maximum potential curd firmness of a given sample after infinite time in the absence of syneresis, curd-firming rate constant (k_{CF}, % x min^{-1}) which measures the relative increment of CF toward CF_P, that is predominant until reaching the maximum curd firmness (CF_{max}, mm; at time t_{max}, min) and before syneresis became prevalent, measured by k_{SR} (% x min^{-1}).

To avoid convergence and estimation problems, CF_P was calculated multiplying CF_{max} by 1.34, that is the coefficient resulting from the linear regression between CF_P and CF_{max} values obtained in a preliminary analysis (Stocco et al., 2016). The other three CF_t model parameters

(RCT_{eq}, k_{CF}, and k_{SR}) were estimated by curvilinear regression using the nonlinear procedure (PROC NLIN) in the SAS software (SAS Institute Inc., Cary, NC).

Individual Cheese Yield and Curd Nutrient Recoveries. A detailed description of the individual model-cheese processing can be found in (Cipolat-Gotet et al., 2013). The phenotypes were obtained through a model cheese-making procedure on 1,500mL of milk for each cow. In brief, the traits analyzed were: i) three %CY traits, expressing the weight (wt) of fresh curd ($\%CY_{CURD}$), of curd dry matter ($\%CY_{SOLIDS}$), and of water retained in the curd ($\%CY_{WATER}$) as percentage of wt of milk processed, and ii) four REC traits representing the proportion of nutrients and energy of the milk retained in the curd (REC_{SOLIDS}, REC_{FAT}, $REC_{PROTEIN}$ and REC_{ENERGY} calculated as the % ratio between the nutrient in curd and the corresponding nutrient in processed milk). The recovery energy in the curd was calculated as the difference between energy in the milk and in the cheese (NRC, 2001).

Statistical Analysis

***Multivariate Factor Analysis*.** To avoid severe multicollinearity problems, 3 out of the 29 phenotypes (CF_{max}, $\%CY_{CURD}$ and $REC_{SOLIDS)}$ were excluded. As mentioned above, CF_{max} is proportional to CF_P. Moreover, the $\%CY_{CURD}$ is the sum of $\%CY_{WATER}$ and $\%CY_{SOLIDS}$, while REC_{SOLIDS} has phenotypic correlation with REC_{ENERGY} and $\%CY_{SOLIDS}$ greater than 0.9. The remaining twenty-six phenotypes were simultaneously analyzed in the following factor model:

$$\boldsymbol{x} = \boldsymbol{\Lambda\xi} + \boldsymbol{\delta}, \qquad [1]$$

where **x** is a vector containing the measured phenotypes, $\boldsymbol{\xi}$ is a vector of the factors, $\boldsymbol{\Lambda}$ contains the

factor loadings (λ) relating the factors to the original variables and **δ** is a vector of the residuals.

At a first step, the difference between Pearson and partial correlations of the measured variables was used as a hint to assess the adequacy of the data for MFA. This difference can be viewed as a way to control if the correlation between 2 variables is mediated by other variables, with a high value indicating the existence of a latent structure. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was applied to quantify this difference (Dziuban and Shirkey, 1974; Kaiser and Rice, 1974). Further, estimatory factor analysis was applied. The factors were extracted based on prior knowledge, biological interpretation and the proportion of original variance explained. Moreover, factor rotation (*varimax*) was used to identify simple structure. Factor loadings > |0.4| were considered as "significant" to explain the factors. Ten factors were extracted and kept for further analysis. The MFA was performed with the *psych* package (Revelle, 2014) in R (R Core Team, 2013).

Mixed model analysis. To estimate sources of variation related to the factors, the 10 factor scores were analyzed with the following model:

$$y_{ijklm} = \mu + dairy\ system_i + herd_j(dairy\ system)_i + parity_k + DIM_l + e_{ijklm} \quad [1]$$

where y_{ijklm} is the observed phenotype (i.e, the factor scores); μ is the overall mean; $dairy\ system_i$ is the fixed effect of the i^{th} dairy system (i=1 to 4); $herd_j(dairy\ system)_i$ is the random effect of the j^{th} herd (j = 1 to 85) $\sim N(0, \mathbf{I}\sigma_h^2)$ nested within the i^{th} dairy system; $parity_k$ is the fixed effect of the k^{th} parity (k = 1 to 4 or more lactations); DIM_l is the l^{th}30-d class of DIM, 11 classes; and e_{ijklm} is the residual random error term $\sim N(0, \mathbf{I}\sigma_e^2)$; where **I** is an identity matrix, σ_h^2, and σ_e^2 are herd/date and residual variances, respectively. The significance of the dairy system was tested on the error line of the herd within dairy system, while for parity and DIM class the error line of the residual variance was used. Orthogonal post hoc contrasts ($P < 0.05$) were built

for dairy system: i) the "Traditional" dairy system was compared with the "Modern" systems; ii) within the modern systems, the "No TMR" herds were compared with the TMR herds; iii) within the TMR herds, those that use silage were compared with those that use water.

Genetic analysis. Non-genetic effects previously described were considered for the estimation of genetic parameters of factor scores.

At first, univariate models were fitted to estimate variance components and heritability for the traits of concern. The model assumed was:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{h} + \mathbf{Z}_2\mathbf{a} + \mathbf{e} \qquad [2]$$

where **y** was the vector of phenotypic records (i.e, the factor scores) with dimension *n;* **X**, $\mathbf{Z}_1$, and $\mathbf{Z}_2$ were appropriate incidence matrices for the systematic effects (**b**), herd/date effects (**h**), and polygenic additive genetic effects (**a**), respectively; and **e** was the vector of residual effects. More specifically, **b** included the non-genetic effects previously described in [1].

All models were analyzed under a standard Bayesian approach. The joint distribution of the parameters in a given model was proportional to:

$$p(\mathbf{b}, \mathbf{h}, \mathbf{a}, \sigma_e^2, \sigma_h^2, \sigma_a^2|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{b}, \mathbf{h}, \mathbf{a}, \sigma_e^2)p(\sigma_e^2)p(\mathbf{b})$$

$$\times p(\mathbf{h}|\sigma_h^2)p(\sigma_h^2)p(\mathbf{a}|\mathbf{A}, \sigma_a^2)p(\sigma_a^2),$$

where **A** was the numerator pedigree relationship matrix between individuals and σ_a^2 the additive genetic variance. The *a priori* distributions of **h** and **a** were assumed to be multivariate normal, as follows:

$$p(\mathbf{h}|\sigma_h^2) \sim N(\mathbf{0}, \mathbf{I}\sigma_h^2)$$

$$p(\mathbf{a}|\sigma_a^2) \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2),$$

The priors for **b** and the variance components were assumed to be flat.

To estimate genetic correlations between the traits, we conducted a set of bivariate analyses that implemented model [1] in its multivariate version. In this case, the involved traits were assumed to jointly follow a multivariate normal distribution along with the additive genetic, herd, and residual effects. For these effects, the corresponding prior distributions were:

$$\mathbf{a}|\mathbf{G}_0, \mathbf{A} \sim MVN(0, \mathbf{G}_0 \otimes \mathbf{A}),$$

$$\mathbf{h}|\mathbf{H}_0, \sim N(0, \mathbf{H}_0 \otimes \mathbf{I}_n) \text{ and}$$

$$\mathbf{e}|\mathbf{R}_0, \sim N(0, \mathbf{R}_0 \otimes \mathbf{I}_m),$$

where $\mathbf{G}_0, \mathbf{H}_0, \mathbf{R}_0$ were the corresponding variance-covariance matrices between the involved traits, and $\mathbf{a}$, $\mathbf{h}$, and $\mathbf{e}$ were vectors of dimension equal to the number of animals in the pedigree times the number of traits considered.

Marginal posterior distributions of all unknowns were estimated by applying the Gibbs Sampling algorithm. The program TM (http://snp.toulouse.inra.fr/~alegarra) was used for all Gibbs sampling procedures. The lengths of the chain and the burn-in period were assessed by visual inspection of trace plots. After some preliminary analysis, chains of 850,000 samples were kept, with a burn in period of 50,000. Subsequently, one in every 200 successive samples was retained. The lower and upper bounds of the highest 95% probability density regions (HPD95%) for the parameters of concern were obtained from the estimated marginal densities. The posterior mean was used as the point estimate for all parameters.

Across-herd (AH) and intra-herd (IH) heritability was computed as following:

$$h^2_{AH} = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_h + \sigma^2_e}$$

$$h^2_{IH} = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_e}$$

Additive genetic correlations (r_g) were estimated as $r_g = \frac{\sigma_{ai,aj}}{\sigma_{ai} \cdot \sigma_{aj}}$, where, $\sigma_{ai,aj}$ is the additive genetic covariance between factor *i* and *j*, and σ_{ai} and σ_{aj} are the additive genetic standard deviations for factor *i* and *j*, respectively.

RESULTS

Factors

Descriptive statistics of the full dataset including all 29 phenotypes are shown in Table 1. The phenotypic Pearson and partial correlations of the 26 traits analyzed in MFA are presented in Figure 1. The average KMO value in our dataset was 0.55. The protein fractions (except α-LA) showed the lowest KMO values (< 0.5) together with REC_{FAT}, $REC_{PROTEIN}$, milk fat and lactose (%). The highest KMO values (> 0.7) were for the fat yield, milk pH, SCS, CF_P and t_{max} (Figure 1). Ten factors, explaining 74% of the original variance were kept for further analysis. The *varimax* rotated factor loadings (sorted by maximum variance explained) are shown in Table 2.

The first factor was heavily and positively loaded on $\%CY_{SOLIDS}$, fat and protein (%) and REC_{ENERGY}, thus it was considered as a factor underlying the latent concept of percentage of cheese yield (F1: %CY). The second factor was linked to all CF_t traits, but CF_P, and the REC_{FAT}, reflecting the curd firmness process (F2: CF_t). Due to positive loadings of the instant rate constants (k_{CF} and k_{SR}) and negative to the time traits (RCT_{eq} and t_{max}), the factor was ascribed to a favorable CF_t meaning. Moreover, this factor was favorably related to REC_{FAT}. The third factor was associated to the milk, fat and protein daily yields of individual cows, hence considered as a descriptor of the milk production (F3: Yield). These three factors were almost equally important and together explained 38% of total variability of the 26 milk traits considered. The following 7 factors were less important, explaining each one between 7 to 4% of total variability. The fourth factor was

heavily, but negatively, associated to β-LG, the quantitatively most important and variable whey protein fraction, and positively to other N compounds, representing a proxy of casein number, and being positively linked also to $REC_{PROTEIN}$. Hence this factor was representative of the N found in the cheese (F4: Cheese N). The fifth factor was primary linked to as_1-CN (positively) and then to the β-CN (negatively) and was considered as representative of as_1- and β-CN importance in milk (F5: as_1-β-CN). The sixth factor was positively associated with lactose and negatively with the SCS and the other N compounds. This factor was considered as indicator of the udder health status of the cow (F6: Udder health). The seventh factor was strongly and positively associated to the κ-CN and negatively, with a weaker loading, with the β-CN (F7: κ-β-CN). Thus, an increase of this factor was associated to an increased importance of κ-CN. The last 3 Fs were 1 trait-1 factor associations and were named according to the phenotype they were linked to as F8: as_2-CN, F9: as_1-CN-Ph and F10: α-LA, respectively.

Sources of Variation

Table 3 summarizes the results of the analysis of variance. The dairy system strongly affected F3: Yield and F1: %CY and had also an effect on udder health. The parity of the cow affected four Fs: F3: Yield, F2: CF_t, F6: Udder health and also F9: as1-CN-Ph. The days in milk influenced all factors but F4: Cheese N, F9: as1-CN-Ph and F7: κ-β-CN.

The F3: Yield and F6: Udder health were the only Fs influenced by all effects in the model. On the contrary, none of the effects that the model accounted for affected F4: Cheese N.

Investigating the effect of the dairy system, Table 4 outlines the least squares means (LSM) of the ten factors within the 4 different dairy systems and the orthogonal contrasts. Our analysis noted major differences between traditional and modern farms, in favor of the second, for F3:

Yield and F1: %CY, and a smaller effect, in favor of the traditional farms, for F6: Udder health. Moreover, for F3: Yield, F1: %CY and F8 as_2-CN, the modern farms using TMR showed greater results than those not using TMR. No difference was found in relation to the source of moisture in the TMR.

The LSM of the Fs on which the DIM had an effect are presented in Figure 2. A "lactation like" pattern (with the peak value in the second month of lactation) was observed for F6: Udder health, and also for F2: α-LA (Figure 2e, g), revealing, after lactation peak, a decrease in lactose, and an increase of SCS. Lactose was the trait on which the factor F6: Udder health was primary loaded. On the contrary, the latent variable F3: Yield was almost linearly decreased during lactation (Figure 2c). This pattern is primarily due to daily yield of fat and protein, and not to fresh milk (that peaked in the second month). The strong decrease at the beginning of lactation of milk fat and protein content (and also of their daily yield), paralleled by REC_{SOLIDS}, is reflected by the inverse "lactation like" pattern of the F1: %CY.

A decrease from 1st to 3rd class of DIM was observed for F2: CF_t followed by a stabilization and a slight increase at the end of the lactation (Figure 2b). An opposite pattern was observed between F8: as_2-CN and F5: as_1-β-CN (Figure 2d, f). The latent variable F8: as_2-CN was increasing rapidly at the beginning of lactation when F5: as_1-β-CN decreased. A more stable situation was observed thereafter (Figure 2f).

Figure 3 presents the LSM across the parity levels for the four Fs affected by parity. Factors F9: as_1-CN-Ph and F3: Yield had a similar pattern with an increase from the 1st to the 3rd parity, and a slight decline from the 4th parity (Figure 3 b, d). The F2: CF_t showed almost no difference between 1st and 3rd parities but an increase after (Figure 3a). A linear decrease was observed for the factor F6: Udder health related to the mammary gland health status (Figure 3c).

Genetic Parameters

Mean (SD) of marginal posterior densities of additive genetic, herd/test-date, and residual variances, across- and intra-herd heritabilities (h^2_{AH} and h^2_{IH}, respectively) and herd contribution for the investigated traits of the 10 Fs are summarized in Table 5. Highest heritability estimate was observed for F7: κ-β-CN (h^2_{AH} = 0.64 and h^2_{IH} = 0.72), F4:Cheese N (h^2_{AH} = 0.35 and h^2_{IH} = 0.62) and F5: as_1-β-CN (h^2_{AH} = 0.41 and h^2_{IH} = 0.57). The lowest values were found for F10: α-LA (h^2_{AH} = 0.08 and h^2_{IH} = 0.15) and F3: Yield (h^2_{AH} = 0.05 and h^2_{IH} = 0.11). The rest of the Fs had moderate values, ranging, for h^2_{AH}, between 0.13-0.25 and from 0.20 to 0.40 for h^2_{IH}. Moreover, the variation of incidence of variance due to herds within dairy system on total variance was large among Fs: the less affected were F7: κ-β-CN, F2: CF_t and F1: %CY, and the more affected were F3: Yield, F10: α-LA, and F9: as_1-CN-Ph.

Additive genetic correlations among the ten Fs are given in Table 6. All the estimates of each genetic correlation, were characterized by a large variability: of the 45 correlations evaluated only 12 presented a distribution of estimates having less than 10% of values with a sign opposite than the mean value (in bold in Table 6). In particular, F9: as_1-CN-Ph was negatively correlated with F4: Cheese N; F3: Yield negatively with F9: as_1-CN-Ph; F7: κ-β-CN positively with F4: Cheese N; F8: as_2-CN positively with F8: as_2-CN, F9: as_1-CN-Ph, F3: Yield, and F7: κ-β-CN; F2: CF_t positively with F1: %CY and negatively with F5: as_1-β-CN, and, lastly, F6: Udder health was negatively related with F10: α-LA and F2: CF_t and positively with F8: as_2-CN.

DISCUSSION

Selection decision in breeding programs relies on a variety of recorded traits related to production, health, reproduction, etc., which in some cases might exceed the 40 regularly recorded phenotypes (Miglior et al., 2005; Banos, 2010). Moreover, breeding aims in improving several traits simultaneously and not one at a time. Recent technological advances such as infrared spectostropy (Ferragina et al., 2015) together with advances in statistics, for e.g. better prediction models, are very likely to increase the number of available phenotypes in the nearest future, and thereby, the number of traits included in breeding goals (Boichard and Brochard, 2012). Therefore, the question on how to handle all the available information, in an efficient but easy way, for selection decision needs to be addressed. For instance, Ceron-Rojas and colleagues (2008) proposed eigen analysis related methods of the phenotypic covariance (correlation) matrix to construct selection indices in plant breeding (Cerón-Rojas et al., 2006; Cerón-Rojas et al., 2008). Close to eigen analysis, factor analysis is a multivariate approach that offers the potential to explore complex correlation structures and to transform them into a simpler space. Multivariate factor analysis has been previously used in dairy cattle giving encouraging results, albeit with smaller datasets (Macciotta et al, 2012; Mele et al, 2016). The objective in our study was to test the performance of MFA in bigger datasets (i.e., > 1,000 samples and > 20 traits) emphasizing on important economic traits for the dairy industry associated to the cheese production. Sources of variation related to the Fs as well as their genetic parameters (heritability and genetic correlations) were investigated.

Interpretation of Latent Variables

Ten orthogonal latent variables were extracted, out of 26 milk (quantity and quality), milk

coagulation and curd firmness indicators, and individual cheese yield traits. The Fs were explaining 74% of the original variability while at the same time drastically reduced the data space. Although the average KMO was not high, it was quite close to the value reported by Manca et al (2016) (0.57) using a similar, but smaller, dataset of 12 milk composition, MCP and udder health phenotypes in dairy sheep. Furthermore, the factor model was able to capture basic concepts of the *cheese-making* process. More precisely, the first four Fs, sorted by variance explained, were able to capture the underlying structure of the cheese yield (%), the curd firmness process, the milk yield and the presence of N into the cheese. Moreover, 4 Fs were associated to the basic milk caseins (as_1-β-CN, κ-β-CN, as_2-CN and as_1-CN-Ph) and 1 factor was related with the, quantitatively, most important 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.

The meaning of some of the Fs is comparable to previous studies that used MFA, albeit with smaller datasets. For e.g., in an experimental design study assessing the effect of DIM, body condition and milk protein genotypes on the MCP and the protein composition in Danish Holstein (n = 39) 8 Fs were extracted out of 28 measured variables (Ostersen et al., 1997). Among the measured phenotypes were milk yield traits (milk, fat, protein and lactose), whey and non- casein contents, proportion of caseins, proportion of alleles in genotyped milk proteins, MCP and energy balance and the body condition. The 8 Fs obtained were representing MCP properties of milk, milk protein genotypes, the energy and the body condition status of the cows, and effects of lactation on milk yield and protein composition. In another study, eleven milk composition, MCP and udder health phenotypes measured in Brown Swiss cows (n = 1,200) were substituted by 4 Fs, describing milk composition, coagulation, acidity and the udder health status (Macciotta et al, 2012). Moving on small ruminants, Todaro et al (2005) replaced 11 traits related to milk composition, MCP and

udder health measured on 117 Girgentana goats with 3 Fs named as “slow milks”, “milk yield” and “curd firmness”. Recently, Vacca and colleagues (2016), by analyzing 9 milk yield, composition and hygiene traits in 1,050 Sardinian goats obtained 4 Fs (quality, hygiene, production, and acidity), while Manca et al. (2016) in a study on 991 Sarda ewes, extracted 4 Fs (composition and cheese yield, udder health status, coagulation and curd characteristics) from a total of 12 measured phenotypes. All the 5 above mentioned studies used an estimatory MFA with a *varimax* rotation. The proportion of variance explained by the factor model in those studies ranged between 51.2% (Todaro et al., 2005) to 97 % in Ostersen et al (1997). Part of this difference, however, could be attributed to the different ways and methods of measuring the phenotypes used in MFA.

Although a direct, factor by factor, comparison between our results and the above mentioned studies is risky to be made due to differences in the measured phenotypes analyzed (number and type), a consistent pattern can be observed: generally, in all the studies MFA clearly distinguishes between milk quality, production, coagulation and health related concepts. These results are encouraging and confirm the ability of MFA to capture basic underlying structure of correlated variables.

Factor F1: %CY was positively related to fat and protein percentage as well as the $\%CY_{SOLIDS}$ and the REC_{ENERGY}. This is in agreement to phenotypic and genetic correlations that have been reported in the literature for these traits (Bittante et al., 2013a). The inverse relationship of RCT_{eq} and t_{max} with k_{CF}, k_{SR} and REC_{FAT} found on F2: CF has been previously pointed out in Cecchinato and Bittante (2016). On factor F5: as_1-β-CN, the as_1- and β- CN were oppositely related. The antagonistic nature of those two caseins has been previously suggested by Bobe et al. (1999) and Bonfatti et al. (2010). Moreover, the factor F6: Udder health was inversely related to

other nitrogen compounds that are present in the milk. It is worth noting that within this nitrogen fraction, urea was also included. Urea is inversely related to lactose (Miglior et al., 2007), which was the trait stronger related to factor F6: Udder health.

To further assess the practical application in breeding programs of the new set of the latent variables, an investigation on the sources affecting variation on those new traits was followed together with estimation of the genetic parameters.

Sources of Variation of the Latent Variables

Effects of Dairy System on the Extracted Factor Scores. Major differences between the different dairy systems were identified for 2 Fs, namely F3: Yield and F1: %CY. Modern farms were associated to a higher milk and solids yield and percentage of cheese. Moreover, within the modern farms, the use of TMR in the feeding system was found beneficial for both latent variables. Only Vacca et al. (2016) on lactating goats studied the relationships between Fs and dairy system and farms characteristics: they did not found any associations between Fs and dairy system, but observed an association between altitude of the farm and the production factor and between flock size and the acidity factor.

Nonetheless, the results of our study are in agreement with previous findings on dairy cattle carried out on the measured daily milk yield production traits instead of Fs (Sturaro et al., 2009; Sturaro et al., 2013; Bittante et al., 2015). Moreover, although individual cheese traits were not included in the analysis of Bittante et al (2015), major differences between the dairy systems were reported for milk fat and protein percentages. These two traits were both loaded on F1: %CY in our study. No significant differences between the dairy systems were observed for the coagulation properties of milk (F2: CF_t). The F2: CF_t was primary loaded to the k_{CF}, but also to RCT_{eq}, k_{SR},

t_{max} and REC_{FAT}. These findings are consistent with Bittante et al (2015), where no significant differences were noted among the dairy systems for k_{CF} and RCT_{eq}. A significant effect of the dairy systems existed in that study for k_{SR} ($P < 0.05$) and t_{max} ($P < 0.01$). However, since a factor is a mixture of phenotypes, and in our case F2: CF_t was dominated by the k_{CF}, this effect was diluted.

Effect of Stage of Lactation and Parity on the Extracted Factor Scores. The stage of lactation significantly influenced most of the latent variables. As expected, the F3: Yield was linearly decreased during lactation, in agreement to Bittante and colleagues (2015). The absence of the peak of lactation is probably an artifact due to the length of the DIM classes. A tendency towards worsening of milk coagulation during lactation is known. This pattern has been reported in Macciotta et al (2012) using a factor (based on the traditional MCP values) as indicator of milk coagulation. Moreover, the same trend has been observed in previous studies using the traditional curd-firmness value at 30 minutes after the beginning of coagulation, generally known as a_{30} (Ikonen et al., 2004; Cipolat-Gotet et al., 2012). Being consistent, the coagulation ability (F2: CF_t) was smoothly decreased (worsening) during lactation, stabilized between 4th and 8th class of DIM, with an evidence of recovery at the end of the lactation. Moreover, the factor related to the percentage of cheese yield (F1: %CY) showed a decrease from the first to the second month of lactation and then a linear increase. Not surprisingly, this pattern during lactation is consistent with the trend of $\%CY_{SOLIDS}$ on which this factor was positively and strongly related (Cipolat-Gotet et al., 2013).

In line to our expectations, the F10: α-LA (α-LA is a whey protein) was decreased during lactation having an opposite trend compared to the F1: %CY. Same pattern has been reported for the measured α-LA in Ostersen et al. (1997). A smooth decrease of F5: as_1-β CN during lactation

and the inverse pattern of F8: as_2-CN was observed, as expected.

The pattern of the health status of the mammary gland mainly related to the lactose (F6: Udder health) was worsened within lactation. As it is known, lactose in milk is inversely related to mastitis or somatic cell count (Shuster et al., 1991) and SCS has an opposite trend within lactation compared to milk yield (Walsh et al., 2007). Similar results have been reported using MFA by Macciotta et al (2012) in dairy cattle and in Manca et al (2016) in dairy sheep. Also in those studies, lactose was the most related trait to the factor underlying the health status of the mammary gland.

The trend of parity effect on F3: Yield was similar to previous results on milk yield traits. More precisely, Bittante et al (2015) showed a significant increase in milk yield from the first to the second parity. Moreover, in the same study the k_{CF} values were lower in first lactation cows compared to following lactations, similar to the effect of parity on the F2: CF_t found in our analysis. Furthermore, a decline of F6: Udder health with the increase of parity is in agreement to the literature where a decrease in milk lactose percentage has been observed in multiparous cows (Yang et al., 2013). Parity also affected F9: as_1-CN-Ph latent variable in our analysis with a pattern among different levels of parities mimicking the trend of F3: Yield.

Genetic Parameters of the Latent Variables

Variance Components and Heritability of the Factor Scores. Previous studies working on MFA with dairy traits were mainly focused on investigating the sources of variation of the latent variables. Hence, in the absence of bibliographic sources, our results on the genetic analysis of the Fs were mainly compared to genetic parameters of the measured single traits in previous studies. The high heritability estimates of F7: κ-β-CN found in our study were in accordance with h^2 values

of κ-CN (h^2_{AH} = 0.63) and β-CN (h^2_{AH} = 0.69) reported in Bonfatti et al. (2011). Moreover, for the 3 Fs that were loading to only 1 trait, i.e. F8: as_2-CN, F9: as_1-CN-Ph and F10: α-LA, heritability estimates were in range with those documented in the literature of the single traits, i.e. h^2_{AH} = 0.11 for α-LA content, h^2_{AH} =0.22 for as_2-CN were reported in Simmental cows (Bonfatti et al., 2011) and intra-herd heritability of 0.13-0.15 for as_1-CN-Ph (with 8 phosphorylated serine residues) in Danish Holstein (Gebreyesus et al., 2016). The low heritability for milk yield (~0.09) that has been previously reported in Brown Swiss cattle (Macciotta et al., 2012; Bittante et al., 2013a) can explain the low heritability found in our study for F3: Yield, a factor that was dominated by milk and protein yields. Factor F1: %CY, that was heavily loaded to $\%CY_{SOLIDS}$, had similar heritability with the single trait $\%CY_{SOLIDS}$ (h^2_{AH} = 0.21 and h^2_{IH} = 0.26) as reported in Bittante et al. (2013a). The F6: Udder health had an across-herd h^2 of 0.13. Similar value of heritability (0.14) has been reported in Macciotta et al (2012) in Brown Swiss cattle for a factor entitled "udder health" (loading on somatic cell count and lactose). In the same study, another factor representing the MCP, but including the traditional MCP traits (RCT, k_{20} and a_{30}) instead of the CF_t values, had a heritability estimate of 0.23. This value is in agreement to heritability estimates (h^2_{AH} = 0.24) that found for F2: CF_t in our study. Concerning the F4: Cheese N (h^2_{AH} = 0.35 and h^2_{IH} = 0.62), comparable values of heritability can be found in the literature (h^2_{AH} = 0.29) for a trait representing the total cheese protein yield (g/l) (Bonfatti et al., 2011). Moreover, F4: Cheese N was also loaded to $REC_{PROTEIN}$. In the literature, heritability estimates of 0.35 (across-herd) and 0.49 (intra-herd) have been reported for the $REC_{PROTEIN}$ (Bittante et al., 2013a).

Additive Genetic Relationships among Factor Scores. Although the Fs were extracted in an orthogonal way, meaning that they were phenotypically uncorrelated, bivariate genetic analysis depicted considerable genetic correlations. The highest genetic correlations > |0.6| were observed

between F10: α-LA and F6: Udder health and F9: as_1-CN-Ph and F3: Yield. The known inverse (medium to strong) genetic relationship between milk yield and the casein percentage (Ikonen et al., 2004) explains the high and negative genetic correlation between F9: as_1-CN-Ph and F3: Yield. Moreover, F9: as_1-CN-Ph was positively correlated with F5: as_1-β CN as expected, since F5: as_1-β CN was mainly loaded on as_1-CN and the as_1-CN-Ph is a posttranslational form of the as_1-CN. In addition, this estimate is comparable to the genetic correlation (~0.80) between as_1-CN and as_1-CN-Ph (with 8 phosphorylated serine residues) reported in Gebreyesus et al. (2016), although in that study the protein fractions were expressed as ratios to the total milk protein %. Moreover, F5: as_1-β-CN was positively related to F7: κ-β-CN. Both Fs contained β-CN as a minor loading, hence the value of 0.38 is comparable to the genetic correlation of 0.12 between as_1- and κ-CN reported in Bonfatti et al. (2011). The positive weak genetic correlation between F5: as_1-β-CN (primary loaded on as_1-CN) and F8: as_2-CN is also in line with Bonfatti and colleagues (2011) that reported genetic correlations between as_1- and as_2-CN of 0.22. A negative-medium r_g was found between F5: as_1-β-CN and F2: CF_t. As has been outlined in Jõudu et al. (2008) a negative phenotypic relationship exists between the relative content of as_1- and β-CN with a firmer curd.

The F6: Udder health was negatively correlated with F10: α-LA and F2: CF (on favor of good coagulation) while positively related to F8: as_2-CN. This means that a healthy mammary gland status is associated with an increase in F8: as_2-CN and an increase of F10: α-LA and F2: CF_t. Indeed, a negative correlation between α-LA and somatic cell count has been reported by Caffin et al. (1985), while a good cow health status results in an increase of caseins. The negative genetic correlation with the F2: CF_t is not clear. However, recently a negative effect of very low SCS to some milk technological traits has been pointed out by Bobbo et al. (2016). More precisely, authors reported a slower coagulation with low values of SCS. Interestingly, F2: CF_t was heavily

loaded to the rate of curd firmness.

The F4: Cheese N was genetically correlated with F9: as_1-CN-Ph (negatively) and F7: κ-β-CN (positively). A competitive synthesis between as_1-CN and β-CN has been previously reported (Bobe et al., 1999; Bonfatti et al., 2010b), while the favorable effect of β- and κ-CN on cheese yield is widely known (Walstra et al., 2014). This also explains the reason that although the F3: Yield is negatively correlated to F9: as_1-CN-Ph it had a positive correlation to F5: as_1-β-CN. Finally, the positive correlation between F1: %CY and F2: CF_t implies that selection on better CF_t characteristics will lead to an increase in the percentage of cheese yield.

CONCLUSION

Our data set comprised of twenty-six variables of economical importance for the dairy industry that were measured in more than 1,000 dairy cows. Using multivariate factor analysis, ten latent variables replaced the original phenotypes, capturing important concepts of the "*cheese-making*" process as well as basic bovine health indicators related to the mammary gland. Results from genetic and ANOVA analyses of the 10 Fs were in agreement to the given name of the factor and reflected the underlying structure that each factor was representing. Moreover, heritability and genetic correlation estimates of the Fs were in favor of the use of Fs for breeding purposes.

Nonetheless, for a successful application in breeding companies further important steps needs to be taken. The use, and potential benefit, of factors in terms of response to selection and economic value in breeding is missing in the literature and should be investigated. Furthermore, being in the genomic era and with a plethora of available phenotypes we propose for future research to test the behavior of latent variables in genome-wide association, and its complement pathway based analysis, or prediction studies.

ACKNOWLEDGMENTS

The authors wish to thank the Trento Province (Italy), the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy) and the Superbrown Consortium of Bolzano and Trento for financial and technical support. C. Dadousis benefitted from financial support of the CARIPARO (Cassa di Risparmio di Padova e Rovigo) Foundation (Padua, Italy).

Table 1. Summary statistics of milk (yield and quality), protein fractions, coagulation (curd firming) and cheese-making (%CY and REC) traits.

Trait[1]	Mean	P1	P99	CV, %
Milk traits				
$Milk_{yield}$, kg/day	24.95	9.30	45.88	31
Fat_{yield}, kg/day	1.09	0.37	2.41	37
$Protein_{yield}$, kg/day	0.92	0.36	1.57	30
Fat, %	4.37	2.48	7.40	20
Protein, %	3.71	2.93	4.68	11
Lactose, %	4.86	4.31	5.22	4
pH	6.64	6.43	6.85	1
SCS, units	2.87	-0.47	7.75	65
Milk protein fractions, %				
αs_1-CN	25.69	21.86	29.79	7
αs_1-CN-Ph	1.45	0.19	3.04	42
αs_2-CN	9.20	6.89	12.60	12
β-CN	32.26	27.38	38.67	8
κ-CN	9.44	4.69	12.17	16
β-LG	8.68	4.63	13.07	18
a-LA	2.39	1.14	3.61	21
Other N compounds	10.89	5.30	15.93	21
Curd Firming				
RCT_{eq}, min	20.96	10.66	40.89	29
CF_P, mm	49.20	25.33	72.28	20
k_{CF}, % × min^{-1}	12.90	4.06	23.70	32
k_{SR}, % × min^{-1}	1.23	0.31	2.60	37
CF_{max}, mm	36.91	18.90	53.94	20
t_{max}, min	41.83	24.00	87.00	30
Cheese yield (%CY)				
$\%CY_{CURD}$	14.95	11.00	19.41	12
$\%CY_{SOLIDS}$	7.17	5.37	9.68	13
$\%CY_{WATER}$	7.77	5.04	11.11	16
Nutrient Recovery (REC, %)				
REC_{SOLIDS}	51.80	43.85	60.27	7
REC_{FAT}	89.75	78.44	95.90	4
$REC_{PROTEIN}$	78.16	72.41	83.44	3
REC_{ENERGY}	67.15	58.92	75.07	5

[1]SCS = $\log_2$ (SCC × 100,000) + 3. Milk protein fractions: CN = casein; LA = lactalbumin and LG = lactoglobulin. Curd firming: RCTeq = estimated RCT; 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 dry matter ($\%CY_{SOLIDS}$) and curd water ($\%CY_{WATER}$) versus the weight of the processed milk (g); REC = ratio of the weight (g) of the curd constituent (dry matter, fat, protein or energy, respectively) versus that of the same constituent in the processed milk (g).

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

Trait[2]	F1 %CY	F2 MCP	F3 Yield	F4 Cheese N	F5 as_1-β-CN	F6 Udder health	F7 κ-β-CN	F8 as_2-CN	F9 as_1-CN-Ph	F10 α-LA	com
Milk traits											
$Milk_{yield}$, kg/day	-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/day	0.28	0.08	**0.89**	-0.02	0.06	0.17	-0.01	-0.01	-0.01	0.07	0.92
$Prot_{yield}$, kg/day	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
Prot, %	**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, units	0.06	-0.02	-0.08	0.04	-0.05	**-0.41**	0.09	0.01	0.03	-0.09	0.20
Milk protein fractions, %											
as_1-CN	0.04	0.01	0.07	-0.14	**0.94**	0.25	-0.04	-0.08	-0.14	-0.04	0.99
as_1-CN-Ph	0.04	-0.02	0.00	0.09	-0.10	0.01	-0.04	-0.04	**0.98**	0.06	1.00
as_2-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 comp.	-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^{-1}	-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^{-1}	-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	0.70	**0.74**	

[1]Factors have been sorted based on proportion of variance explained. F1: %CY = Factor related to the percentage of individual cheese yield; F2: CF_t = 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: as_1-β-CN = Factor related to the as_1- 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: κ-β-CN = Factor related to the κ- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; F8: as_2-CN = Factor related to the milk as_2-CN, expressed as relative content to the total milk nitrogen; F9: as_1-CN-Ph = Factor related to the milk as_1-CN-Ph expressed as content to the total milk nitrogen; F10: α-LA = Factor related to the milk α-LA.

[2]SCS = $\log_2$ (SCC × 100,000) + 3. Milk protein fractions: CN = casein; LA = lactalbumin and LG = lactoglobulin. Curd firming: RCTeq = estimated RCT; CF_P = asymptotical potential value of CF; k_{CF} = curd-firming instant rate constant; k_{SR} = syneresis instant rate constant; and t_{max} = time at achievement of CF_{max}. %CY = ratios of the weight (g) of the curd dry matter ($\%CY_{SOLIDS}$) and curd water ($\%CY_{WATER}$) versus the weight of the processed milk (g); REC = ratio of the weight (g) of the curd constituent (fat, protein or energy, respectively) versus that of the same constituent in the processed milk (g).

Table 3. Analysis of variance (*F*-values and significance) of the10 extracted factors[1].

Item	Dairy system	Parity	DIM	Residual
	F-value	*F-value*	*F-value*	RMSE
DF	3	3	10	-
F1: %CY	6.25***	2.41	13.34***	0.83
F2: CF_t	1.78	7.38***	5.96***	0.84
F3: Yield	19.54***	46.37***	30.42***	0.59
F4: Cheese N	1.18	0.75	1.74	0.76
F5: as_1-β-CN	1.43	0.74	7.47***	0.84
F6: Udder health	3.39*	32.88***	14.40***	0.71
F7: κ-β-CN	0.81	0.23	1.28	0.93
F8: as_2-CN	1.68	1.62	4.16***	0.78
F9: as_1-CN-Ph	0.48	3.60*	1.14	0.73
F10: α-LA	1.71	1.60	11.09***	0.72

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

[1]F1: %CY = Factor related to the percentage of individual cheese yield; F2: CF_t = 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: as_1-β-CN = Factor related to the as_1- 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: κ-β-CN = Factor related to the κ- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; F8: as_2-CN = Factor related to the milk as_2-CN, expressed as relative content to the total milk nitrogen; F9: as_1-CN-Ph = Factor related to the milk as_1-CN-Ph expressed as content to the total milk nitrogen; F10: α-LA = Factor related to the milk α-LA.

Table 4. Effects of the dairy system (traditional with tied cows, modern with loose cows), the use of total mixed ration (TMR) within modern farms, and of the moisture source of TMR on the 10 extracted latent factors[1].

	Dairy system LSM:				Orthogonal contrasts *F-values*		
Dairy system:	Traditional	Modern					
Feed distribution:	-	No TMR	TMR		Modern *vs* Traditional[2]	TMR *vs* No TMR[3]	Silage *vs* Water[4]
Moisture source:	-	-	Silage	Water			
Herds	29	30	9	17	-	-	-
F1: %CY	-0.25	-0.11	0.40	0.18	13.21***	10.22**	1.37
F2: CF_t	0.15	-0.13	0.02	0.04	2.48	1.50	0.01
F3: Yield	-0.53	0.16	0.59	0.60	53.31***	8.38**	0.12
F4: Cheese N	0.19	-0.02	-0.26	-0.04	3.29	0.52	0.63
F5: as_1-β-CN	0.21	-0.002	0.02	-0.12	3.34	0.10	0.42
F6: Udder health	0.28	-0.25	-0.18	0.05	5.98*	1.28	0.90
F7: κ-β-CN	-0.002	-0.08	0.19	-0.01	0.10	1.75	1.17
F8: as_2-CN	-0.12	-0.21	0.22	0.14	1.22	4.66*	0.09
F9: as_1-CN-Ph	0.08	0.02	-0.24	0.08	0.52	0.23	1.14
F10: α-LA	-0.12	-0.11	0.35	0.20	2.43	3.95	0.26

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

F1: %CY = Factor related to the percentage of individual cheese yield; F2: CF_t = 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: as_1-β-CN = Factor related to the as_1- 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: κ-β-CN = Factor related to the κ- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; F8: as_2-CN = Factor related to the milk as_2-CN, expressed as relative content to the total milk nitrogen; F9: as_1-CN-Ph = Factor related to the milk as_1-CN-Ph expressed as content to the total milk nitrogen; F10: α-LA = Factor related to the milk α-LA.

[2]Contrast between the "Traditional" dairy system vs the three "Modern" ones.

[3]Contrast between the "Modern No TMR" dairy system vs the two "Modern TMR" ones.

[4]Contrast between the "Modern TMR Silage" dairy system vs the "Modern TMR Water" one.

Table 5. Mean (SD) of marginal posterior densities of additive genetic ($\boldsymbol{\sigma^2_A}$), herd/test-date ($\boldsymbol{\sigma^2_H}$), and residual ($\boldsymbol{\sigma^2_E}$), variances, across-herd ($\boldsymbol{h^2_{AH}}$), and intra-herd ($\boldsymbol{h^2_{IH}}$) heritabilities and herd contribution for the investigated traits.

Item[1]	σ^2_a	σ^2_h	σ^2_e	h^2_{AH}	h^2_{IH}	Herd[2], %
F1: %CY	$0.186_{(0.07)}$	$0.204_{(0.04)}$	$0.502_{(0.06)}$	$0.207_{(0.07)}$	$0.268_{(0.09)}$	22
F2: CF_t	$0.217_{(0.07)}$	$0.174_{(0.03)}$	$0.513_{(0.06)}$	$0.239_{(0.08)}$	$0.295_{(0.09)}$	19
F3: Yield	$0.039_{(0.02)}$	$0.477_{(0.08)}$	$0.324_{(0.02)}$	$0.047_{(0.02)}$	$0.108_{(0.06)}$	56
F4: Cheese N	$0.339_{(0.06)}$	$0.419_{(0.07)}$	$0.207_{(0.05)}$	$0.352_{(0.07)}$	$0.618_{(0.10)}$	43
F5: as_1-β-CN	$0.416_{(0.08)}$	$0.207_{(0.05)}$	$0.315_{(0.07)}$	$0.414_{(0.08)}$	$0.565_{(0.11)}$	27
F6: Udder health	$0.108_{(0.05)}$	$0.329_{(0.06)}$	$0.418_{(0.04)}$	$0.126_{(0.06)}$	$0.204_{(0.09)}$	38
F7: κ-β-CN	$0.687_{(0.13)}$	$0.125_{(0.03)}$	$0.257_{(0.11)}$	$0.639_{(0.10)}$	$0.723_{(0.11)}$	11
F8: as_2-CN	$0.250_{(0.06)}$	$0.378_{(0.07)}$	$0.377_{(0.05)}$	$0.248_{(0.06)}$	$0.397_{(0.09)}$	37
F9: as_1-CN-Ph	$0.172_{(0.06)}$	$0.510_{(0.09)}$	$0.363_{(0.05)}$	$0.165_{(0.06)}$	$0.318_{(0.10)}$	48
F10: α-LA	$0.075_{(0.03)}$	$0.486_{(0.08)}$	$0.432_{(0.03)}$	$0.075_{(0.03)}$	$0.147_{(0.07)}$	49

[1]F1: %CY = Factor related to the percentage of individual cheese yield; F2: CF_t = 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: as_1-β-CN = Factor related to the as_1- 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: κ-β-CN = Factor related to the κ- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; F8: as_2-CN = Factor related to the milk as_2-CN, expressed as relative content to the total milk nitrogen; F9: as_1-CN-Ph = Factor related to the milk as_1-CN-Ph expressed as content to the total milk nitrogen; F10: α-LA = Factor related to the milk α-LA.

[2]The variance of herd/date within dairy system and season is expressed as ratio with the total phenotypic variance ($\sigma^2_a + \sigma^2_h + \sigma^2_e$).

Table 6. Additive genetic correlations for the 10 factors extracted from the bivariate analysis[1].

Item[2]	F1: %CY	F2: CF_t	F3: Yield	F4: Cheese N	F5: as_1-β-CN	F6: Udder health	F7: κ-β-CN	F8: as_2-CN	F9: as_1-CN-Ph	F10: α-LA
F1: %CY	-	**0.372(92)**	0.216(72)	0.066(62)	-0.237(84)	-0.322(81)	-0.012(51)	-0.198(80)	-0.110(66)	-0.222(74)
F2: CF_t		-	0.159(65)	0.164(79)	**-0.387(97)**	**-0.411(90)**	0.102(69)	-0.161(77)	-0.127(66)	-0.032(53)
F3: Yield			-	-0.022(51)	**0.423(91)**	0.370(81)	-0.123(63)	-0.161(70)	**-0.603(95)**	-0.361(78)
F4: Cheese N				-	0.094(71)	-0.126(68)	**0.228(94)**	0.184(86)	**-0.373(97)**	0.057(69)
F5: as_1-β-CN					-	-0.068(58)	**0.380(98)**	**0.444(98)**	**0.574(99)**	-0.211(77)
F6: Udder health						-	-0.078(60)	**0.436(95)**	0.187(68)	**-0.667(97)**
F7: κ-β-CN							-	0.095(71)	0.174(76)	-0.228(77)
F8: as_2-CN								-	-0.062(60)	0.181(70)
F9: as_1-CN-Ph									-	-0.102(62)
F10: α-LA										-

[1]Mean of the marginal posterior density of the additive genetic correlation; the posterior probability (%) for positive correlations greater than 0 or for negative correlations lower than 0 is given within parentheses. Boldface indicates additive genetic correlations with ≥ 90% of posterior probability accumulated above 0 (positive estimates) or below 0 (negative estimates).

[2]F1: %CY = Factor related to the percentage of individual cheese yield; F2: CF_t = 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: as_1-β-CN = Factor related to the as_1- 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: κ-β-CN = Factor related to the κ- and β-CN contents in milk, expressed as relative contents to the total milk nitrogen; F8: as_2-CN = Factor related to the milk as_2-CN, expressed as relative content to the total milk nitrogen; F9: as_1-CN-Ph = Factor related to the milk as_1-CN-Ph expressed as content to the total milk nitrogen; F10: α-LA = Factor related to the milk α-LA. Factors have been orthogonally extracted, hence all their pairwise phenotypic correlations are equal to zero.

Figure 1. Pearson (above the diagonal) and partial (under the diagonal) phenotypic correlations and among the 25 traits used in the factor analysis. On the diagonal the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy per trait.

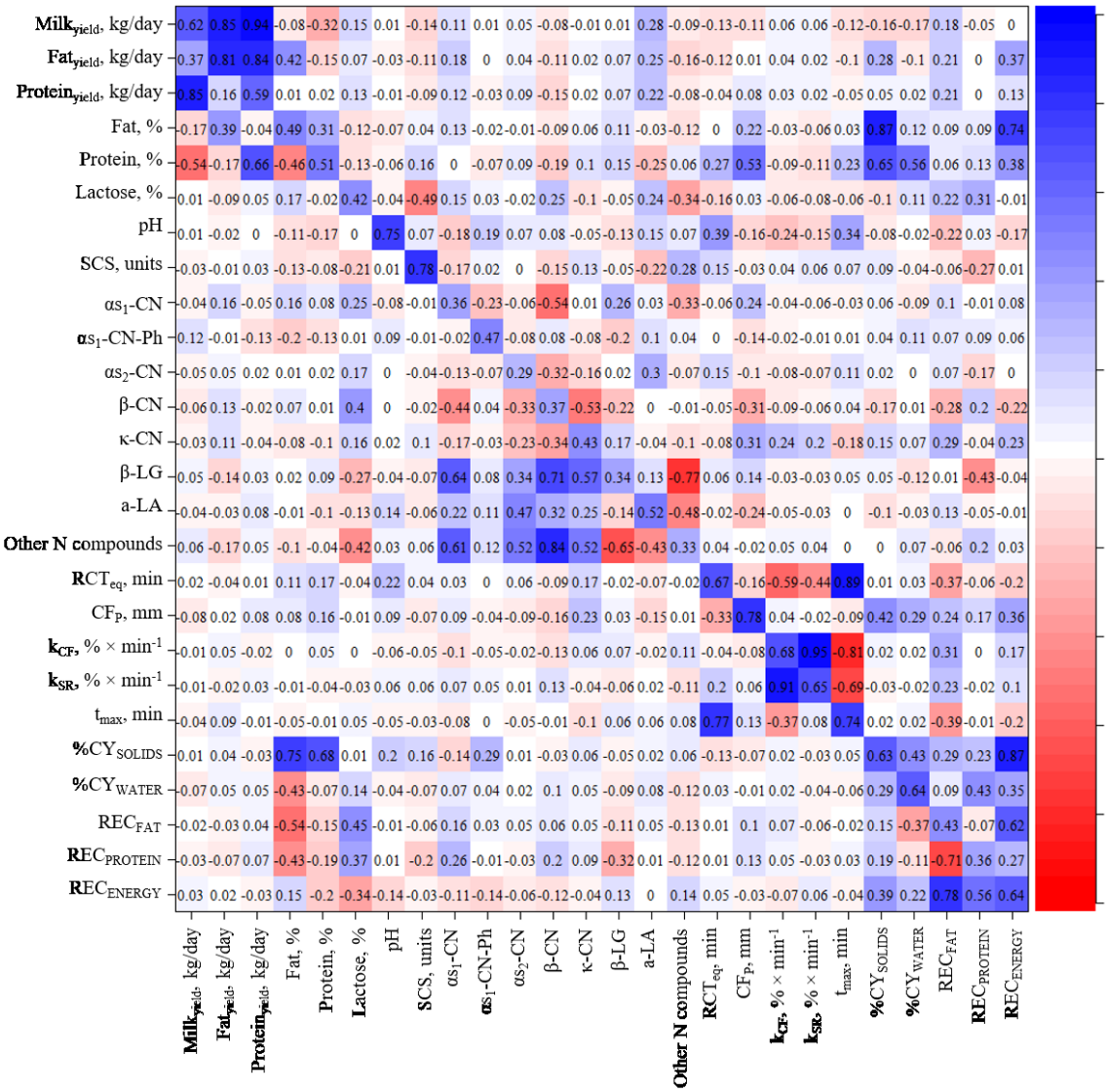
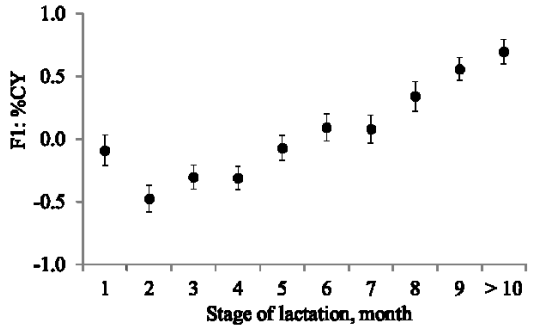


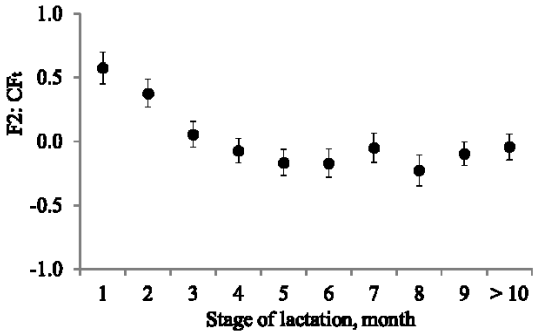
Figure 2 Least squares means of the seven factors across stage of lactation.

Description: F1: %CY = Factor related to the percentage of individual cheese yield; F2: CF_t = Factor related to the curd firmness; F3: Yield = Factor related to the milk yield; F5: as_1-β-CN = Factor related to the as_1- 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; F8: as_2-CN = Factor related to the milk as_2-CN, expressed as relative content to the total milk nitrogen F10: α-LA = Factor related to the milk α-LA.

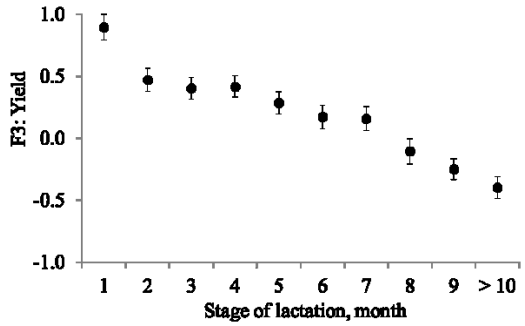
[a]


F1: %CY
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 > 10
Stage of lactation, month

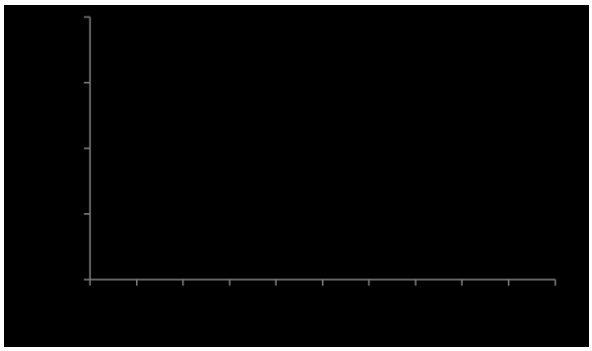
[b]


F2: CFt
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 > 10
Stage of lactation, month

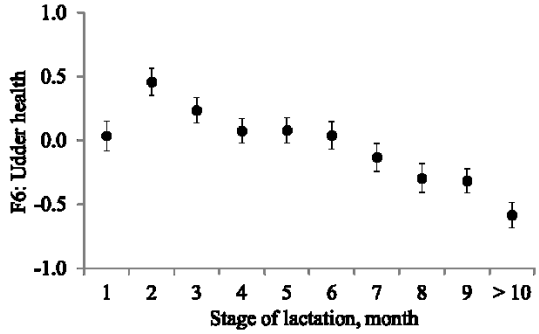
[c]


F3: Yield
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 > 10
Stage of lactation, month

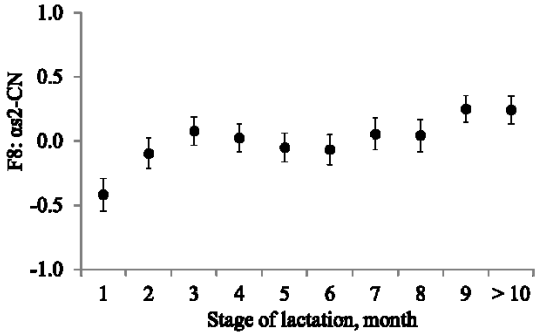
[d]



[e]


F6: Udder health
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 > 10
Stage of lactation, month

[f]


F8: αs2-CN
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 > 10
Stage of lactation, month

[g]

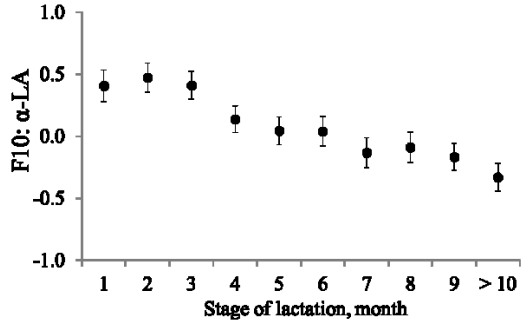
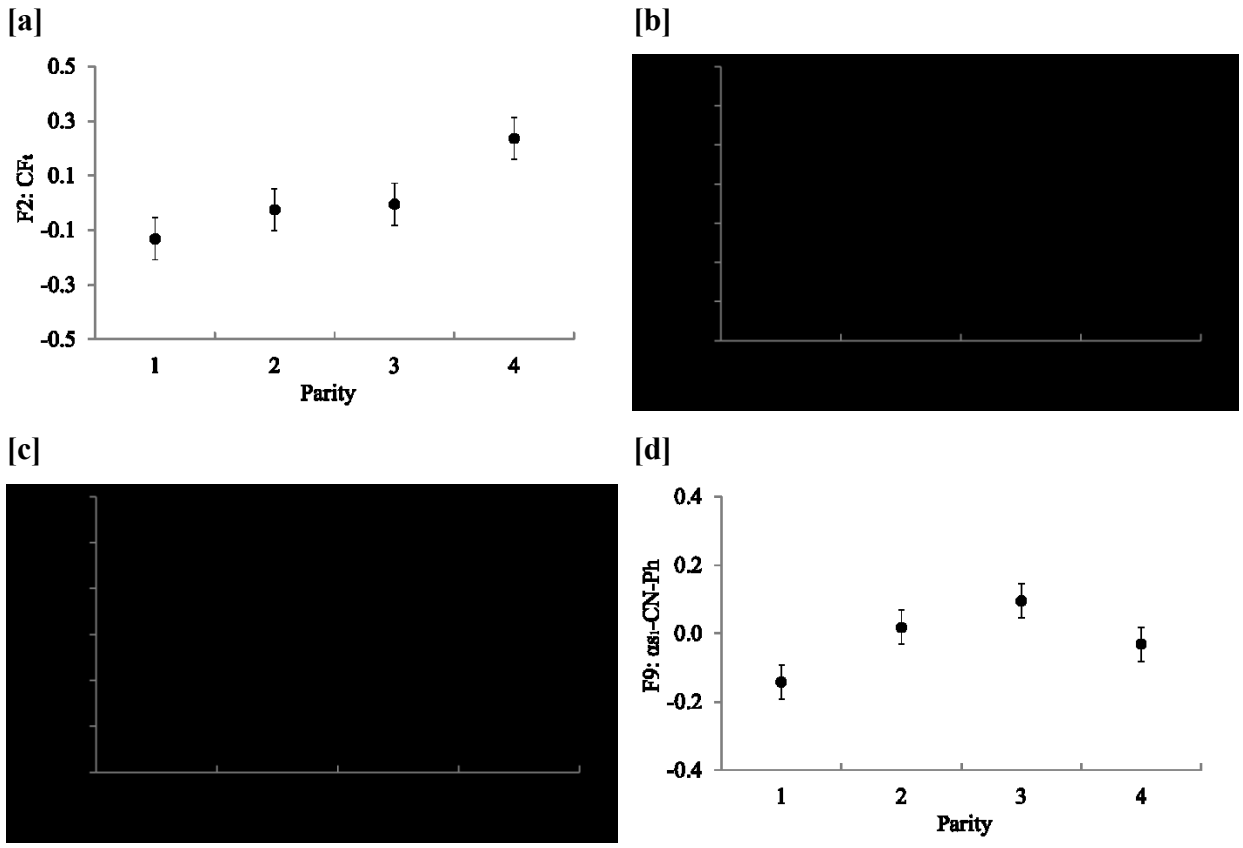

F10: α-LA
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 > 10
Stage of lactation, month

Figure 3 Least squares means of the four factors across the levels of parity.

Description: F2: CF_t = Factor related to the curd firmness; F3: Yield = Factor related to the milk yield; F6: Udder health = Factor related to the udder health of a cow; F9: as_1-CN-Ph = Factor related to the milk as_1-CN-Ph expressed as content to the total milk nitrogen.



CHAPTER 5

Genome-wide association and pathway-based analysis using latent variables related to milk yield and quality, protein composition and cheese-making traits in dairy cattle.

Christos Dadousis[1], Sara Pegolo[1], Giovanni Bittante[1], Alessio Cecchinato[1]

[1] Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

ABSTRACT

Genome wide associations (GWAS) and gene-set enrichment analyses were conducted with cheese-related latent variables (factors, Fs). Factor analysis (FA) was applied to identify latent structures of 26 traits related to bovine milk quantity and quality, protein fractions [αs_1-, αs_2-, β- and κ- casein (CN), β-lactoglobulin and α-lactalbumin (α-LA)], coagulation and curd firming at time t (CF_t) and individual cheese properties [cheese yield (%CY) and nutrient recovery in the curd]. Cows (n=1,152) were genotyped with the Illumina BovineSNP50 Bead Chip v.2. Single marker regression GWAS were fitted. Gene-set enrichment analysis was run on GWAS results, using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway databases, to reveal ontologies/pathways associated with the Fs.

Ten orthogonal Fs were extracted, explaining 74% of the original variability. $F1_{\%CY}$ underlined the %CY, $F2_{CFt}$ was related to the CF_t process, $F3_{Yield}$ considered descriptor of milk and solids yield while $F4_{Cheese\ N}$ described the presence of nitrogen (N) into the cheese. Four factors were related to the milk caseins ($F5_{\alpha s_1\text{-}\beta\text{-}CN}$, $F7_{\beta\text{-}\kappa\text{-}CN}$, $F8_{\alpha s_2\text{-}CN}$, and $F9_{\alpha s_1\text{-}CN\text{-}Ph}$) and 1 factor was linked to the whey protein ($F10_{\alpha\text{-}LA}$). One factor underlined the udder health status ($F6_{Udder\ health}$). All Fs, but $F5_{Yield}$, showed significance ($P < 5 \times 10^{-5}$) in GWAS. Signals in 10 *Bos taurus autosomes* (BTA) were detected. High peaks on BTA6 (~87Mbp) were found for $F6_{\beta\text{-}\kappa\text{-}CN}$, $F5_{\alpha s_1\text{-}\beta\text{-}CN}$ and at the tail of BTA11 (~104Mbp) for $F1_{Cheese\ N}$. After false discovery rate correction (at 5%), 33 GO terms and 6 KEGG categories were mainly enriched for $F8_{\alpha s_2\text{-}CN}$, but also for $F1_{\%CY}$, $F4_{Cheese\ N}$, and $F10_{\alpha\text{-}LA}$, including terms related to ion transport and homeostasis, neuron function/part, tight junction and GnRH signaling pathway. Results support the potential application of Fs in dairy cattle genomic studies. Pathway analysis indicated a key role of αs_2-CN.

Key words: Factor analysis, milk protein, coagulation, cheese yield, GWAS, gene-set enrichment

INTRODUCTION

Cheese production is the main target of bovine milk in many countries worldwide, while recent studies revealed an important role of animal genetics in bovine cheese yield (Bittante et al., 2013a). Nevertheless, cheese manufacture is a complicated process with many interrelated factors involved, such as milk components (e.g. fat, protein, minerals), milk acidity and microbial flora, coagulation properties (MCP), etc. In addition, animal breeding aims in the simultaneous improvement of several traits in consequent generations, including a wide range of phenotypes related with production, reproduction and health, etc., that might exceed the 40 regularly recorded phenotypes (Miglior et al., 2005; Banos, 2010). However, the large number of traits of interest and their complex phenotypic and genetic correlation structure pose restrictions in selection decision, as well as in computations. Dealing with large amount of variables, factor analysis (FA) is commonly adopted to identify latent structures (factors; Fs) of correlated variables. Based on the observed covariance structure, the objective of FA is to replace n measured variables with p ($p < n$) Fs, where the measured variables are expressed as linear functions of the Fs, and the Fs capture the underlying latent concept that the original variables represent (Bollen, 2014).

In dairy cattle, the potential use of Fs obtained from FA has been investigated for a variety of traits, such as milk quality, milk technological properties, e.g., MCP and cheese-related traits (CY) (Macciotta et al., 2012), as well as milk fatty acids (Conte et al., 2016). However, those studies were focused on the sources of variation related to the Fs and their genetic parameters. In addition, Fs have been used within the framework of structural equation modelling for the analysis of bovine mastitis (Detilleux et al., 2013). Despite this, the potential use of Fs in genome wide associations (GWAS) has not been yet explored.

In recent years, GWAS have become a valuable tool in dairy cattle breeding programs. Generally, each trait is separately analyzed in a classical GWAS approach. However, in the case of complex phenotypes, e.g. the 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 to 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. However, all principal components need to be tested for associations, hence no benefit in data reduction is achieved (Aschard et al., 2014). Besides, 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).

On top of the GWAS, it is becoming common to complement association analyses with gene-set enrichment and pathway analyses to alleviate problems related to GWAS, and to deepen the understanding of the biological pathways affecting quantitative traits (Gambra et al., 2013; Peñagaricano et al., 2013; Abdalla et al., 2016; Iso-Touru et al., 2016). For instance, GWAS ignores the fact that genes work together in networks in the various biological pathways. Moreover, correction for multiple testing may result into stringent thresholds, while strong linkage disequilibrium, as found in livestock species, can mask the causal genomic region, thus important markers might be undetected (Peng et al., 2010; Hayes, 2013; Ha et al., 2015).

Integration of Fs, GWAS and pathways analyses might address part of the aforementioned issues and therefore appears attractive in dairy cattle breeding. This combination has been adopted in human studies (Fanous et al., 2012), while its potential application in animal breeding is still unexplored.

Our objective was to conduct genome-wide associations and gene ontology and pathway analysis using a set of latent variables obtained from 26 traits related to milk yield and quality, curd firming, and individual cheese properties in a sample of 1,152 Brown Swiss cows genotyped with a 50k SNP chip.

MATERIALS AND METHODS

Animals and sampling

A detailed description of the sampling procedure was previously reported (Cipolat-Gotet et al, 2012). In brief, 1,264 Italian Brown Swiss cattle were sampled once (evening milking). Animals belonged to 85 herds. In total, 29 traits involved in cheese-making were included in the analyses. In Fig 1, a schematic representation of the basic features of the cheese-making process is presented.

Phenoypic data

Milk quality and 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, Hillerød, Denmark) and somatic cell scores were obtained through logarithmic transformation [(SCS = $\log_2$SCC/100,000) – 3] (Ali and Shook, 1980). Casein (CN) fractions (αs_1-, αs_2-, β- and κ- CN) and whey proteins [β-lactoglobulin (β-LG) and α-lactalbumin (α-LA)] were measured using a validated reversed-phase high-performance liquid chromatography (RP-HPLC) method (Bonfatti et al., 2008). Each fraction was expressed as ratio to the total milk nitrogen (N) content. Moreover, the phosphorylated form of the αs_1-casein 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 (CF_t) and derived from the CF modeling (Bittante et al., 2013b) were included in our analysis: rennet coagulation time (RCT_{eq}, min), maximum curd firmness (CF_{max}, mm; at time t_{max}, min) and time to reach CF_{max} (t_{max}, min), potential asymptotical curd firmness in the absence of syneresis (CF_P, mm), the rate constants of curd-firming (k_{CF}, % x min^{-1}) and syneresis (k_{SR}, % x min^{-1}). 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., 2016). The three CF_t

model parameters (RCT_{eq}, k_{CF}, and k_{SR}) were obtained through curvilinear regression (PROC NLIN; SAS Institute Inc., Cary, NC).

Individual cheese yield and curd nutrients recovery

Individual cow cheese phenotypes, obtained through a model cheese-making procedure (Cipolat-Gotet et al., 2013), were included in the analysis. Individual cheese yields, expressed as percentage of the weight of the total milk processed, comprised: the weight of the curd dry matter ($\%CY_{SOLIDS}$) and water ($\%CY_{WATER}$) as well as their sum (fresh curd; $\%CY_{CURD}$). 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 REC_{SOLIDS}, REC_{FAT} and $REC_{PROTEIN}$. Finally, the energy within the curd (REC_{ENERGY}), calculated as the difference between energy in the milk and in the cheese (NRC, 2001), was also obtained.

Genotyping

Cows (n=1,152) 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. Primary to factor analysis, 3 out of the 29 phenotypes (CF_{max}, $\%CY_{CURD}$ and REC_{SOLIDS}) were excluded to avoid severe multicollinearity problems: i) $\%CY_{CURD}$ is the sum of $\%CY_{SOLIDS}$ and $\%CY_{WATER;}$ ii) CF_{max} is proportional to CF_P; the phenotypic correlation coefficients of REC_{SOLIDS} with REC_{ENERGY} and $\%CY_{SOLIDS}$ were greater than 0.9 (Bittante et al., 2013a).

The following factor model was used to simultaneously analyze the remaining twenty-six

phenotypes:

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

where **x** is a vector of the 26 phenotypes and **ξ** is the factor vector. The factor loadings, relating the factors to the original variables, are contained in **Λ** and **δ** is the residual vector.

Following, 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 control 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 (Schaefer et al., 2013). Further, 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 Fs. To explain the Fs, a threshold of factor loadings > |0.4| was considered as "significant" (Fanous et al., 2012). The FA was implemented using the *psych* package (Revelle, 2014) in the R environment (R Core Team, 2013).

Genome-wide associations. A single marker regression was fitted for GWAS using the GenABEL package in R (GenABEL project developers, 2013; R Core Team, 2013) 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: Firstly, 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. Finally, the genomic control corrects for conservativeness of the procedure (Svishcheva et al., 2012). The polygenic model was:

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

where **y** is a vector containing the latent variables; β is a vector with the fixed effects of i) days in milk of the cow (classes of 30 days each), ii) parity level of each cow (with classes 1, 2, 3, ≥4), iii) the effect of the pendulum (considered only for the CF_t traits), and iv) herd-date effect (n=85); **X** is an incidence matrix connecting each observation to specific levels of factors in β. The non-genetic effects have been previously studied in the same dataset (Cipolat-Gotet et al., 2013; Bittante et al., 2013a). The two random terms in the model were the animal and the residuals, which were assumed to be normally distributed as $\boldsymbol{a} \sim N(0, \mathbf{G}\sigma_g^2)$ and $\boldsymbol{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where **G** and **I** are the genomic relationship and the identity matrix, respectively, σ_g^2 is the additive genomic and σ_e^2 the residual variance. 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 R package "qqman" (Turner, 2014).

Gene-set Enrichment and Pathway-based Analysis. Nominal *P*-values < 0.05 obtained from the GWAS were used to select the significant SNP 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; Durinck et al., 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, while the background genes were the genes associated to those SNP. The Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata et al., 1999) and the Gene Ontology (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 three 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, false discovery rate (FDR) 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

Factors' extraction

Ten factors were extracted from 26 variables explaining 74% of the original variability. Summary statistics of all the 29 phenotypes included in this study are shown in Table 1, while the Pearson and partial correlations among the 26 variables finally used in FA are presented in Figure S1. The average KMO value in our dataset was 0.55. The factor loadings with their given names (sorted by maximum variance explained) are shown in Table 2.

A full description of the factors is given in CHAPTER 4. In brief, the first factor ($F1_{\%CY}$), in order of variance, was loaded on $\%CY_{SOLIDS}$, fat and protein (%) and REC_{ENERGY}, representing the percentage of cheese yield. The second factor ($F2_{CFt}$) underlined the curd firmness process. It was associated with REC_{FAT} and all CF_t traits, but CF_P, with positive loadings with the curd firmness rate constants (k_{CF} and k_{SR}) and negative relations with the time traits (RCT_{eq} and t_{max}). The following factor was associated with the milk yield traits ($F3_{Yield}$). Factor 4 ($F4_{Cheese\ N}$) was considered representative of the nitrogen found in the cheese, with heavy and negative association with β-LG and positive relations to other N compounds in milk and the $REC_{PROTEIN}$. The fifth factor ($F5_{as_1\text{-}\beta\text{-}CN}$) was representative of as_1-CN (positive association) and β-CN (negative association). The sixth factor ($F6_{Udder\ health}$) reflected the cow's udder health status, being associated with lactose (positively), and with the SCS and the remaining milk N compounds (albeit with a weaker and negative relation). Factor 7 ($F7_{\kappa\text{-}\beta\text{-}CN}$) was considered indicator of the κ- and β- CN, having a positive association with

κ-CN and negative with the β-CN. Finally, factors 8, 9 and 10 were heavily loaded to only 1 trait and were named accordingly ($F8_{as_2\text{-CN}}$, $F9_{as_1\text{-CN-Ph}}$ and $F10_{\alpha\text{-LA}}$, respectively).

Genome-wide associations

The GWAS results of the 10 latent variables are summarized in Table 3 and more details can be found in Table S1. In total, 149 SNP were significant on 10 chromosomes (1, 2, 6, 9, 10, 11, 19, 20, 25 and 27). Three of those SNP had unknown position on the genome. All latent variables showed signals except $F5_{Yield}$. Shared signals among traits were found. The strongest signals were detected on BTA6 (~87,4 Mbp) and BTA11 (~104,3 Mbp). More precisely, the marker Hapmap52348-rs29024684 located on 87,396,306bp on 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 $F5_{\alpha s_1\text{-}\beta\text{-CN}}$ ($P = 2.84 \times 10^{-47}$). Moreover, $F5_{\alpha s_1\text{-}\beta\text{-CN}}$ 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 $F8_{\alpha s_2\text{-CN}}$ ($P = 8.34 \times 10^{-11}$ at 87,201,599 bp; $P = 1.67 \times 10^{-10}$ at 87,245,049 bp). The same was observed for $F9_{\alpha s_1\text{-CN-Ph}}$ with a $P = 3.86 \times 10^{-11}$ at 87,201,599bp and $P = 7.80 \times 10^{-11}$ at 87,245,049bp. All casein Fs showed signals on BTA6 in the region 6e (~77.2-89.1 Mbp) (Fig 2b). On BTA11, marker ARS-BFGL-NGS-104610 (104,293,559bp) was strongly linked to $F4_{\text{Cheese N}}$ ($P = 9.81 \times 10^{-26}$). On BTA 6, 11, 20 and 27 signals were distributed in more than one chromosomic regions.

On BTA6, 8 sub-regions were overall detected (Table 3, Fig 2). In regions 6a (~40 Mbp), 6b (~46.6 Mbp) and 6c (~68.5 Mbp), weak signals were detected for the factors $F6_{\text{Udder health}}$, $F2_{CFt}$ and $F8_{\alpha s_2\text{-CN}}$, respectively. The region 6d (~71-74.6 Mbp) was associated to both $F8_{\alpha s_2\text{-CN}}$ and $F7_{\kappa\text{-}\beta\text{-CN}}$. The denser region (6e) was found between ~77-89 Mbp and included 71 significant SNP. In this genomic area, all factors, but $F10_{\alpha\text{-LA}}$ and $F3_{Yield}$, showed associations with a peak at ~87,4 Mbp. Especially for $F7_{\kappa\text{-}\beta\text{-CN}}$, the proportion of additive genetic variance explained by Hapmap52348-

rs29024684 (~87.4 Mbp) reached 74.2%. In addition, the marker Hapmap28023-BTC-060518, located at ~87.2 Mbp, explained ~53% of the additive genetic variance for F5 $_{\alpha s_1\text{-}\beta\text{-}CN}$. In both cases the SNP effects were negative and considerably large, around 1 standard deviation from the mean (Table 4).

Close to region 6e, at ~90.7-92.6 Mbp (region 6f), 8 SNP were significant for 3 Fs:F8$_{\alpha s_2\text{-}CN,}$ F7$_{\kappa\text{-}\beta\text{-}CN}$ and F5$_{\alpha s_1\text{-}\beta\text{-}CN}$. Moreover, F8$_{\alpha s_2\text{-}CN}$ and F7$_{\kappa\text{-}\beta\text{-}CN}$ were associated to a region at ~94.2Mbp (region 6g). A weak association, close the significance threshold, was detected at ~114.2 Mbp for F1$_{\%CY}$ (region 6h).

Five distinct genomic regions were identified on BTA11 (Table 3, Fig 3). The regions at ~4.4 Mbp, ~77.5 Mbp, ~87.7 Mbp and ~97.8 Mbp were associated with F6$_{\text{Udder health}}$, F9$_{\alpha s_1\text{-}CN\text{-}Ph}$, F2$_{CFt}$ and F4$_{\text{Cheese N}}$, respectively. In the range ~101.3-106,5 Mbp (region 11e), 18 significant SNP were detected for F4$_{\text{Cheese N}}$ and F2$_{\alpha\text{-}LA}$, with a peak at ~104,3 Mbp.

Apart from BTA6 and 11, significant associations were detected on other chromosomes, albeit at a weaker strength (Table 3). Two regions were detected on BTA20 at ~7.9 Mbp and ~46.7 Mbp. The first region was associated to F10$_{\alpha\text{-}LA}$ and the second to F9$_{\alpha s_1\text{-}CN\text{-}Ph}$. Moreover, on BTA27, 2 chromosomic regions were detected. Although close to each other, they were associated with different Fs. More precisely, F1$_{\%CY}$ was associated to one marker at ~42.1 Mbp, while 3 SNP were linked to F10$_{\alpha\text{-}LA}$ in the range ~43.4-43.9 Mbp. The rest of the signals were 1 trait-1 factor associations and close to the significance threshold. F7$_{\kappa\text{-}\beta\text{-}CN}$ was associated to 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$_{\alpha s_1\text{-}\beta\text{-}CN}$. One marker at ~10.7 Mbp on BTA10 was associated with F8$_{\alpha s_2\text{-}CN}$. At the beginning of BTA19 (~1.8 Mbp), a weak signal was detected for F1$_{\%CY}$. On BTA25, one marker was associated with F6$_{\text{Udder health}}$ at ~5.4 Mbp.

Gene-set enrichment and pathway-based analysis

Out of 37,568 tested SNP in GWAS, 10,094 were located in annotated genes or in the 15 kbp window. In total, 13,269 background genes were annotated in the *Bos taurus* UMD3.1 assembly. On average, 1,550 SNP were significant per factor based on nominal *P*-value. From those SNP, 529 were assigned to genes and 454 genes were mapped (average values per factor) (Table S2).

After FDR control (5%), 33 GO terms and 6 KEGG categories were associated with 4 of the 10 tested Fs, namely $F1_{\%CY}$, $F4_{Cheese\ N}$, $F8_{\alpha s_2\text{-}CN}$ and $F10_{\alpha\text{-}LA}$, with the vast majority being associated with $F8_{\alpha s_2\text{-}CN}$. Results of the gene-set enrichment and pathway-based analysis are outlined in Fig 4 and Table S3. A total of 117 genes spanning all BTA but 21 and 29 were included into the significantly enriched GO and KEGG categories (Table S4). $F4_{Cheese\ N}$ was associated with the arrhythmogenic right ventricular cardiomyopathy (ARVC; KEGG: bta05412). The tight junction pathway (KEGG: bta04530) was enriched for $F1_{\%CY}$ and $F10_{\alpha\text{-}LA}$. Three KEGG categories were enriched for $F8_{\alpha s_2\text{-}CN}$, 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 $F8_{\alpha s_2\text{-}CN}$: 12 GO_BP related to cell communication and ion transport, 11 GO_CC belonging to neuron part/function and 10 GO_MF related to ion transport.

DISCUSSION

Extraction of factors

Using FA, we condensed 26 cheese-making traits into ten Fs. Although the average KMO value was not high, it is 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 Fs in our study represented basic concepts of the "*cheese-making*" process while retained 74% of the original variability. In a similar dataset, but with eleven MCP and udder health phenotypes, the total variance explained by the 4, in total, extracted Fs was 70% (Macciotta et al., 2012). The same factor scores

were previously tested in an ANOVA and genetic analysis with results being coherent to the given name of the factors (CHAPTER 4). Given the wide use of genomics in several breeding programs nowadays, a further step was to explore the behavior of the extracted Fs within the framework of GWAS, and gene-set enrichment and pathway-based analyses.

Genome wide associations

Previous GWAS studies detected several chromosomic regions related to bovine milk protein components (Schopen et al., 2011; Bijl et al., 2014), MCP and CF_t characteristics (Gregersen et al., 2015; Dadousis et al., 2016a), and individual CY traits (Dadousis et al., 2016b). Major effects are known on BTA6 for milk technological traits and protein variants, in a region spanning between ~82 – 88 Mbp (Schopen et al., 2011; Gregersen et al., 2015; Dadousis et al., 2016a), including the casein cluster, two potential QTL (quantitative trait locus) have been suggested at ~82.6 and ~88.4 and at the tail of BTA11 (at ~87 and ~104 Mbp) (Schopen et al., 2011; Dadousis et al., 2016a). The location of the casein genes on BTA6 is widely known (Caroli et al., 2009), while the signals on BTA11 were mainly attributed to β-lactoglobulin gene (*BLG*). Moreover, the effect of milk protein variants in 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 explained ~74% of the total additive genetic variance for $F7_{\kappa\text{-}\beta\text{-CN}}$, having a strong and negative effect (Table 4). In previous studies, this marker was strongly linked with a trait describing the potential asymptotical curd firmness (Dadousis et al., 2016a) and with the REC_{FAT} (Dadousis et al., 2016b). In our study, the same marker was also associated with $F6_{\text{Udder health}}$, albeit at a much weaker strength compared to $F7_{\kappa\text{-}\beta\text{-CN}}$. The general region roughly between 83.4-88.9 Mbp has been associated with

clinical mastitis in Nordic Holstein (Sahana et al., 2013). It is worth mentioning that SCS was a minor loading on $F6_{Udder\ health}$. Moreover, the region at ~88.8Mbp has been associated with SCS in US Holstein cows (Cole et al., 2011). Close to Hapmap52348-rs29024684, at ~87,2Mbp, the Hapmap28023-BTC-060518 was associated with $F5_{\alpha s_1\text{-}\beta\text{-}CN}$, $F7_{\kappa\text{-}\beta\text{-}CN}$, $F8_{\alpha s_2\text{-}CN}$ and $F9_{\alpha s_1\text{-}CN\text{-}Ph}$ (Table 4, Table S1). The highest effect of this SNP was found for $F5_{\alpha s_1\text{-}\beta\text{-}CN}$ and explained ~53% of the additive genetic variability. This marker is located within the histatherin gene (*HSTN*) and there is an evidence that this gene underlies QTL related with CF_t traits and REC_{FAT} (Dadousis et al., 2016a; Dadousis et al., 2016b). In the region (~88.4 Mbp), $F2_{CFt}$ and $F7_{\kappa\text{-}\beta\text{-}CN}$ were associated with BTA-122637-no-rs (*P-value* = 6.91×10^{-6} and 2.46×10^{-10}, respectively; Table S1). Notably, the same marker has been previously associated with RCT_{eq}, while in the broader region ~87.2-88.8 Mbp, the CF_{max}, k_{CF} and protein percentage, also had signals (Dadousis et al., 2016a). The last 2 traits were the major loadings of $F2_{CFt.}$ This marker is located within the *SLC4A4* (electrogenic sodium bicarbonate cotransporter 1) protein coding (~88.2-88.5 Mbp). The product of *SLC4A4* (sodium bicarbonate cotransporter) is involved in intracellular pH, and regulates the secretion and absorption of bicarbonate. Very close to this region, the *GC* (group-specific component) protein coding is mapped (~88.69–88.74 Mbp). The *GC* encodes a vitamin-D binding protein. The protein is involved in the metabolism of the vitamin D, lipids and lipoproteins and belongs in the albumin family. In a recent fine mapping study on BTA6, using sequencing data in Norwegian Red cattle, the *GC* has been suggested as a candidate gene in this region related with milk production and clinical mastitis (Olsen et al., 2016). Factor $F1_{\%CY}$ was also associated in the region 6e, with a peak at ~82,7Mbp (Hapmap53172-rs29012675). The same marker has been previously associated with $\%CY_{SOLIDS}$, $\%CY_{CURD}$, REC_{FAT}, REC_{ENERGY} and REC_{SOLIDS} (Dadousis et al., 2016b). Not surprisingly, $F1_{\%CY}$ was primarily loaded to $\%CY_{SOLIDS}$, as well as to REC_{ENERGY}.

In the region 6b, a weak signal for $F2_{CFt}$ was detected (Hapmap23226-BTA-159656, ~46.6Mbp). The same region was previously associated with t_{max} (Dadousis et al., 2016a). The t_{max}

was strongly related to $F2_{CFt}$, but it was not the heaviest loading of this factor. The region 6h, at ~114.2Mbp, was exclusively associated to the $F1_{\%CY}$, 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., 2016a). Interestingly, this factor was loaded to milk protein (%), although with the weaker relation (0.59) among the rest of the traits describing the factor.

BTA11

Overall, five of the ten Fs were linked to five regions on BTA11. The strongest association was found between $F4_{Cheese\ N}$ and ARS-BFGL-NGS-104610 (104,293,559 bp). The same marker has been strongly related to $REC_{PROTEIN}$ (Dadousis et al., 2016b). Notably, $F4_{Cheese\ N}$ was loaded on $REC_{PROTEIN}$. Part of the region 11e (around 104,46 Mbp) was also associated to $F10_{\alpha\text{-}LA}$. However, there is no known QTL on BTA11 related with α-LA (Schopen et al., 2011). The region 11d associated with $F4_{Cheese\ N}$ is in close proximity to the region 96,2-98,5 Mbp where signals have been previously detected for $REC_{PROTEIN}$ (Dadousis et al., 2016b). In both cases same peak was observed at ~97 Mbp. $F2_{CFt}$ 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., 2016a). A weak association at ~4.4 Mbp was found for $F6_{Udder\ health}$. Although far from this region, signals for SCS have been reported in US Holstein cows at the beginning of BTA11 at ~0.28 and ~2.75 Mbp (Cole et al., 2011).

Signals on chromosomes other than BTA6 and BTA11

Our study detected weak associations in 8 additional chromosomes (Table 3). With the exception of BTA20 and 27, the rest of the chromosomes were linked to only 1 factor. The SNPs associated to $F1_{\%CY}$ on BTA19 and 27 have been significantly related to $\%CY_{SOLIDS}$ (Dadousis et al., 2016b), while the marker on BTA2 is ~6Mbp downstream to the one associated to $\%CY_{SOLIDS}$ (Dadousis et al., 2016b). On BTA25, $F6_{Udder\ health}$ was associated to a SNP at ~5.4 Mbp, in close proximity to the ~5.3 Mbp region which 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 since neither of these casein fractions have been associated to this region on BTA1. Moreover, individual GWAS for κ- and β-CN didn't result in significant associations on BTA1 (results not shown). Further, based on the fact that only 1 SNP passed the significance threshold while the rest of the markers in the same region showed much lower *P*-values, we could hypothesize for a spurious association. Similarly, we found significant associations on BTA9 for $F5_{\alpha s_1\text{-}\beta\text{-CN}}$ but no associations on BTA9 have been previously reported for αs_1- or β-CN. However, GWAS analysis using the individual αs_1-CN content detected the same marker with similar *P-value* (results not shown). On BTA10, two genomic regions are known to be related to αs_2-CN, at ~51.4 and ~91.8Mbp (Schopen et al., 2011). In our analysis, $F8_{\alpha s_2\text{-CN}}$ was associated to a region at ~10.7Mbp. No QTL is known at this position affecting the αs_2-CN. Moreover, no justification of QTL could be found for the associations on BTA20 and 27b with $F10_{\alpha\text{-LA}}$.

Gene-set enrichment and pathway analysis

Four Fs, $F1_{\%CY}$, $F4_{\text{Cheese N}}$, $F8_{\alpha s_2\text{-CN}}$ and $F10_{\alpha\text{-LA}}$, out of 10 tested, were associated with biological pathways and ontologies in the KEGG and GO databases (Fig 5, Table S4). The majority of the significantly enriched terms were associated with $F8_{\alpha s_2\text{-CN}}$. To support these findings on $F8_{\alpha s_2\text{-CN}}$, we re-run the GWAS and gene-set enrichment analysis on the measured αs_2-CN, as well as the rest of the caseins. Gene-set enrichment results for αs_2-CN were similar (results not shown), but the KEGG categories as well as 13 out of the 33 GO terms were not enriched. Moreover, GO and KEGG categories were not enriched for the rest of the caseins, being consistent with the Fs 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 traits (Dadousis et al., 2016c), 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 the cheese process (Walstra et al., 2014), they also have bioactive and antimicrobial properties (Zucht et al., 1995; Silva and Malcata, 2005; López-Expósito et al., 2006). Moreover, an antimicrobial role of α-LA has also 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 $F8_{\alpha s_2\text{-CN}}$ 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 as_2-CN is known to be as rather sensitive to Ca^{+2} (Walstra et al., 2014). 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 intramammary infection, 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 $F10_{\alpha\text{-LA.}}$ Besides, it has been reported that milk with high cell count has lower casein content (Haenlein et al., 1973; Auldist and Hubble, 1998). Pathways related to mammary gland and mastitis, including the tight junction, have been previously associated with the REC_{ENERGY} (Dadousis et al., 2016c), a trait that was strongly related with $F1_{\%CY}$ in the FA.

Moreover, for $F8_{\alpha s_2\text{-CN}}$, GO terms related to cell communication and signalling (e.g. GO_BP:0023052, GO_BP:0007154) were enriched in our study. Those categories have been shown to have a role in human milk fat globule transcriptome, while among a set of genes in milk fat globule, the *CSN2* (casein beta) was also expressed [56].

Enriched pathways and ontologies for $\mathbf{F8_{\alpha s_2\text{-CN}}}$ ***related to reproduction.*** Seven GO_CC categories relative to neuron functions were enriched for $F8_{\alpha s_2\text{-CN}}$. A possible explanation can be due to 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 (Grattan, 2002; Akers, 2016). Interestingly, in a recent gene enrichment and pathway study, using the individual CF_t traits, 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., 2016c). Moreover, associations of the synapse part (GO:0044456) and the postsynapse (GO:0098794) with the k_{CF} were detected in that study. The cellular response to stimulus category (GO_BP:0051716) was also significantly enriched for $F8_{\alpha s_2\text{-CN}}$. 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 gonadotropin-releasing hormone (GnRH) signaling pathway (KEGG_bta04912) was enriched for $F8_{\alpha s_2\text{-CN}}$. The GnRH is synthesized and released in the hypothalamus from the GnRH neurons. As it is widely known, GnRH is 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., 2016c). 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 oestrus and multiple ovulation rate and milk production

in dairy cattle. More precisely, high production is associated with shorter oestrus duration and double ovulation rate (Wiltbank et al., 2006).

In our study ARVC was enriched for $F4_{\text{Cheese N}}$. 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., 2016c). Notably, in a transcriptome study of the swine mammary gland, ARVC was associated to the mammary gland functionality of pregnant sows (Zhao et al., 2013).

CONCLUSION

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 to the expected signals based on the factor names. Results of gene-set enrichment analysis were also in line with previous findings based on the individual traits, 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 $F8_{\alpha s_2\text{-CN}}$ suggests that, perhaps, more focus should be given in αs_2-CN.

We conclude that factor analysis could be successfully implemented in genomic studies in dairy cattle. Reduce of dimensionality without substantial loss of information mark factor analysis as an attractive tool for dairy cattle breeding.

ACKNOWLEDGMENTS

The authors wish to thank the Trento Province (Italy), the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy) and the Superbrown Consortium of Bolzano and Trento for financial and technical support. C. Dadousis benefitted from financial support of the CARIPARO (Cassa di Risparmio di Padova e Rovigo) Foundation (Padua, Italy).

Table 1. Summary statistics of milk (yield and quality), protein fractions, curd firming and cheese-making (%CY and REC) traits.

Trait[1]	Mean	CV, %
Milk traits		
$Milk_{yield}$, kg/day	24.95	31
Fat_{yield}, kg/day	1.09	37
$Protein_{yield}$, kg/day	0.92	30
Fat, %	4.37	20
Protein, %	3.71	11
Lactose, %	4.86	4
pH	6.64	1
SCS, units	2.87	65
Milk protein fractions, %		
αs_1-CN	25.69	7
αs_1-CN-Ph	1.45	42
αs_2-CN	9.20	12
β-CN	32.26	8
κ-CN	9.44	16
β-LG	8.68	18
a-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^{-1}	12.90	32
k_{SR}, % × min^{-1}	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 = $\log_2$ (SCC × 100,000) + 3. Milk protein fractions: CN = casein; LA = lactalbumin and LG = lactoglobulin. Curd firming: RCTeq = estimated RCT; 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 dry matter ($\%CY_{SOLIDS}$) and curd water ($\%CY_{WATER}$) versus the weight of the processed milk (g); REC = ratio of the weight (g) of the curd constituent (dry matter, fat, protein or energy, respectively) versus that of the same constituent in the processed milk (g).

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

Traits[3]	F1$_{\%CY}$	F2$_{CFt}$	F3$_{Yield}$	F4$_{Cheese\ N}$	F5 $_{\alpha s_1\text{-}\beta\text{-}CN}$	F6$_{Udder\ health}$	F7$_{\kappa\text{-}\beta\text{-}CN}$	F8$_{\alpha s_2\text{-}CN}$	F9$_{\alpha s_1\text{-}CN\text{-}Ph}$	F10$_{\alpha\text{-}LA}$	com
Milk traits											
Milk$_{yield}$, kg/day	-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/day	0.28	0.08	**0.89**	-0.02	0.06	0.17	-0.01	-0.01	-0.01	0.07	0.92
Prot$_{yield}$, kg/day	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
Prot, %	**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, units	0.06	-0.02	-0.08	0.04	-0.05	**-0.41**	0.09	0.01	0.03	-0.09	0.20
Milk protein fractions, %											
as$_1$-CN	0.04	0.01	0.07	-0.14	**0.94**	0.25	-0.04	-0.08	-0.14	-0.04	0.99
as$_1$-CN-Ph	0.04	-0.02	0.00	0.09	-0.10	0.01	-0.04	-0.04	**0.98**	0.06	1.00
as$_2$-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 comp.	-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^{-1}	-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^{-1}	-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. $F1_{\%CY}$ = 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; $F5_{\alpha s_1\text{-}\beta\text{-}CN}$ = Factor related to the as_1- 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; $F8_{\alpha s_2\text{-}CN}$ = Factor related to the milk as_2-CN, expressed as relative content to the total milk nitrogen; $F9_{\alpha s_1\text{-}CN\text{-}Ph}$ = Factor related to the milk as_1-CN-Ph expressed as content to the total milk nitrogen; $F10_{\alpha\text{-}LA}$ = Factor related to the milk α-LA.[3]SCS = $\log_2$ (SCC × 100,000) + 3. Milk protein fractions: CN = casein; LA = lactalbumin and LG = lactoglobulin. Curd firming: RCT_{eq} = estimated RCT; CF_P = asymptotical potential value of CF; k_{CF} = curd-firming instant rate constant; k_{SR} = syneresis instant rate constant; and t_{max} = time at achievement of CF_{max}. %CY = ratios of the weight (g) of the curd dry matter ($\%CY_{SOLIDS}$) and curd water ($\%CY_{WATER}$) versus the weight of the processed milk (g); REC = ratio of the weight (g) of the curd constituent (fat, protein or energy, respectively) versus that of the same constituent in the processed milk (g).

Table 3 Summary results of the genome wide association analyses.

BTA	# SNP (signals)	Interval (Mbp)	*P*-value (range)	Top SNP	Top SNP location, bp	Top SNP MAF	Trait[1]
0*	5 (3)	-	$(4.66 \times 10^{-05}, 9.47 \times 10^{-17})$	BTA-76907-no-rs	0	0.26	F5 αs_1-β-CN, **F7κ-β-CN**, F9αs_1-CN-Ph
1	1	-	4.19×10^{-05}	BTB-00041036	90,156,001	0.01	F7κ-β-CN
2	1	-	4.43×10^{-05}	ARS-BFGL-NGS-101039	122,509,616	0.34	F1%CY
6a	3 (3)	39,503–40,378	$(7.89 \times 10^{-06}, 1.19 \times 10^{-06})$	Hapmap26618-BTC-070864	39,597,740	0.03	F6Udder health
6b	1	-	1.07×10^{-05}	Hapmap23226-BTA-159656	46,599,570	0.24	F2CFt
6c	1	-	2.18×10^{-05}	ARS-BFGL-NGS-111636	68,546,212	0.05	F8αs_2-CN
6d	8 (7)	71,154-74,607	$(2.87 \times 10^{-05}, 1.12 \times 10^{-07})$	Hapmap29639-BTC-041962	71,350,048	0.02	F7κ-β-CN, F8αs_2-CN
6e	140 (71)	77,186-89,104	$(4.89 \times 10^{-05}, 9.81 \times 10^{-56})$	Hapmap52348-rs29024684	87,396,306	0.19	F1%CY, F2CFt, F4Cheese N, F5 αs_1-β-CN, F6Udder health, **F7κ-β-CN**, F8αs_2-CN, F9αs_1-CN-Ph
6f	10 (8)	90,730-92,579	$(3.23 \times 10^{-05}, 1.63 \times 10^{-09})$	Hapmap43045-BTA-76998	90,730,485	0.01	F5 αs_1-β-CN, F7κ-β-CN, **F8αs_2-CN**
6g	5 (3)	94,229-94,360	$(3.77 \times 10^{-06}, 1.36 \times 10^{-08})$	BTB-01687386	94,360,125	0.02	F7κ-β-CN, **F8αs_2-CN**
6h	1	-	4.85×10^{-05}	BTB-02092741	114,223,059	0.01	F1%CY
9	1	-	4.24×10^{-05}	BTA-21753-no-rs	36,790,663	0.01	F5 αs_1-β-CN
10	1	-	3.22×10^{-05}	Hapmap41952-BTA-73370	10,659,761	0.38	F8αs_2-CN
11a	1	-	4.72×10^{-05}	BTB-01723556	4,419,032	0.06	F6Udder health
11b	1	-	3.40×10^{-05}	ARS-BFGL-NGS-56195	77,493,775	0.04	F9αs_1-CN-Ph
11c	12 (12)	85,367-88,214	$(3.32 \times 10^{-05}, 3.82 \times 10^{-07})$	BTA-110429-no-rs	87,670,344	0.42	F2CFt
11d	5 (5)	94,687-97,845	$(4.51 \times 10^{-05}, 2.63 \times 10^{-07})$	Hapmap56906-rs29014970	97,844,929	0.31	F4Cheese N
11e	21 (21)	101,301-106,543	$(3.04 \times 10^{-05}, 2.08 \times 10^{-26})$	ARS-BFGL-NGS-104610	104,293,559	0.45	**F4Cheese N**, F10α-LA
19	1	-	4.32×10^{-05}	ARS-BFGL-NGS-102974	1,822,133	0.34	F1%CY
20a	1	-	8.02×10^{-06}	BTA-51080-no-rs	7,881,875	0.01	F10α-LA
20b	1	-	1.84×10^{-05}	Hapmap51592-BTA-41521	46,709,345	0.36	F9αs_1-CN-Ph
25	1	-	3.12×10^{-06}	Hapmap31994-BTC-065943	5,385,729	0.14	F6Udder health
27a	1	-	2.38×10^{-05}	ARS-BFGL-NGS-87845	42,118,037	0.03	F1%CY
27b	3 (3)	43,436-43,902	$(4.84 \times 10^{-05}, 2.84 \times 10^{-05})$	ARS-BFGL-NGS-24170	43,459,156	0.48	F10α-LA

BTA = *Bos taurus* autosome chromosome; #SNP (signals) = number of the single nucleotide polymorphisms significantly associated to the trait. In parenthesis 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) = The *P*-value of the highest significant SNP adjusted for genomic control and the range of the *P*-values when multiple SNP were significantly associated to one 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;

[1] $F1_{\%CY}$ = 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; $F5_{as1\text{-}\beta\text{-}CN}$ = Factor underlying the αs_1 and β caseins; $F6_{Udder\ health}$= Factor underlying the udder health condition of a cow; $F7_{\kappa\text{-}\beta\text{-}CN}$ = Factor underlying the κ and β caseins; $F8_{as2\text{-}CN}$ = Factor underlying the αs_2-casein; $F9_{as1\text{-}CN\text{-}Ph}$ = Factor underlying the phosphorylated αs_1-casein; $F10_{\alpha\text{-}LA}$ = Factor underlying the α-Lactalbumin. In bold the trait with the highest *P*-value in each genomic region.

*Undefined chromosome and position on the genome

Table 4 Highest significant SNP in the region 6e on chromosome 6 (sorted by location) per factor, *P-value*, effect and proportion of the additive genetic variance explained by the highest significant SNP

Traits[1]	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
$F5_{\alpha s_1\text{-}\beta\text{-}CN}$			2.84×10^{-47}	-0.90	52.8
$F8_{\alpha s_2\text{-}CN}$	Hapmap28023-BTC-060518	87,201,599	8.34×10^{-11}	-0.35	16.0
$F9_{\alpha s_1\text{-}CN\text{-}Ph}$			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
$F7_{\kappa\text{-}\beta\text{-}CN}$			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
$F2_{CFt}$	BTA-122637-no-rs	88,442,145	6.91×10^{-6}	-0.40	13.2

Top SNP = the highest significant SNP in the region 6e on chromosome 6 per trait; Top SNP location (bp) = position of the highest significant SNP on the chromosome in base pairs on UMD3.1; *P*-value = The *P*-value of the highest significant SNP adjusted for genomic control; Top SNP effect = the effect of the highest significant SNP. The factor scores are standardized with zero mean and standard deviation 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 one allele, q = 1-p is the frequency of the second allele and a denotes the additive genetic effect).

[1] $F1_{\%CY}$ = Factor underlying the percentage of individual cheese yield; $F2_{CFt}$ = Factor underlying the milk curd firmness; $F4_{\text{Cheese N}}$= Factor underlying the protein in the cheese; $F5_{\alpha s_1\text{-}\beta\text{-}CN}$ = Factor underlying the αs_1 and β caseins; $F6_{\text{Udder health}}$= Factor underlying the udder health condition of a cow; $F7_{\kappa\text{-}\beta\text{-}CN}$ = Factor underlying the κ and β caseins; $F8_{\alpha s_2\text{-}CN}$ = Factor underlying the αs_2-casein; $F9_{\alpha s_1\text{-}CN\text{-}Ph}$ = Factor underlying the phosphorylated αs_1-casein.

Figure 1 Flowchart of the basic milk and coagulation components included in the cheese-making process.

Description: Milk protein fractions: CN = casein; LA = lactalbumin and LG = lactoglobulin. Coagulation: RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Curd firming: RCTeq = estimated RCT; 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 dry matter ($\%CY_{SOLIDS}$) and curd water ($\%CY_{WATER}$) versus the weight of the processed milk (g); REC = ratio of the weight (g) of the curd constituent (dry matter, fat, protein or energy, respectively) versus that of the same constituent in the processed milk (g).

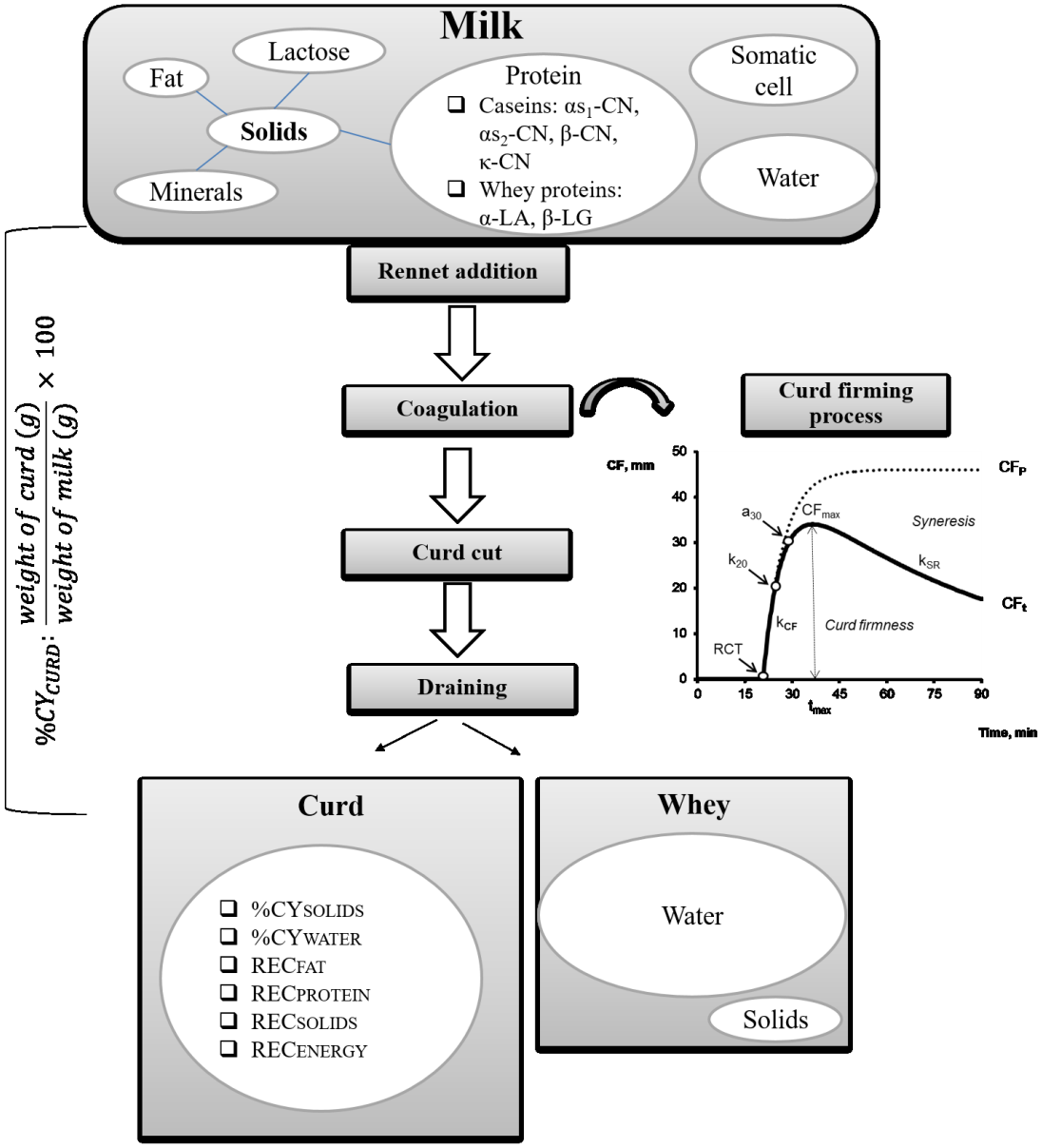
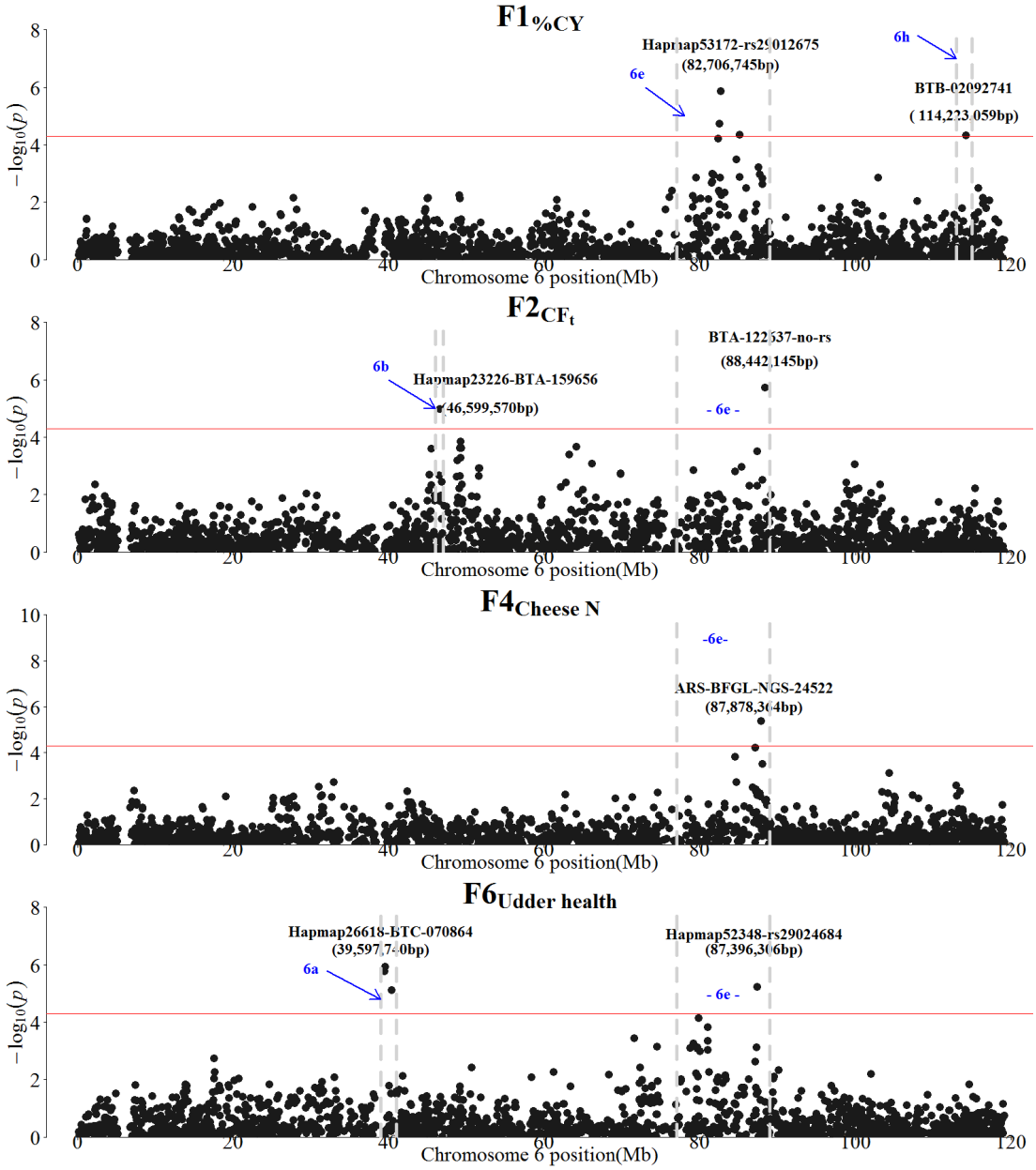

Milk
Lactose
Fat
Solids
Minerals
Protein
Caseins: αs1-CN, αs2-CN, β-CN, κ-CN
Whey proteins: α-LA, β-LG
Somatic cell
Water
Rennet addition
Coagulation
Curd cut
Draining
Curd firming process
CF, mm
CFP
CFt
a30
CFmax
Syneresis
k20
kSR
kCF
Curd firmness
RCT
tmax
Time, min
Curd
%CYSOLIDS
%CYWATER
RECFAT
RECPROTEIN
RECSOLIDS
RECENERGY
Whey
Water
Solids
$\%CY_{CURD}: \frac{weight\ of\ curd\ (g)}{weight\ of\ milk\ (g)} \times 100$

Figure 2 Manhattan plots of *P*-values for the genome-wide association studies on *Bos taurus* autosome 6 (BTA6).

Description: **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)** $F5_{as1\text{-}\beta\text{-}CN}$ = Factor underlying the αs_1 and β caseins; $F7_{\kappa\text{-}\beta\text{-}CN}$ = Factor underlying the κ and β caseins; $F8_{as2\text{-}CN}$ = Factor underlying the αs_2-casein; $F9_{as1\text{-}CN\text{-}Ph}$ = Factor underlying the phosphorylated αs_1-casein. The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}). The highest significant marker on BTA6 per trait is also presented.

a)


F1%CY
Hapmap53172-rs29012675
(82,706,745bp)
6e
6h
BTB-02092741
(114,223,059bp)
F2CFt
6b
Hapmap23226-BTA-159656
(46,599,570bp)
BTA-122537-no-rs
(88,442,145bp)
- 6e -
F4Cheese N
-6e-
ARS-BFGL-NGS-24522
(87,878,304bp)
F6Udder health
Hapmap26618-BTC-070864
(39,597,740bp)
6a
Hapmap52348-rs29024684
(87,396,306bp)
- 6e -
−log10(p)
Chromosome 6 position(Mb)

b)

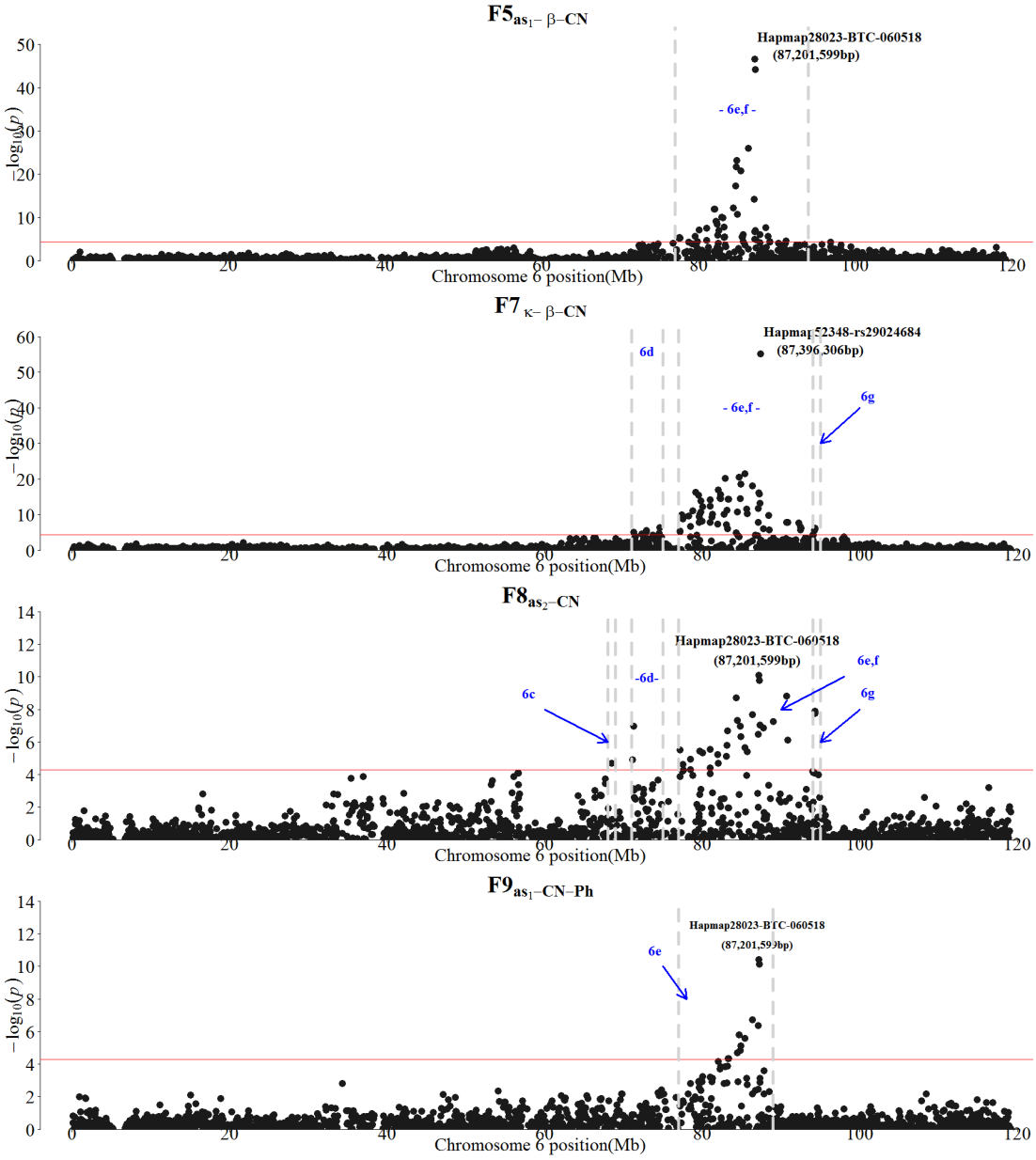

F5as1– β–CN
Hapmap28023-BTC-060518
(87,201,599bp)
- 6e,f -
−log10(p)
Chromosome 6 position(Mb)
F7 κ– β–CN
Hapmap52348-rs29024684
(87,396,306bp)
6d
- 6e,f -
6g
F8as2–CN
Hapmap28023-BTC-060518
(87,201,599bp)
6c
-6d-
6e,f
6g
F9as1–CN–Ph
Hapmap28023-BTC-060518
(87,201,599bp)
6e

Figure 3 Manhattan plots of *P*-values for the genome-wide association studies on *Bos taurus* autosome 11 (BTA11).

Description: $F2_{CFt}$= 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; $F9_{as1\text{-}CN\text{-}Ph}$ = Factor underlying the phosphorylated αs_1-casein; $F10_{\alpha\text{-}LA}$ = Factor underlying the α-Lactalbumin. Red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}).The highest significant marker on BTA11 per trait is also presented.

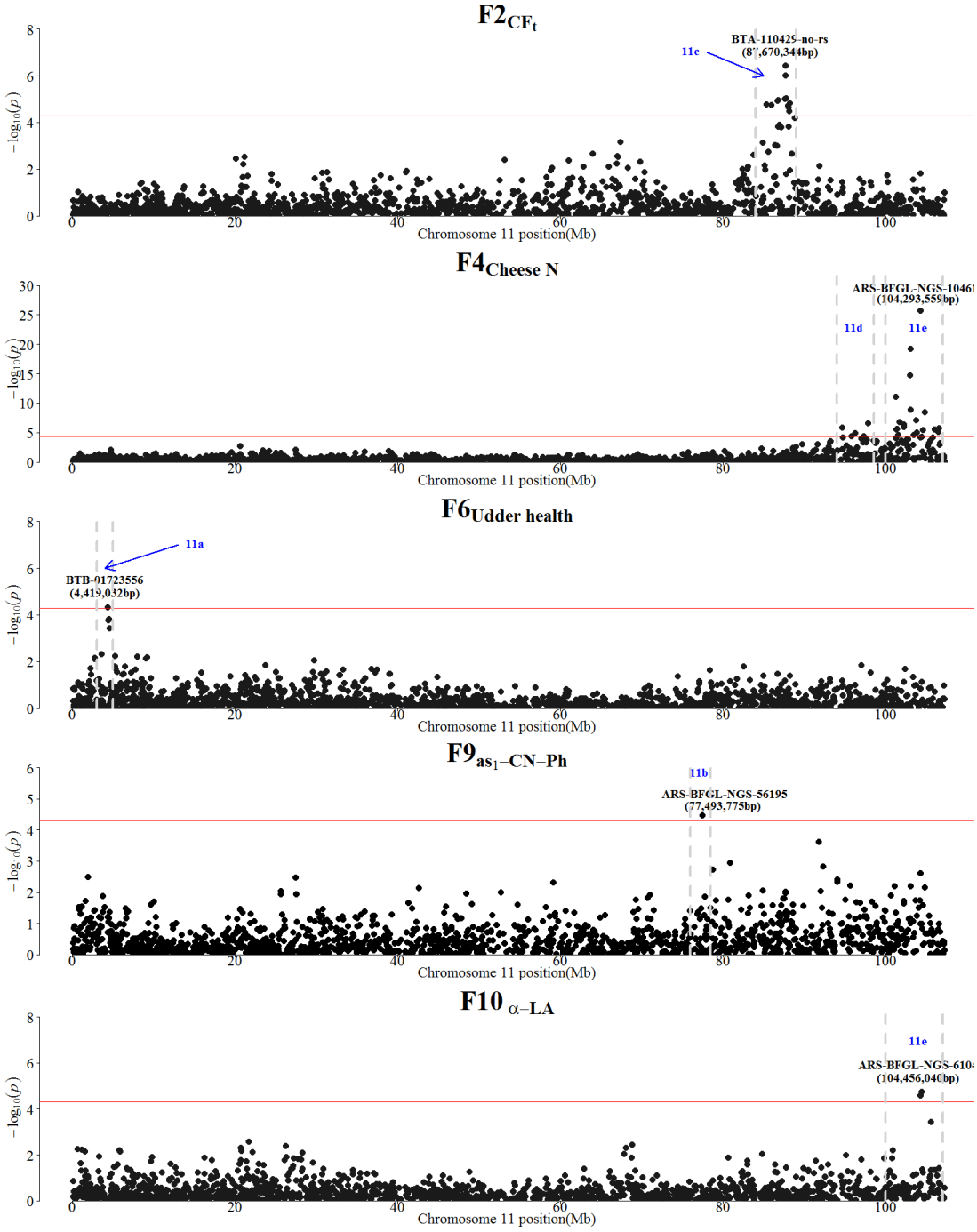
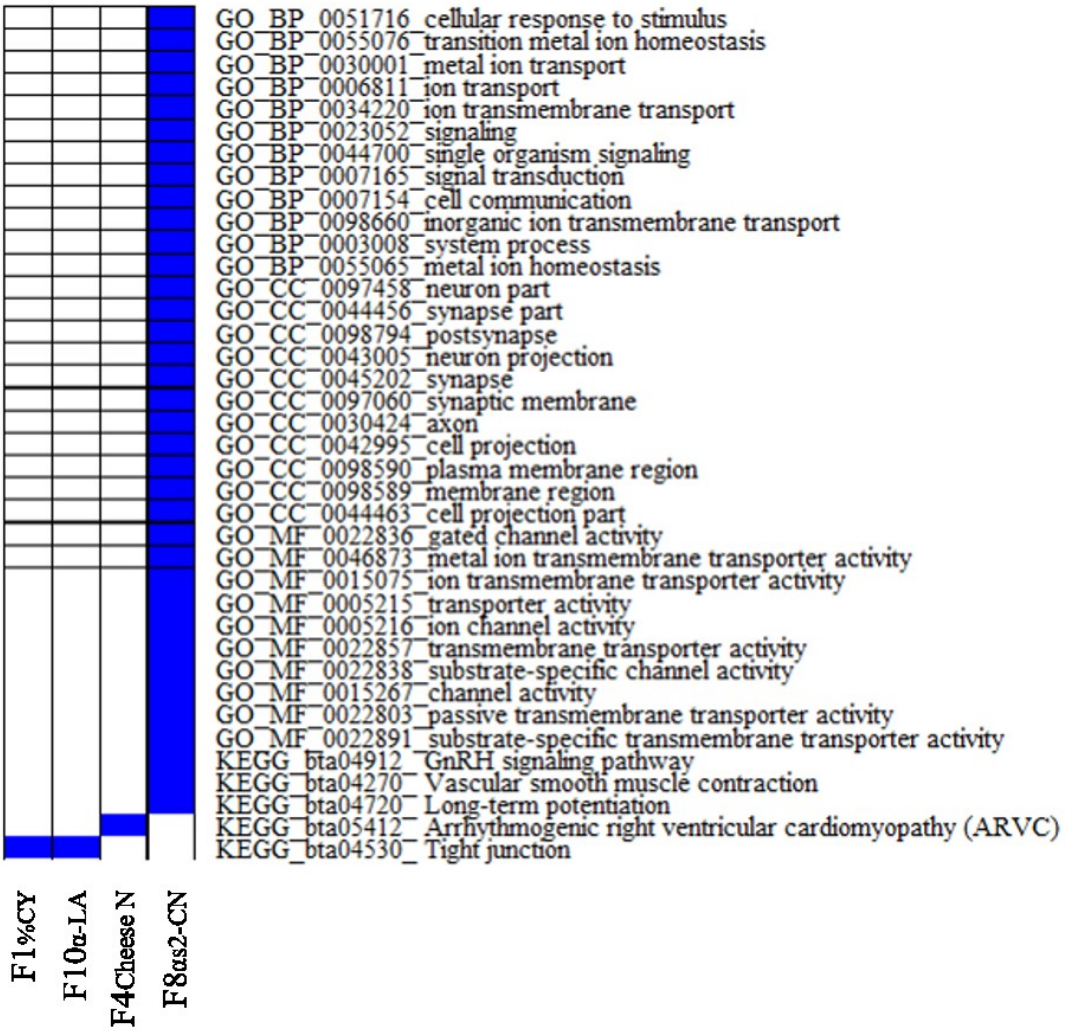

F2CFt
11c
BTA-110425-no-rs
(87,670,344bp)
−log10(p)
Chromosome 11 position(Mb)
F4Cheese N
11d
11e
ARS-BFGL-NGS-1046
(104,293,559bp)
F6Udder health
11a
BTB-01723556
(4,419,032bp)
F9as1−CN−Ph
11b
ARS-BFGL-NGS-56195
(77,493,775bp)
F10 α−LA
11e
ARS-BFGL-NGS-610
(104,456,040bp)

Figure 4. Gene ontology (GO) terms and Kyoto encyclopedia of genes and genomes (KEGG) pathways significantly enriched.

Description: Genes containing significant SNPs (P<0.05) or mapping at 15kbp up- and downstream the significant SNPs (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; $F8_{\alpha s_2\text{-}CN}$ = Factor underlying the αs_2-casein (CN); $F10_{\alpha\text{-}LA}$ = Factor underlying the α-LA. KEGG: KEGG pathway; GO_BP: GO biological process; GO_CC: GO cellular component; GO_MF: GO molecular function.



GENERAL DISCUSSION

Breeding goal: Selecting for bovine cheese production

In recent years, the proportion of bovine milk used for cheese production, is steadily increasing in many countries worldwide (Food and Agriculture Organization of the,United Nations, 2015). Nevertheless, in almost all dairy cattle breeding associations around the world the percentage of cheese yield produced by an individual cow is indirectly selected by measuring the major cheese components that are present in milk, i.e. protein and fat of the milk (Miglior et al., 2005; Banos, 2010). However, this simplicity assumes that i) protein and fat are the major determinants of cheese yield, ii) their recovery from milk to cheese is approximately constant and iii) this recovery is not genetically controlled. Instead, it has recently been shown that although fat and protein contents have high genetic correlations with percentage cheese yields, these values are significantly lower than 100% (Bittante et al., 2013; Cipolat-Gotet et al., 2013).

On the other hand, milk coagulation properties consist technological features of the milk that can be measured during the coagulation process. Coagulation is the transformation of milk from liquid to solid, at the beginning of the cheese process. Therefore, MCP provide with an extra information on cheese compared to the simple milk content components (i.e. fat and protein).

However, MCP are also indicator traits of the cheese-making process. Moreover, traditional MCP traits have some limitations. To overcome the problems related to the classical single point estimates of MCP, such as late and non-coagulating milk samples and low repeatability, it has been proposed to model the CF as a function of time. Thus, for a better description of the CF a set of 6 parameters has been developed. Yet, these parameters are also indirect descriptors of cheese yield. Contrary, the percentage cheese yield is a direct measure of cheese production. In other words, %CY is the only measure that provides the direct answer to the

question on "*how much cheese each cow produces*?". Nevertheless, for the moment and in the nearest future, it does not seem feasible to directly measure the CY traits at a population level due to:

1. Scarcity of appropriate procedures for model – cheese production
2. Complexity of cheese making
3. The frequent use of fat and protein (or casein) contents of the milk as a proxy for cheese yield
4. Time and money highly consuming procedure

One way to deal with this, and to overcome those limitations, is through the use of spectral data (Ferragina et al., 2013). Indeed, much of the research in the dairy industry has been recently focused in the use of spectra. An alternative, is the identification of potential genomic regions affecting the aforementioned traits, that might be useful for implementing gene-assisted selection programs, genomic predictions or gene-editing technology. This research focused in the second alternative.

The primary objective of this PhD thesis was to unveil the genomic background of a plethora of bovine milk technological and cheese-related traits. To this purpose, the first 2 chapters were focused in conducting several GWAS analyses with a set of 11 milk technological traits related to the coagulation process and 7 cheese-related phenotypes.

GWAS results

Sharp peaks were detected on BTA6, especially for REC_{FAT} and CF_P, in a region spanning between 77.5-88.4 Mbp and more precisely at 84 to 88 Mbp, with a peak at ~87.4 Mbp (Hapmap52348-rs29024684) (Figure 1). This region harbors the casein genes and more precisely *CSN3*. Evidence of other quantitative trait loci at 82.6 and 88.4 Mbp on the same chromosome was found. Sharp peaks were also detected on BTA11 were marker ARS-BFGL-NGS-104610 (~104.3 Mbp) was highly associated with $REC_{PROTEIN}$. Also, on BTA11 in a region between 85.9-88 Mbp the traits RCT_{eq} and t_{max} showed signals (Figure 2). Apart from those genomic regions on BTA6 and BTA11, signals in 15 more chromosomes (1, 2, 9, 12, 13, 14, 15, 16, 18, 19, 20, 23, 26, 27 and 28) were detected for the MCP-CF_t and CY-REC traits (Figure 3). Nevertheless, the majority of those associations were weak and on the significance threshold.

Our analysis showed that the importance of the region ~88 Mbp on BTA6 is related, apart from codification of protein genetic variants (Schopen et al., 2011), with the recovery of the milk fat into the curd, and not to the recovery of protein. This was not surprising, since in healthy cows all casein variants are essential for the structure and activity of milk protein micelles (Holt et al., 2013) and participate to the formation of coagula (Caroli et al., 2009) and curd syneresis (Pearse and Mackinlay, 1989; Everard et al., 2011). However, the efficiency of capturing milk fat globules depends on the rapidity of clotting and the strength of the curd that in turn depend on casein genetic variants (Alipanah and Kalashnikova, 2007). The strong signal on BTA11 confirms previous findings, in which significant associations were identified for the casein variants (α, β and κ) as well as for β-lactoglobulin content and casein index (Schopen et al., 2011). The $REC_{PROTEIN}$ is phenotypically very similar to the casein index. Nevertheless, since part of the caseins are lost in the whey, $REC_{PROTEIN}$ is able to capture more variation compared to the casein index.

Albeit the relative success of GWAS to identify chromosomic regions related to the phenotypes under study, association studies alone may provide with a limited understanding of the complex nature of quantitative traits. Basic problems and limitations related to GWAS were summarized in the 3rd chapter. One way to alleviate some of the problematic features of GWAS and to deepen the knowledge on the biological background of the phenotypes under study is to move up the analysis, from the SNP to the gene level, conducting gene-set enrichment and pathway-based analysis. The pathway analysis constituted the 3rd contribution.

Pathway analysis

Our pathway analysis, based on the previous GWAS results on the MCP-CF_t and the CY-REC phenotypes, revealed 21 GO and 17 KEGG categories significantly with 7 of the traits (RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS} and REC_{ENERGY}). The significantly enriched categories included calcium signaling pathway, salivary secretion, metabolic pathways, carbohydrate digestion and absorption, the tight junction and the phosphatidylinositol pathways, as well as pathways related to the bovine mammary gland health status, and contained a total of 150 genes located in all chromosomes but 9, 20, and 27.

Nevertheless, pathway analysis has, in turn, its own restrictions. For instance, publicly available ontologies and pathways in cattle are still limited (compared to human) and/or not all are well described. Thus, some of our results may be misleading, especially when the detected genes are involved in various biological processes. It is likely that, when more complete gene sets become available, more competitive pathways might be detected and the power to identify genomic regions influencing these traits might increase (Fan et al., 2015). In this respect, transcriptomic methods (e.g. RNA-seq) may represent a useful tool to complement the present

analysis and validate the achieved biological information. More details in the methodological aspects, novel applications in livestock as well as further considerations concerning the gene-set enrichment and pathway analysis can be found in review studies (Cantor et al., 2010; Wang et al., 2011; Charitou et al., 2016) as well as in the recent works of (Fontanesi, 2016) and (Kent, 2016).

Investigating the use of latent variables, underlying cheese-related traits, in breeding programs

Our genomic work was a step-by-step integrated approach for the detection of genomic regions associated to cheese-related phenotypes. Nevertheless, the large number of variables required to describe the cow's ability to produce cheese poses restrictions in developing selection indices, and thereby in selection decisions for breeding purposes. It is interesting to note that in some countries breeding goals include more than 40 traits that are commonly recorded (Banos, 2010). With the recent technological advances, such as the spectral data, it is very likely that this number will drastically increase in the nearest future, (Boichard and Brochard, 2012). To tackle the problem of phenotypic dimensionality and complexity, factor analysis was proposed (chapters 4 and 5). The question to be addressed in those two chapters was: if a small set of factors could replace measured phenotypes in a breeding goal targeting in high cheese yield production. To this purpose, a set of 26 traits related to milk yield and quality (including milk protein fractions), MCP-CF_t and CY-REC traits was analyzed. Ten mutual orthogonal factors were extracted. Those latent variables captured basic concepts of the "*cheese-making*" process and retained 74% of the original variability.

Further, ANOVA, genetic analyses, GWAS and gene-set enrichment and pathways-based analyses were carried out using the obtained factors as phenotypes. In general, results were coherent to the given name of the Fs. Moreover, GWAS and pathways results were in accordance

to the previous findings based on the individual MCP-CF_t and CY-REC traits. Interestingly, using Fs as phenotypes in GWAS, same high peaks were detected on BTA6 (~87.4Mbp) and at the tail of BTA11 (~104.3Mbp) associated with $F6_{\beta\text{-}\kappa\text{-CN}}$ and $F1_{\text{Cheese N}}$. Furthermore, biological pathways associated with $F1_{\%CY}$, $F4_{\text{Cheese N}}$, $F8_{\alpha s_2\text{-CN}}$ and $F10_{\alpha\text{-LA}}$ were mainly related to the broader categories of ion activity, neurons and the tight junction. An important finding is that traits with lower loadings on the factors retained their signals in the genome. This enhance the evidence that the factors can be considered as a mixture of phenotypes and not only descriptors of the phenotype(s) with the highest loadings on them. This finding, however, questions the threshold (loading > |0.6|is a commonly used value) of the factor loadings which is used as a criterion to name the factors.

Future perspectives

Our genomic analyses was based on a panel of 50k SNP. Higher density SNP panels or sequence data, together with an increased number of observations, are expected to help in narrowing down the associated genomic regions in GWAS (Bouwman et al., 2014). Moreover, our GWAS model was based on single marker regression, which assumes a simplified theory of the genomic background of phenotypes. However, the polygenic and complex nature of quantitative traits, as possibly are the MCP-CF_t and CY-REC traits, is probably the rule rather the exception for traits commonly recorded in dairy cattle breeding. This complex genomic content is, however, ignored by single marker regression GWAS. To address this limitation, non-additive effects, e.g. by applying GWAS incorporating SNP by SNP interactions, could be explored. This might provide with an extra level of knowledge of the genomic background of the traits (Boichard and Brochard, 2012; Aliloo et al., 2015; Bolormaa et al., 2015). Consequently, denser marker

panels (or sequence data) together with non-additive GWAS, might provide a better substrate for biological pathway analysis. Moreover, more complete livestock gene ontology and pathway databases, is also expected to improve pathway analysis in livestock animals (Fan et al., 2015).

For future research, integrating all the available information into a structural equation model is suggested. Structural equation models are commonly used to study relationships among phenotypes, simultaneously. Moreover, they can handle latent variables, thus modeling complex correlation structures in a reduced data space. In animal breeding structural equation models have been used providing with encouraging results, for e.g., in analyzing carcass traits in swine (Peñagaricano et al., 2015) and studying bovine mastitis (Detilleux et al., 2013). In addition, our results could be further used and tested in genomic prediction models.

To our knowledge this is the first – large scale – genomic study investigating milk coagulation, curd firmness and individual bovine cheese yield traits. The road for fine mapping of genomic regions associated to bovine cheese yield is ahead. Partition of a complex process, as is the cheese yield, into different components has been shown as a useful tool for connecting phenomics to their genomic counterpart. Replication of all of our results from independent, and ideally of large scale, studies remains crucial.

Figure 1 Manhattan plot of *P*-values for the genome wide association studies (GWAS) on *Bos taurus* autosome 6 (BTA6).

Description: Traits showed significant associations on BTA6 were RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Prot % = protein percent; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}).

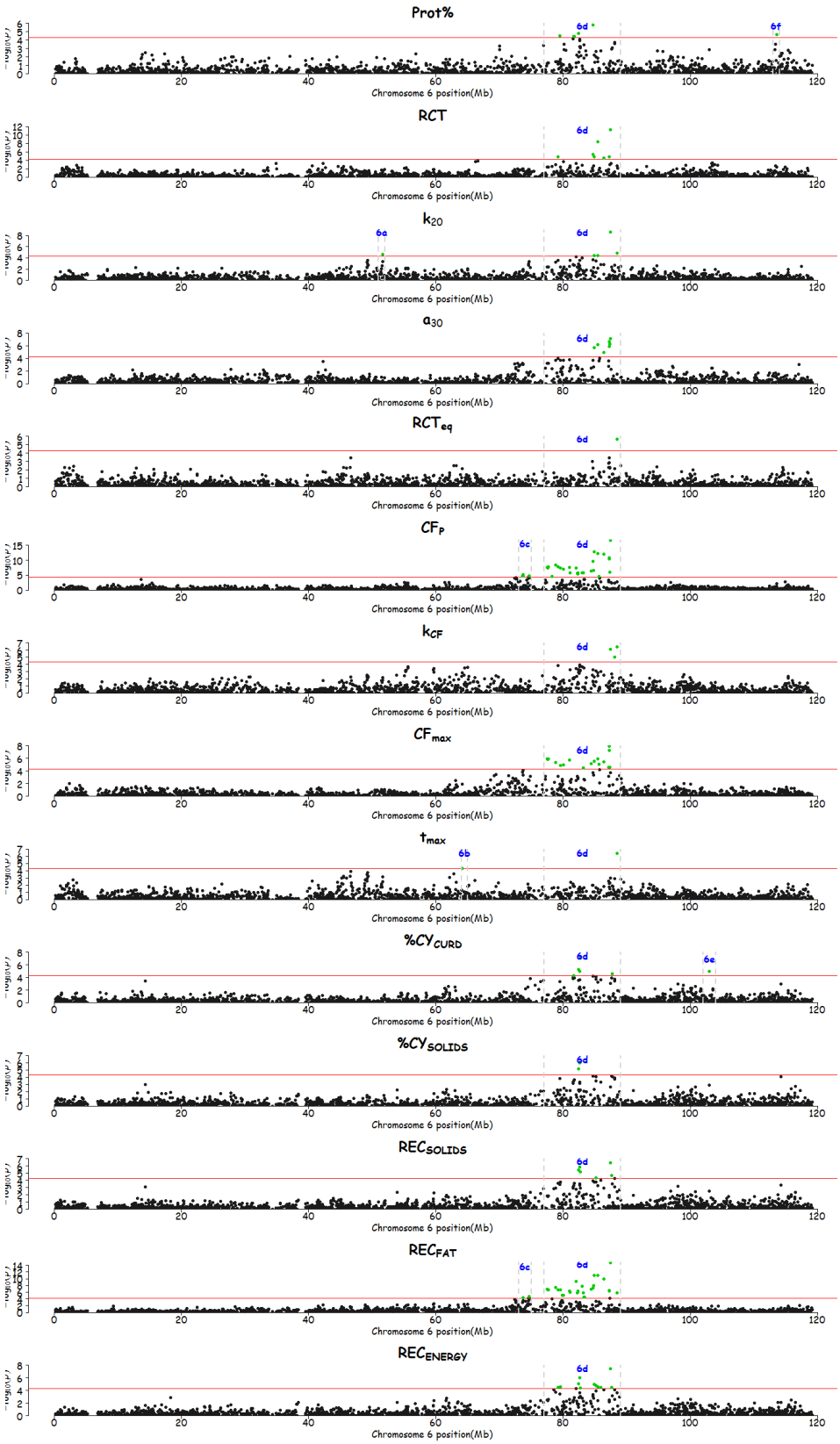

Prot%
6d
6f
RCT
6d
k20
6a
6d
a30
6d
RCTeq
6d
CFP
6c
6d
kCF
6d
CFmax
6d
tmax
6b
6d
%CYCURD
6d
6e
%CYSOLIDS
6d
RECSOLIDS
6d
RECFAT
6c
6d
RECENERGY
6d
Chromosome 6 position(Mb)

Figure 2. Manhattan plot of *P*-values for the genome wide association studies (GWAS) on *Bos taurus* autosomes (BTA) 11.

Description: RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; t_{max} = time to maximum curd firmness (min); $REC_{PROTEIN}$= Protein of the curd as percentage of the protein of the milk processed.

The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}).

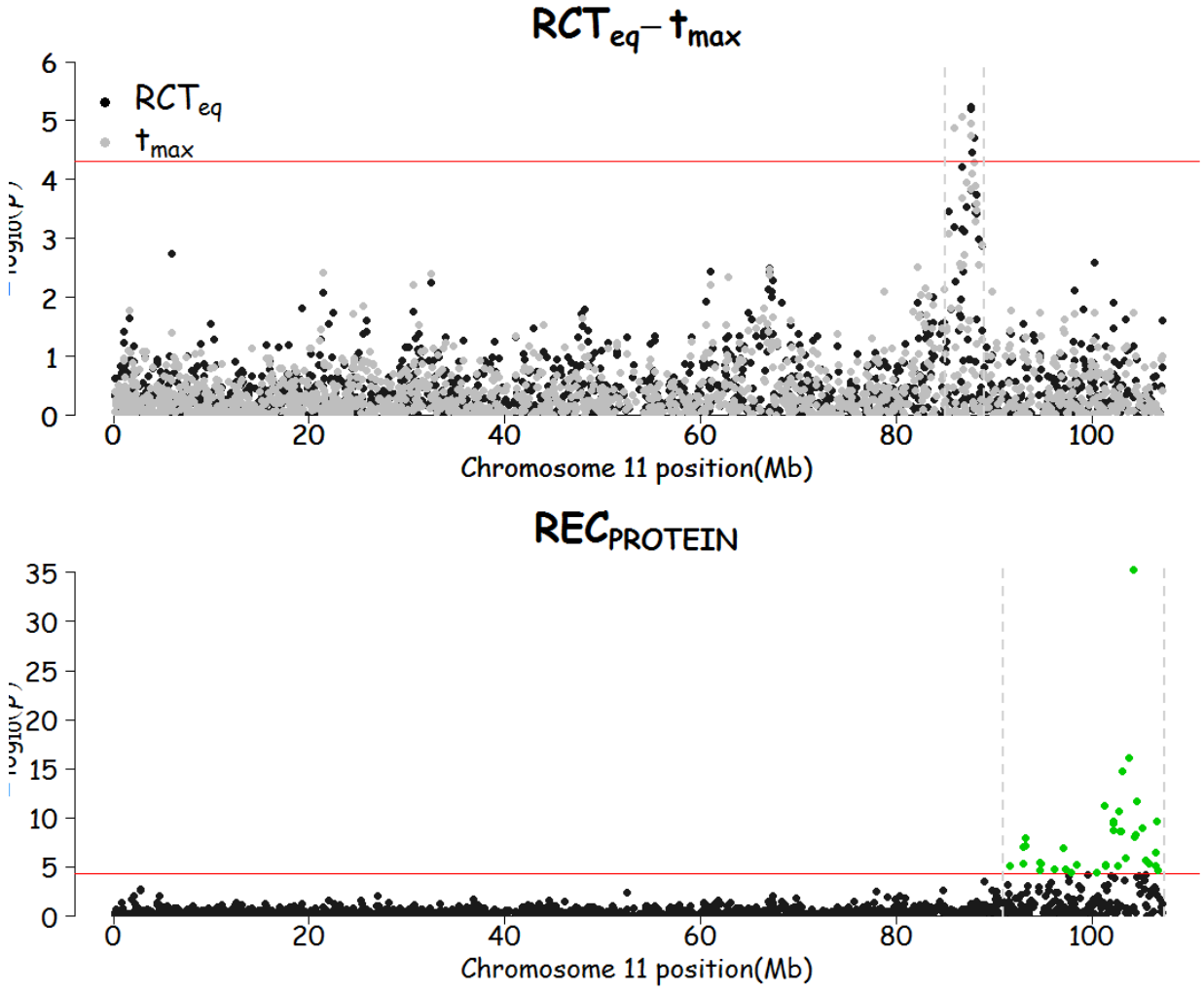
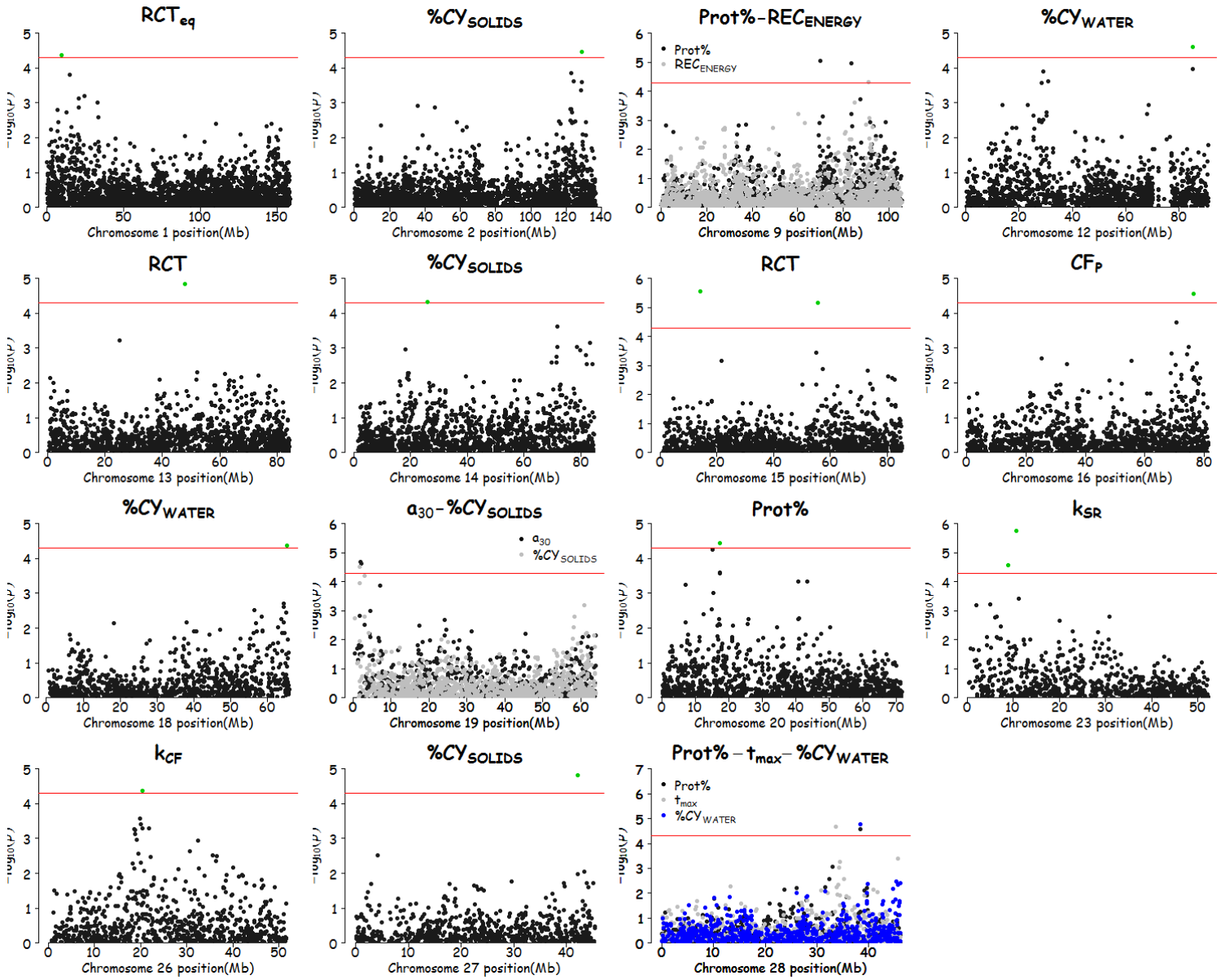


Figure 3 Manhattan plot of *P*-values for the genome wide association studies (GWAS) on *Bos taurus* autosomes (BTA) 1, 2, 9, 12, 13, 14, 15, 16, 18, 19, 20, 23, 26, 27 and 28.

Description: RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Prot % = protein percent; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); t_{max} = time to maximum curd firmness (min); $\%CY_{SOLIDS}$ = weight of curd solids as percentage of weight of milk processed; $\%CY_{WATER}$ = weight of curd water as percentage of weight of milk processed; REC_{ENERGY} = energy of the curd as percentage of the energy of the milk processed.

The red horizontal lines indicate a –log10(*P*-values) of 4.30 (corresponding to *P*-value = 5×10^{-5}).


RCT$_{eq}$
%CY$_{SOLIDS}$
Prot%-REC$_{ENERGY}$
Prot%
REC$_{ENERGY}$
%CY$_{WATER}$
RCT
%CY$_{SOLIDS}$
RCT
CF$_P$
%CY$_{WATER}$
a$_{30}$-%CY$_{SOLIDS}$
a$_{30}$
%CY$_{SOLIDS}$
Prot%
k$_{SR}$
k$_{CF}$
%CY$_{SOLIDS}$
Prot% – t$_{max}$– %CY$_{WATER}$
Prot%
t$_{max}$
%CY$_{WATER}$
−log$_{10}$(P)
Chromosome 1 position(Mb)
Chromosome 2 position(Mb)
Chromosome 9 position(Mb)
Chromosome 12 position(Mb)
Chromosome 13 position(Mb)
Chromosome 14 position(Mb)
Chromosome 15 position(Mb)
Chromosome 16 position(Mb)
Chromosome 18 position(Mb)
Chromosome 19 position(Mb)
Chromosome 20 position(Mb)
Chromosome 23 position(Mb)
Chromosome 26 position(Mb)
Chromosome 27 position(Mb)
Chromosome 28 position(Mb)

REFERENCES

Abdalla, E., F. Peñagaricano, T. Byrem, K. Weigel and G. Rosa. 2016. Genome-wide association mapping and pathway analysis of leukosis incidence in a US holstein cattle population. Anim. Genet.

Akers, R. M. 2016. Lactation and the Mammary Gland. John Wiley & Sons, .

Ali, A. and G. Shook. 1980. An optimum transformation for somatic cell concentration in milk. J. Dairy Sci. 63:487-490.

Aliloo, H., J. E. Pryce, O. González-Recio, B. G. Cocks and B. J. Hayes. 2015. Validation of markers with non-additive effects on milk yield and fertility in holstein and jersey cows. BMC Genetics. 16:1.

Alipanah, M. and L. Kalashnikova. 2007. Influence of k-casein genetic variant on cheese making ability. J Anim Vet Advances. 6:855-857.

Amin, N., C. M. Van Duijn and Y. S. Aulchenko. 2007. A genomic background based method for association analysis in related individuals. PloS One. 2:1274.

Aschard, H., B. Vilhjálmsson, N. Greliche, P. Morange, D. Trégouët and P. Kraft. 2014. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. The American Journal of Human Genetics. 94:662-676.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight and J. T. Eppig. 2000. Gene ontology: Tool for the unification of biology. Nat. Genet. 25:25-29.

Auldist, M. and I. Hubble. 1998. Effects of mastitis on raw milk and dairy products. Aust. J. Dairy Technol. 53:28.

Banks, J. M. 2007. Cheese yield. Pages 100-114 in Cheese Problems Solved. P. L. H. McSweeny ed. Woodhead PublishingLtd., Cambridge, UK.

Banos, G. 2010. Past, present and future of international genetic evaluations of dairy bulls. in Proceedings of the 9th world congress of genetics applied to livestock production, Leipzig, Germany.

Berry, S. D., N. Lopez-Villalobos, E. M. Beattie, S. R. Davis, L. F. Adams, N. L. Thomas, A. Ankersmit-Udy, A. M. Stanfield, K. Lehnert, H. E. Ward, J. A. Arias, R. J. Spelman and R. G. Snell. 2010. Mapping a quantitative trait locus for the concentration of beta-lactoglobulin in milk, and the effect of beta-lactoglobulin genetic variants on the composition of milk from holstein-friesian x jersey crossbred cows. N. Z. Vet. J. 58:1-5.

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.

Bittante, G. 2011. Modeling rennet coagulation time and curd firmness of milk. J. Dairy Sci. 94:5821-5832.

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.

Bittante, G., C. Cipolat-Gotet, F. Malchiodi, E. Sturaro, F. Tagliapietra, S. Schiavon and A. Cecchinato. 2015. Effect of dairy farming system, herd, season, parity, and days in milk on modeling of the coagulation, curd firming, and syneresis of bovine milk. J. Dairy Sci. 98:2759-2774.

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.
Bittante, G., A. Ferragina, C. Cipolat-Gotet and A. Cecchinato. 2014. Comparison between genetic parameters of cheese yield and nutrient recovery or whey loss traits measured from individual model cheese-making methods or predicted from unprocessed bovine milk samples using fourier-transform infrared spectroscopy. J. Dairy Sci. 97:6560-6572.
Bittante, G., M. Penasa and A. Cecchinato. 2012. Invited review: Genetics and modeling of milk coagulation properties. J. Dairy Sci. 95:6843-6870.
Bland, J. H., A. S. Grandison and C. C. Fagan. 2015. Evaluation of milk compositional variables on coagulation properties using partial least squares. J. Dairy Res. 82:8-14.
Bobbo, T., C. Cipolat-Gotet, G. Bittante and A. Cecchinato. 2016a. The nonlinear effect of somatic cell count on milk composition, coagulation properties, curd firmness modeling, cheese yield, and curd nutrient recovery. J. Dairy Sci. 99:5104-5119.
Bobbo, T., C. Cipolat-Gotet, G. Bittante and A. Cecchinato. 2016b. The nonlinear effect of somatic cell count on milk composition, coagulation properties, curd firmness modeling, cheese yield, and curd nutrient recovery. J. Dairy Sci.
Bobe, G., D. Beitz, A. Freeman and G. Lindberg. 1999. Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. J. Dairy Sci. 82:2797-2804.
Boichard, D. and M. Brochard. 2012. New phenotypes for new breeding goals in dairy cattle. Animal. 6:544-550.
Bollen, K. A. 2014. Structural Equations with Latent Variables. John Wiley & Sons, .
Bolormaa, S., J. E. Pryce, Y. Zhang, A. Reverter, W. Barendse, B. J. Hayes and M. E. Goddard. 2015. Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. Genetics Selection Evolution. 47:1.
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.
Bonfatti, V., G. Di Martino, A. Cecchinato, L. Degano and P. Carnier. 2010a. Effects of β-κ-casein (CSN2-CSN3) haplotypes, β-lactoglobulin (BLG) genotypes, and detailed protein composition on coagulation properties of individual milk of simmental cows. J. Dairy Sci. 93:3809-3817.
Bonfatti, V., G. Di Martino, A. Cecchinato, D. Vicario and P. Carnier. 2010b. 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.
Bonfatti, V., M. Giantin, M. Gervaso, A. Coletta, M. Dacasto and P. Carnier. 2012. Effect of CSN1S1-CSN3 (α S1-κ-casein) composite genotype on milk production traits and milk coagulation properties in mediterranean water buffalo. J. Dairy Sci. 95:3435-3443.
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. Journal of Chromatography A. 1195:101-106.
Bonfatti, V., G. Di Martino, A. Cecchinato, L. Degano and P. Carnier. 2010c. Effects of β-κ-casein (CSN2-CSN3) haplotypes, β-lactoglobulin (BLG) genotypes, and detailed protein composition on coagulation properties of individual milk of simmental cows. J. Dairy Sci. 93:3809-3817.

Bouwman, A. C., M. H. Visker, J. M. van Arendonk and H. Bovenhuis. 2014. Fine mapping of a quantitative trait locus for bovine milk fat composition on bos taurus autosome 19. J. Dairy Sci. 97:1139-1149.
Burton, P. R., D. G. Clayton, L. R. Cardon, N. Craddock, P. Deloukas, A. Duncanson, D. P. Kwiatkowski, M. I. McCarthy, W. H. Ouwehand and N. J. Samani. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 447:661-678.
Caffin, J. P., B. Poutrel and P. Rainard. 1985. Physiological and pathological factors influencing bovine α-lactalbumin and β-lactoglobulin concentrations in milk. J. Dairy Sci. 68:1087-1094.
Cantor, R. M., K. Lange and J. S. Sinsheimer. 2010. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. The American Journal of Human Genetics. 86:6-22.
Caravaca, F., J. L. Ares, J. Carrizosa, B. Urrutia, F. Baena, J. Jordana, B. Badaoui, A. Sànchez, A. Angiolillo and M. Amills. 2011. Effects of α s1-casein (CSN1S1) and κ-casein (CSN3) genotypes on milk coagulation properties in murciano-granadina goats. J. Dairy Res. 78:32-37.
Caroli, A. M., S. Chessa and G. J. Erhardt. 2009a. Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. J. Dairy Sci. 92:5335-5352.
Caroli, A. M., S. Chessa and G. J. Erhardt. 2009b. Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. J. Dairy Sci. 92:5335-5352.
Cecchinato, A., C. Ribeca, M. Penasa, C. Cipolat Gotet, M. De Marchi, A. Maurmayr and G. Bittante. 2012a. Associations between single nucleotide polymorphisms in multiple candidate genes on milk yield, composition, coagulation properties and individual cheese yield in brown swiss cows. J. Dairy Sci. 95(Suppl. 2):402-b (Abstr.).
Cecchinato, A., A. Albera, C. Cipolat-Gotet, A. Ferragina and G. Bittante. 2015a. Genetic parameters of cheese yield and curd nutrient recovery or whey loss traits predicted using fourier-transform infrared spectroscopy of samples collected during milk recording on holstein, brown swiss, and simmental dairy cows. J. Dairy Sci. 98:4914-4927.
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.
Cecchinato, A., A. Albera, C. Cipolat-Gotet, A. Ferragina and G. Bittante. 2015b. Genetic parameters of cheese yield and curd nutrient recovery or whey loss traits predicted using fourier-transform infrared spectroscopy of samples collected during milk recording on holstein, brown swiss, and simmental dairy cows. J. Dairy Sci. 98:4914-4927.
Cecchinato, A., S. Chessa, C. Ribeca, C. Cipolat-Gotet, T. Bobbo, J. Casellas and G. Bittante. 2015c. Genetic variation and effects of candidate-gene polymorphisms on coagulation properties, curd firmness modeling and acidity in milk from brown swiss cows. Animal. 9:1104-1112.
Cecchinato, A., C. Cipolat-Gotet, J. Casellas, M. Penasa, A. Rossoni and G. Bittante. 2013a. Genetic analysis of rennet coagulation time, curd-firming rate, and curd firmness assessed over an extended testing period using mechanical and near-infrared instruments. J. Dairy Sci. 96:50-62.
Cecchinato, A., C. Cipolat-Gotet, J. Casellas, M. Penasa, A. Rossoni and G. Bittante. 2013b. Genetic analysis of rennet coagulation time, curd-firming rate, and curd firmness assessed

over an extended testing period using mechanical and near-infrared instruments. J. Dairy Sci. 96:50-62.
Cecchinato, A., M. De Marchi, L. Gallo, G. Bittante and P. Carnier. 2009. Mid-infrared spectroscopy predictions as indicator traits in breeding programs for enhanced coagulation properties of milk. J. Dairy Sci. 92:5304-5313.
Cecchinato, A., M. Penasa, M. De Marchi, L. Gallo, G. Bittante and P. Carnier. 2011. Genetic parameters of coagulation properties, milk yield, quality, and acidity estimated using coagulating and noncoagulating milk information in brown swiss and holstein-friesian cows. J. Dairy Sci. 94:4205-4213.
Cecchinato, A., C. Ribeca, A. Maurmayr, M. Penasa, M. De Marchi, N. P. P. Macciotta, M. Mele, P. Secchiari, G. Pagnacco and G. Bittante. 2012b. Short communication: Effects of β-lactoglobulin, stearoyl-coenzyme A desaturase 1, and sterol regulatory element binding protein gene allelic variants on milk production, composition, acidity, and coagulation properties of brown swiss cows. J. Dairy Sci. 95:450-454.
Cerón-Rojas, J. J., J. Crossa, J. Sahagún-Castellanos, F. Castillo-González and A. Santacruz-Varela. 2006. A selection index method based on eigenanalysis. Crop Sci. 46:1711-1721.
Cerón-Rojas, J. J., J. Sahagún-Castellanos, F. Castillo-González, A. Santacruz-Varela and J. Crossa. 2008. A restricted selection index method based on eigenanalysis. Journal of Agricultural, Biological, and Environmental Statistics. 13:440-457.
Charitou, T., K. Bryan and D. J. Lynn. 2016. Using biological networks to integrate, visualize and analyze genomics data. Genetics Selection Evolution. 48:1-12.
Chessa, S., O. Bulgari, R. Rizzi, L. Calamari, P. Bani, S. Biffani and A. M. Caroli. 2014. Selection for milk coagulation properties predicted by fourier transform infrared spectroscopy in the italian holstein-friesian breed. J. Dairy Sci. 97:4512-4521.
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.
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.
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 and L. K. Matukumalli. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US holstein cows. BMC Genomics. 12:1.
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. Livestock Science. 188:124-132.
Dadousis, C., S. Biffani, C. Cipolat-Gotet, E. Nicolazzi, A. Rossoni, E. Santus, G. Bittante and A. Cecchinato. 2016a. 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.
Dadousis, C., S. Biffani, C. Cipolat-Gotet, E. Nicolazzi, G. Rosa, D. Gianola, A. Rossoni, E. Santus, G. Bittante and A. Cecchinato. 2016b. Genome-wide association study for cheese yield and curd nutrient recovery in dairy cows. J. Dairy Sci. http://dx.doi.org/10.3168/jds.2016-11586
Dadousis, C., G. Bittante, C. Cipolat-Gotet and A. Cecchinato. 2016c. Pathway-based genome-

wide association analysis of milk coagulation properties, curd firmness, cheese yield and curd nutrient recovery in dairy cattle. J. Dairy Sci. http://dx.doi.org/10.3168/jds.2016-11587.
Dadousis, C., C. Cipolat-Gotet, S. Biffani, E. L. Nicolazzi, G. J. M. Rosa, D. Gianola, G. Bittante and A. Cecchinato. 2015. Genome-wide association study for cheese yield and curd nutrients recovery in bovine milk. Page 358 in Book of abstracts of the 66th Annual meeting of the european federation of animal science, Warsaw, Poland. Wageningen Academic Publishers, The Netherlands.
de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler and M. P. L. Calus. 2013. Whole genome regression and prediction methods applied to plant and animal breeding. Genetics. 193:327-345.
Detilleux, J., L. Theron, J. Duprez, E. Reding, M. Humblet, V. Planchon, C. Delfosse, C. Bertozzi, J. Mainil and C. Hanzen. 2013. Structural equation models to estimate risk of infection and tolerance to bovine mastitis. Genetics Selection Evolution. 45:1.
Devold, T. G., R. Nordbø, T. Langsrud, C. Svenning, M. J. Brovold, E. S. Sørensen, B. Christensen, T. Ådnøy and G. E. Vegarud. 2011. Extreme frequencies of the αs1-casein "null" variant in milk from norwegian dairy goats—Implications for milk composition, micellar size and renneting properties. Dairy Science & Technology. 91:39-51.
Dosogne, H., F. Vangroenweghe, J. Mehrzad, A. M. Massart-Leën and C. Burvenich. 2003. Differential leukocyte count method for bovine low somatic cell count milk. J. Dairy Sci. 86:828-834.
Duchemin, S. I., M. Glantz, D. De Koning, M. Paulsson and W. F. Fikse. 2016. Identification of QTL on chromosome 18 associated with non-coagulating milk in swedish red cows. Frontiers in Genetics. 7:57.
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. Nature Protocols. 4:1184-1191.
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.
Dziuban, C. D. and E. C. Shirkey. 1974. When is a correlation matrix appropriate for factor analysis? some decision rules. Psychol. Bull. 81:358.
Edwards, S. M., B. Thomsen, P. Madsen and P. Sørensen. 2015. Partitioning of genomic variance reveals biological pathways associated with udder health and milk production traits in dairy cattle. Genetics Selection Evolution. 47:1.
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.
Everard, C. D., D. J. O'Callaghan, M. J. Mateo, M. Castillo, F. A. Payne and C. P. O'Donnell. 2011. Effects of milk composition, stir-out time, and pressing duration on curd moisture and yield. J. Dairy Sci. 94:2673-2679.
Fan, H., Y. Wu, X. Zhou, J. Xia, W. Zhang, Y. Song, F. Liu, Y. Chen, L. Zhang, X. Gao, H. Gao and J. Li. 2015. Pathway-based genome-wide association studies for two meat production traits in simmental cattle. Sci. Rep. 5:18389.
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. 2012. Genome-wide association study of clinical dimensions of schizophrenia: Polygenic effect on disorganized symptoms. Am. J. Psychiatry. 169:1309-1317.
Ferragina, A., C. Cipolat-Gotet, A. Cecchinato and G. Bittante. 2013. The use of fourier-transform infrared spectroscopy to predict cheese yield and nutrient recovery or whey loss traits from unprocessed bovine milk samples. J. Dairy Sci. 96:7980-7990.
Ferragina, A., G. de los Campos, A. I. Vazquez, A. Cecchinato and G. Bittante. 2015. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. J. Dairy Sci. 98:8133-8151.
Ferreiro, T., S. Martínez, L. Gayoso and J. Rodríguez-Otero. 2016. Evolution of phospholipid contents during the production of quark cheese from buttermilk. J. Dairy Sci. 99:4154-4159.
Fontanesi, L. 2016. Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. Animal Frontiers. 6:73-79.
Food and Agriculture Organization of the,United Nations. 2015. FAOSTAT.
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.
Gambra, R., F. Peñagaricano, J. Kropp, K. Khateeb, K. Weigel, J. Lucey and H. Khatib. 2013. Genomic architecture of bovine κ-casein and β-lactoglobulin. J. Dairy Sci. 96:5333-5343.
Gebreyesus, G., M. Lund, L. Janss, N. Poulsen, L. Larsen, H. Bovenhuis and A. Buitenhuis. 2016. Short communication: Multi-trait estimation of genetic parameters for milk protein composition in the danish holstein. J. Dairy Sci.
GenABEL project developers. 2013. GenABEL: Genome-wide SNP association analysis. .R package version 1.8-0.:http://CRAN.R-project.org/package=GenABEL;.
Gengler, N., G. R. Wiggans, J. R. Wright, H. D. Norman and C. W. Wolfe. 1997. Application of canonical transformation with missing values to multitrait evaluation of jersey type. J. Dairy Sci. 80:2563-2571.
Glantz, M., H. Lindmark Månsson, H. Stålhammar and M. Paulsson. 2011. Effect of polymorphisms in the leptin, leptin receptor, and acyl-coenzyme A:Diacylglycerol acyltransferase 1 (DGAT1) genes and genetic polymorphism of milk proteins on cheese characteristics. J. Dairy Sci. 94:3295-3304.
Glantz, M., H. Lindmark Månsson, M. Paulsson and H. Stålhammar. 2012. Genomic selection in relation to bovine milk composition and processability. J. Dairy Res. 79:53-59.
Goddard, M. E. and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10:381-391.
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 and M. Paulsson. 2015. Bovine chromosomal regions affecting rheological traits in rennet-induced skim milk gels. J. Dairy Sci. 98:1261-1272.
Gustavsson, F., M. Glantz, A. J. Buitenhuis, H. Lindmark-Månsson, H. Stålhammar, A. Andrén and M. Paulsson. 2014a. Factors influencing chymosin-induced gelation of milk from individual dairy cows: Major effects of casein micelle size and calcium. Int. Dairy J. 39:201-208.
Gustavsson, F., M. Glantz, N. A. Poulsen, L. Wadsö, H. Stålhammar, A. Andrén, H. L. Månsson, L. B. Larsen, M. Paulsson and W. Fikse. 2014b. Genetic parameters for rennet-and acid-induced coagulation properties in milk from swedish red dairy cows. J. Dairy Sci. 97:5219-

5229.
Ha, N., J. J. Gross, A. van Dorland, J. Tetens, G. Thaller, M. Schlather, R. Bruckmaier and H. Simianer. 2015. Gene-based mapping and pathway analysis of metabolic traits in dairy cows. PloS One. 10:e0122325.
Haenlein, G., L. Schultz and J. Zikakis. 1973. Composition of proteins in milk with varying leucocyte contents. J. Dairy Sci. 56:1017-1024.
Hayes, B. 2013. Overview of statistical methods for genome-wide association studies (GWAS). Genome-Wide Association Studies and Genomic Prediction. 149-169.
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.
Holland, J. W., H. C. Deeth and P. F. Alewood. 2005. Analysis of o-glycosylation site occupancy in bovine κ-casein glycoforms separated by two-dimensional gel electrophoresis. Proteomics. 5:990-1002.
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 foods1. J. Dairy Sci. 96:6127-6146.
Houle, D., D. R. Govindaraju and S. Omholt. 2010. Phenomics: The next challenge. Nat. Rev. Genet. 11:855-866.
Hurtaud, C., H. Rulquin, M. Delaite and R. Vérité. 1995. Appréciation de l'aptitude fromagère des laits de vaches individuels. tests d'aptitude fromagère et rendement fromager de fabrication. Ann. Zootech. 44:385-398.
Ikonen, T., S. Morri, A. Tyrisevä, O. Ruottinen and M. Ojala. 2004a. Genetic and phenotypic correlations between milk coagulation properties, milk production traits, somatic cell count, casein content, and pH of milk. J. Dairy Sci. 87:458-467.
Ikonen, T., K. Ahlfors, R. Kempe, M. Ojala and O. Ruottinen. 1999. Genetic parameters for the milk coagulation properties and prevalence of noncoagulating milk in finnish dairy cows. J. Dairy Sci. 82:205-214.
Ikonen, T., S. Morri, A. -. Tyrisevä, O. Ruottinen and M. Ojala. 2004b. Genetic and phenotypic correlations between milk coagulation properties, milk production traits, somatic cell count, casein content, and pH of milk. J. Dairy Sci. 87:458-467.
International Committee for Animal Recording (ICAR). 2012. International agreement of recording practices—Guidelines approved by the general assembly held in cork, ireland on june 2012. ICAR, Rome, Italy; 2012.
Iso-Touru, T., G. Sahana, B. Guldbrandtsen, M. Lund and J. Vilkki. 2016. Genome-wide association analysis of milk yield traits in nordic red cattle using imputed whole genome sequence variants. BMC Genetics. 17:1.
Jenko, J., G. Gorjanc, M. A. Cleveland, R. K. Varshney, C. B. A. Whitelaw, J. A. Woolliams and J. M. Hickey. 2015. Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. Genetics Selection Evolution. 47:1.
Jolliffe, I. T. 2002. Principal Component Analysis. Springer, .
Jõudu, I., M. Henno, T. Kaart, T. Püssa and O. Kärt. 2008. The effect of milk protein contents on the rennet coagulation properties of milk from individual dairy cows. Int. Dairy J. 18:964-967.
Kaiser, H. F. and J. Rice. 1974. Little jiffy, mark IV. Educational and Psychological Measurement.

Kent, J. W. 2016. Pathway-based analyses. BMC Genetics. 17:43.
Kitchen, B. 1981. Review of the progress of dairy science: Bovine mastitis: Milk compositional changes and related diagnostic test. J. Dairy Res. 48:167-188.
Law, B. A. and A. Y. Tamime. 2011. Technology of Cheesemaking. John Wiley & Sons, .
Lilbæk, H. M., M. Broe, E. Høier, T. Fatum, R. Ipsen and N. Sørensen. 2006. Improving the yield of mozzarella cheese by phospholipase treatment of milk. J. Dairy Sci. 89:4114-4125.
Lopez, C., V. Briard-Bion, O. Ménard, E. Beaucher, F. Rousseau, J. Fauquant, N. Leconte and B. Robert. 2011. Fat globules selected from whole milk according to their size: Different compositions and structure of the biomembrane, revealing sphingomyelin-rich domains. Food Chem. 125:355-368.
López-Expósito, I., J. Á Gómez-Ruiz, L. Amigo and I. Recio. 2006. Identification of antibacterial peptides from ovine α s2-casein. Int. Dairy J. 16:1072-1080.
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.
Malacarne, M., P. Franceschi, P. Formaggioni, S. Sandri, P. Mariani and A. Summer. 2014. Influence of micellar calcium and phosphorus on rennet coagulation properties of cows milk. J. Dairy Res. 81:129-136.
Manca, M., J. Serdino, G. Gaspa, P. Urgeghe, I. Ibba, M. Contu, P. Fresi and N. Macciotta. 2016. Derivation of multivariate indices of milk composition, coagulation properties, and individual cheese yield in dairy sheep. J. Dairy Sci. 99:4547-4557.
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.
McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis and J. N. Hirschhorn. 2008. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. Nature Reviews Genetics. 9:356-369.
Mele, M., N. P. P. Macciotta, A. Cecchinato, G. Conte, S. Schiavon and G. Bittante. under review. Multivariate factor analysis of detailed milk fatty acid profile: Effects of dairy system, feeding, herd, parity, and stage of lactation . J Dairy Sci.
Miglior, F., B. Muir and B. Van Doormaal. 2005. Selection indices in holstein cattle of various countries. J. Dairy Sci. 88:1255-1263.
Miglior, F., A. Sewalem, J. Jamrozik, J. Bohmanova, D. Lefebvre and R. Moore. 2007. Genetic analysis of milk urea nitrogen and lactose and their relationships with other production traits in canadian holstein cattle. J. Dairy Sci. 90:2468-2479.
Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet and D. Lee. 2002. BLUPF90 and related programs (BGF90). in World congress on genetics applied to livestock production, .
Molee, A., C. Poompramun and P. Mernkrathoke. 2015. Effect of casein genes - beta-LGB, DGAT1, GH, and LHR - on milk production and milk composition traits in crossbred holsteins. Genet. Mol. Res. 14:2561-2571.
Mrode, R. A. 2005. Linear Models for the Prediction of Animal Breeding Values. 2nd ed.
Nguyen, D. D. 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.
NRC. 2001. Nutrient Requirements of Dairy Cattle. . 7th rev. ed. ed. Natl. Acad. Press,

Washington, DC.
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 and S. R. Dahl. 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. Genetics Selection Evolution. 48:79.
Ostersen, S., J. Foldager and J. E. Hermansen. 1997. Effects of stage of lactation, milk protein genotype and body condition at calving on protein composition and renneting properties of bovine milk. J. Dairy Res. 64:207-219.
Othmane, M. H., J. A. Carriedo, L. F. de la Fuente Crespo and F. San Primitivo. 2002. An individual laboratory cheese-making method for selection in dairy ewes. Small Ruminant Research. 45:67-73.
Pearse, M. J. and A. G. Mackinlay. 1989. Biochemical aspects of syneresis: A review. J. Dairy Sci. 72:1401-1407.
Pellegrini, A., U. Thomas, N. Bramaz, P. Hunziker and R. von Fellenberg. 1999. Isolation and identification of three bactericidal domains in the bovine α-lactalbumin molecule. Biochimica Et Biophysica Acta (BBA)-General Subjects. 1426:439-448.
Peñagaricano, F., B. Valente, J. Steibel, R. Bates, C. Ernst, H. Khatib and G. Rosa. 2015. Searching for causal networks involving latent variables in complex traits: Application to growth, carcass, and meat quality traits in pigs. J. Anim. Sci. 93:4617-4623.
Peñagaricano, F., K. Weigel, G. J. Rosa and H. Khatib. 2013. Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. Frontiers in Genetics. 3:307.
Peng, G., L. Luo, H. Siu, Y. Zhu, P. Hu, S. Hong, J. Zhao, X. Zhou, J. D. Reveille and L. Jin. 2010. Gene and pathway-based second-wave analysis of genome-wide association studies. European Journal of Human Genetics. 18:111-117.
Pickrell, J. K., J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J. Veyrieras, M. Stephens, Y. Gilad and J. K. Pritchard. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 464:768-772.
Poulsen, N., H. Jensen and L. Larsen. 2016. Factors influencing degree of glycosylation and phosphorylation of caseins in individual cow milk samples. J. Dairy Sci. 99:3325-3333.
Proudfoot, C., D. F. Carlson, R. Huddart, C. R. Long, J. H. Pryor, T. J. King, S. G. Lillico, A. J. Mileham, D. G. McLaren and C. B. A. Whitelaw. 2015. Genome edited sheep and cattle. Transgenic Res. 24:147-153.
R Core Team. 2013. R: A language and environment for statistical computing.2013:URL http://www.R-project.org/.
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. 9:1.
Raven, L., B. G. Cocks, M. E. Goddard, J. E. Pryce and B. J. Hayes. 2014. Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genetics Selection Evolution. 46:1.
Revelle, W. 2014. Psych: Procedures for personality and psychological research. Northwestern

University, Evanston.R Package Version. 1:.
Rombaut, R., K. Dewettinck and J. Van Camp. 2007. Phospho-and sphingolipid content of selected dairy products as determined by HPLC coupled to an evaporative light scattering detector (HPLC–ELSD). Journal of Food Composition and Analysis. 20:308-312.
Rutten, M. J. M., H. Bovenhuis, K. A. Hettinga, H. J. F. van Valenberg and J. A. M. van Arendonk. 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. J. Dairy Sci. 92:6202-6209.
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.
Schaefer, J., R. Opgen-Rhein, V. Zuber, M. Ahdesmäki, A. Duarte Silva and K. Strimmer. 2013. Corpcor: Efficient estimation of covariance and (partial) correlation. R Package Version. 1:.
Schneider, F., W. Tomek and C. Gründker. 2006. Gonadotropin-releasing hormone (GnRH) and its natural analogues: A review. Theriogenology. 66:691-709.
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.
Shuster, D., R. Harmon, J. Jackson and R. Hemken. 1991. Suppression of milk production during endotoxin-induced mastitis. J. Dairy Sci. 74:3763-3774.
Silva, S. V. and F. X. Malcata. 2005. Caseins as source of bioactive peptides. Int. Dairy J. 15:1-15.
Stelwagen, K. and K. Singh. 2014. The role of tight junctions in mammary gland function. J. Mammary Gland Biol. Neoplasia. 19:131-138.
Stocco, G., C. Cipolat-Gotet, T. Bobbo, A. Cecchinato and G. and Bittante. 2016. Herd productivity and breed of cows affect milk composition, coagulation properties and curd firming and syneresis modeling. J. Dairy Sci. under review:.
Sturaro, E., E. Marchiori, G. Cocca, M. Penasa, M. Ramanzin and G. Bittante. 2013. Dairy systems in mountainous areas: Farm animal biodiversity, milk production and destination, and land use. Livestock Science. 158:157-168.
Sturaro, E., G. Cocca, L. Gallo, M. Mrad and M. Ramanzin. 2009. Livestock systems and farming styles in eastern italian alps: An on-farm survey. Italian Journal of Animal Science. 8:541-554.
Svishcheva, G. R., T. I. Axenovich, N. M. Belonogova, C. van Duijn M. and Y. S. Aulchenko. 2012. Rapid variance components-based method for whole-genome association analysis. Nat. Genet. 44:1166-1170.
The Bovine Genome Sequencing and Analysis Consortium, C. G. Elsik, R. L. Tellam and K. C. Worley. 2009. The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science. 324:522-528.
Todaro, M., M. L. Scatassa and P. Giaccone. 2005. Multivariate factor analysis of girgentana goat milk composition. Italian Journal of Animal Science. 4:403-410.
Turner, S. D. 2014. Qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv.
Tyrisevä, A., K. Elo, A. Kuusipuro, V. Vilva, I. Jänönen, H. Karjalainen, T. Ikonen and M. Ojala. 2008. Chromosomal regions underlying noncoagulation of milk in finnish ayrshire cows. Genetics. 180:1211-1220.
Van Eenennaam, A. L., K. A. Weigel, A. E. Young, M. A. Cleveland and J. C. Dekkers. 2014.

Applied animal genomics: Results from the field. Annu.Rev.Anim.Biosci. 2:105-139.
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.
Visscher, P. M., M. A. Brown, M. I. McCarthy and J. Yang. 2012. Five years of GWAS discovery. Am. J. Hum. Genet. 90:7-24.
Walsh, S., F. Buckley, D. Berry, M. Rath, K. Pierce, N. Byrne and P. Dillon. 2007. Effects of breed, feeding system, and parity on udder health and milking characteristics. J. Dairy Sci. 90:5767-5779.
Walstra, P., P. Walstra, J. T. Wouters and T. J. Geurts. 2014. Dairy Science and Technology. CRC press, .
Wang, L., P. Jia, R. D. Wolfinger, X. Chen and Z. Zhao. 2011. Gene set analysis of genome-wide association studies: Methodological issues and perspectives. Genomics. 98:1-8.
Wedholm, A., L. B. Larsen, H. Lindmark-Månsson, A. H. Karlsson and A. Andrén. 2006. Effect of protein composition on the cheese-making properties of milk from individual dairy cows. J. Dairy Sci. 89:3296-3305.
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.
Wickenden, J. A. and C. J. Watson. 2010. Key signalling nodes in mammary gland development and cancer. signalling downstream of PI3 kinase in mammary epithelium: A play in 3 akts. Breast Cancer Research. 12:1.
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.
Yang, L., Q. Yang, M. Yi, Z. Pang and B. Xiong. 2013. Effects of seasonal change and parity on raw milk composition and related indices in chinese holstein cows in northern china. J. Dairy Sci. 96:6863-6869.
Yellen, G. 2002. The voltage-gated potassium channels and their relatives. Nature. 419:35-42.
Young, M. D., M. J. Wakefield, G. K. Smyth and A. Oshlack. 2010. Method gene ontology analysis for RNA-seq: Accounting for selection bias. Genome Biol. 11:1.
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.
Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell and T. S. Sonstegard. 2009. A whole-genome assembly of the domestic cow, bos taurus. Genome Biol. 10:1.
Zucht, H., M. Raida, K. Adermann, H. Mägert and W. Forssmann. 1995. Casocidin-I: A casein-αs2 derived peptide exhibits antibacterial activity. FEBS Lett. 372:185-188.

APPENDIX I: Supplementary files

CHAPTER 1

Table S1. Genetic correlations among traits used in the first multi-trait animal model (MCP-set)

Trait[1]	Milk pH	RCT (min)	k_{20} (min)	a_{30} (mm)
Milk pH				
RCT (min)	$0.56^{(0.02)}$			
k_{20} (min)	$0.36^{(0.03)}$	$0.65^{(0.02)}$		
a_{30} (mm)	$-0.48^{(0.03)}$	$-0.88^{(0.01)}$	$-0.93^{(0.01)}$	

[1] RCT = rennet coagulation time of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness at 30 min after enzyme addition.
In parenthesis the standard errors of the estimates.

Table S2. Genetic correlations among traits used in the second multi-trait animal model (CF_t -set)

Trait[1]	Protein %	RCT_{eq} (min)	CF_P (mm)	k_{CF} (% x min^{-1})	k_{SR} (% x min^{-1})	CF_{max} (mm)	t_{max} (min)
Protein, %							
RCT_{eq} (min)	$0.00^{(0.12)}$						
CF_P (mm)	$0.53^{(0.11)}$	$0.70^{(0.11)}$					
k_{CF} (% x min^{-1})	$0.07^{(0.14)}$	$-0.74^{(0.09)}$	$-0.35^{(0.13)}$				
k_{SR} (% x min^{-1})	$-0.31^{(0.18)}$	$0.05^{(0.16)}$	$-0.79^{(0.08)}$	$0.21^{(0.17)}$			
CF_{max} (mm)	$0.51^{(0.12)}$	$-0.08^{(0.13)}$	$0.67^{(0.08)}$	$0.34^{(0.12)}$	$-0.74^{(0.08)}$		
t_{max} (min)	$-0.01^{(0.12)}$	$0.87^{(0.03)}$	$0.58^{(0.09)}$	$-0.87^{(0.03)}$	$-0.25^{(0.15)}$	$-0.11^{(0.12)}$	

[1] CF_t -set observations were obtained by modeling the curd-firming process over time: RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

In parenthesis the standard errors of the estimates.

Table S3. List of all significant SNP identified in the GWAS analyses sorted by chromosome

N	SNP	BTA	Location (bp)	Pc1df	LOG	effB	snp.V_P (%)	snp.V_a (%)	MAF	Trait[1]
1	BTA-76907-no-rs	0*	0	5.31E-06	5.27	-1.61	2%	14%	0.26	CF_P
2	ARS-BFGL-NGS-41048	1	9484167	4.27E-05	4.37	1.58	3%	11%	0.09	RCT_{eq}
3	ARS-BFGL-NGS-36707	6	86354888	1.23E-05	4.91	-2.94	1%	5%	0.14	a_{30}
4	Hapmap41098-BTA-86027	6	84889974	1.99E-06	5.70	-3.16	1%	6%	0.13	a_{30}
5	Hapmap24184-BTC-070077	6	87245049	1.29E-06	5.89	-2.87	1%	6%	0.17	a_{30}
6	BTA-111108-no-rs	6	85424500	6.17E-07	6.21	-3.30	2%	6%	0.13	a_{30}
7	Hapmap28023-BTC-060518	6	87201599	2.25E-07	6.65	-3.03	2%	6%	0.17	a_{30}
8	Hapmap33631-BTC-043555	6	87327708	5.26E-07	6.28	-2.44	2%	6%	0.35	a_{30}
9	Hapmap52348-rs29024684	6	87396306	7.26E-08	7.14	-2.85	2%	7%	0.23	a_{30}
10	Hapmap41098-BTA-86027	6	84889974	3.89E-05	4.41	-3.43	19%	112%	0.13	k_{20}
11	BTA-111108-no-rs	6	85424500	3.73E-05	4.43	-3.46	20%	116%	0.13	k_{20}
12	BTB-01451336	6	51669513	2.57E-05	4.59	-3.67	20%	118%	0.12	k_{20}
13	BTA-122637-no-rs	6	88442145	1.35E-05	4.87	-5.03	22%	127%	0.06	k_{20}
14	Hapmap52348-rs29024684	6	87396306	2.64E-09	8.58	-3.95	40%	232%	0.23	k_{20}
15	Hapmap52479-rs29018853	6	79203343	1.47E-05	4.83	1.07	1%	4%	0.24	RCT
16	Hapmap24184-BTC-070077	6	87245049	1.66E-05	4.78	1.22	1%	4%	0.17	RCT
17	Hapmap28023-BTC-060518	6	87201599	1.32E-05	4.88	1.21	1%	4%	0.17	RCT
18	ARS-BFGL-NGS-36707	6	86354888	2.39E-05	4.62	1.35	1%	4%	0.14	RCT
19	Hapmap41098-BTA-86027	6	84889974	1.39E-05	4.86	1.38	1%	4%	0.13	RCT
20	ARS-BFGL-NGS-114609	6	84713584	3.98E-06	5.40	1.18	1%	5%	0.23	RCT
21	BTA-111108-no-rs	6	85424500	3.92E-09	8.41	1.84	2%	7%	0.13	RCT
22	Hapmap52348-rs29024684	6	87396306	5.88E-12	11.23	1.72	2%	10%	0.23	RCT
23	Hapmap38629-BTA-76891	6	78289101	2.98E-05	4.53	-1.37	2%	12%	0.35	CF_P
24	Hapmap52482-ss46526125	6	73639744	1.55E-05	4.81	-1.66	2%	13%	0.20	CF_P
25	Hapmap43767-BTA-113302	6	85646902	3.18E-05	4.50	-1.39	2%	13%	0.38	CF_P
26	ARS-BFGL-NGS-17026	6	73734080	1.36E-05	4.87	-1.73	2%	13%	0.19	CF_P
27	Hapmap60182-rs29025531	6	74606760	1.96E-05	4.71	-1.72	2%	13%	0.19	CF_P

28	Hapmap43042-BTA-76779	6	73688640	8.14E-06	5.09	-1.75	2%	14%	0.20	CF_P
29	Hapmap23419-BTC-059652	6	82314634	3.73E-06	5.43	-1.62	3%	16%	0.29	CF_P
30	Hapmap26317-BTC-059618	6	82253271	3.04E-06	5.52	-1.63	3%	16%	0.29	CF_P
31	ARS-BFGL-NGS-63312	6	82965163	2.00E-06	5.70	-1.91	3%	16%	0.19	CF_P
32	Hapmap49297-BTA-76961	6	83147633	1.57E-06	5.80	-1.74	3%	16%	0.25	CF_P
33	Hapmap31932-BTC-042947	6	82350917	1.75E-06	5.76	-1.67	3%	17%	0.30	CF_P
34	Hapmap51938-BTA-21491	6	81057816	1.46E-06	5.83	-1.80	3%	17%	0.23	CF_P
35	Hapmap33631-BTC-043555	6	87327708	1.21E-06	5.92	-1.67	3%	18%	0.35	CF_P
36	ARS-BFGL-NGS-70112	6	84448550	3.25E-07	6.49	-1.97	3%	19%	0.22	CF_P
37	BTB-01393607	6	80062406	8.58E-08	7.07	-2.13	3%	20%	0.19	CF_P
38	Hapmap46932-BTA-111719	6	84819700	2.39E-07	6.62	-2.06	3%	20%	0.21	CF_P
39	ARS-BFGL-NGS-42175	6	79544981	4.55E-08	7.34	-2.15	3%	21%	0.20	CF_P
40	Hapmap60030-rs29013992	6	77585276	3.78E-08	7.42	-2.35	3%	21%	0.16	CF_P
41	BTA-114800-no-rs	6	77520815	2.52E-08	7.60	-2.37	4%	22%	0.16	CF_P
42	Hapmap23387-BTC-072905	6	82078166	3.80E-08	7.42	-2.49	4%	22%	0.14	CF_P
43	Hapmap32475-BTC-050530	6	82047313	3.80E-08	7.42	-2.49	4%	22%	0.14	CF_P
44	BTB-00264815	6	81019581	2.16E-08	7.67	-2.27	4%	22%	0.18	CF_P
45	ARS-BFGL-NGS-80068	6	77650126	1.88E-08	7.73	-2.39	4%	22%	0.16	CF_P
46	Hapmap52479-rs29018853	6	79203343	1.34E-08	7.87	-2.10	4%	23%	0.24	CF_P
47	ARS-BFGL-NGS-27643	6	78786848	4.46E-09	8.35	-2.45	4%	25%	0.17	CF_P
48	ARS-BFGL-NGS-114609	6	84713584	2.91E-10	9.54	-2.40	5%	30%	0.23	CF_P
49	Hapmap24184-BTC-070077	6	87245049	3.65E-11	10.44	-2.78	5%	31%	0.17	CF_P
50	Hapmap28023-BTC-060518	6	87201599	1.41E-11	10.85	-2.79	5%	32%	0.18	CF_P
51	BTA-111108-no-rs	6	85424500	6.62E-13	12.18	-3.35	6%	37%	0.13	CF_P
52	ARS-BFGL-NGS-36707	6	86354888	9.44E-13	12.02	-3.39	6%	38%	0.13	CF_P
53	Hapmap41098-BTA-86027	6	84889974	1.41E-13	12.85	-3.47	6%	40%	0.13	CF_P
54	Hapmap52348-rs29024684	6	87396306	1.62E-17	16.79	-3.19	9%	53%	0.24	CF_P
55	Hapmap49297-BTA-76961	6	83147633	3.16E-05	4.50	-2.96	19%	35%	0.25	CF_{max}
56	Hapmap27109-BTC-060711	6	87152621	2.66E-05	4.58	-2.77	20%	36%	0.32	CF_{max}
57	Hapmap33631-BTC-043555	6	87327708	2.23E-05	4.65	-2.84	21%	39%	0.35	CF_{max}
58	ARS-BFGL-NGS-42175	6	79544981	1.15E-05	4.94	-3.41	22%	39%	0.20	CF_{max}

59	BTB-01393607	6	80062406	9.52E-06	5.02	-3.49	22%	40%	0.19	CF_{max}
60	ARS-BFGL-NGS-70112	6	84448550	6.28E-06	5.20	-3.43	23%	42%	0.22	CF_{max}
61	Hapmap43767-BTA-113302	6	85646902	8.94E-06	5.05	-2.91	23%	42%	0.38	CF_{max}
62	ARS-BFGL-NGS-27643	6	78786848	4.58E-06	5.34	-3.76	24%	43%	0.17	CF_{max}
63	BTB-00264815	6	81019581	1.77E-06	5.75	-3.80	25%	46%	0.18	CF_{max}
64	Hapmap41098-BTA-86027	6	84889974	2.66E-06	5.58	-4.34	25%	46%	0.13	CF_{max}
65	Hapmap60030-rs29013992	6	77585276	1.34E-06	5.87	-4.06	26%	47%	0.16	CF_{max}
66	BTA-114800-no-rs	6	77520815	1.23E-06	5.91	-4.04	26%	47%	0.16	CF_{max}
67	ARS-BFGL-NGS-36707	6	86354888	3.03E-06	5.52	-4.37	26%	47%	0.13	CF_{max}
68	ARS-BFGL-NGS-80068	6	77650126	1.05E-06	5.98	-4.07	26%	47%	0.16	CF_{max}
69	BTA-111108-no-rs	6	85424500	1.03E-06	5.99	-4.55	28%	51%	0.13	CF_{max}
70	Hapmap24184-BTC-070077	6	87245049	4.70E-08	7.33	-4.54	34%	62%	0.17	CF_{max}
71	Hapmap28023-BTC-060518	6	87201599	1.09E-08	7.96	-4.67	37%	67%	0.18	CF_{max}
72	ARS-BFGL-NGS-112872	6	88069548	9.97E-06	5.00	1.62	2%	11%	0.14	k_{CF}
73	Hapmap52348-rs29024684	6	87396306	7.46E-07	6.13	-1.51	3%	14%	0.24	k_{CF}
74	BTA-122637-no-rs	6	88442145	4.01E-07	6.40	-2.70	3%	15%	0.07	k_{CF}
75	BTA-110240-no-rs	6	81652194	3.21E-05	4.49	0.07	0%	0%	0.20	Prot
76	Hapmap56688-rs29025335	6	81767374	3.44E-05	4.46	0.07	0%	0%	0.19	Prot
77	Hapmap55384-rs29026113	6	79525009	3.04E-05	4.52	0.07	0%	0%	0.23	Prot
78	Hapmap47844-BTA-115673	6	1.14E+08	2.20E-05	4.66	-0.07	0%	0%	0.25	Prot
79	Hapmap26275-BTC-043486	6	82409949	1.60E-05	4.80	0.09	0%	0%	0.11	Prot
80	ARS-BFGL-NGS-27958	6	84689991	1.64E-06	5.79	0.09	0%	0%	0.16	Prot
81	BTA-122637-no-rs	6	88442145	2.39E-06	5.62	2.32	4%	17%	0.07	RCT_{eq}
82	Hapmap43353-BTA-76584	6	64179687	4.65E-05	4.33	4.32	2%	12%	0.05	t_{max}
83	BTA-122637-no-rs	6	88442145	3.63E-07	6.44	4.75	4%	19%	0.07	t_{max}
84	ARS-BFGL-NGS-88859	9	83575446	1.06E-05	4.97	-0.06	0%	0%	0.38	Prot
85	Hapmap53034-rs29011422	9	69696334	9.04E-06	5.04	-0.07	0%	0%	0.34	Prot
86	ARS-BFGL-NGS-116951	11	87793629	3.47E-05	4.46	-1.00	3%	12%	0.39	RCT_{eq}
87	ARS-BFGL-NGS-37074	11	88028793	1.96E-05	4.71	-1.02	3%	13%	0.42	RCT_{eq}
88	BTA-110431-no-rs	11	87692024	6.47E-06	5.19	-1.09	3%	15%	0.40	RCT_{eq}
89	BTA-110429-no-rs	11	87670344	5.72E-06	5.24	-1.09	3%	15%	0.42	RCT_{eq}

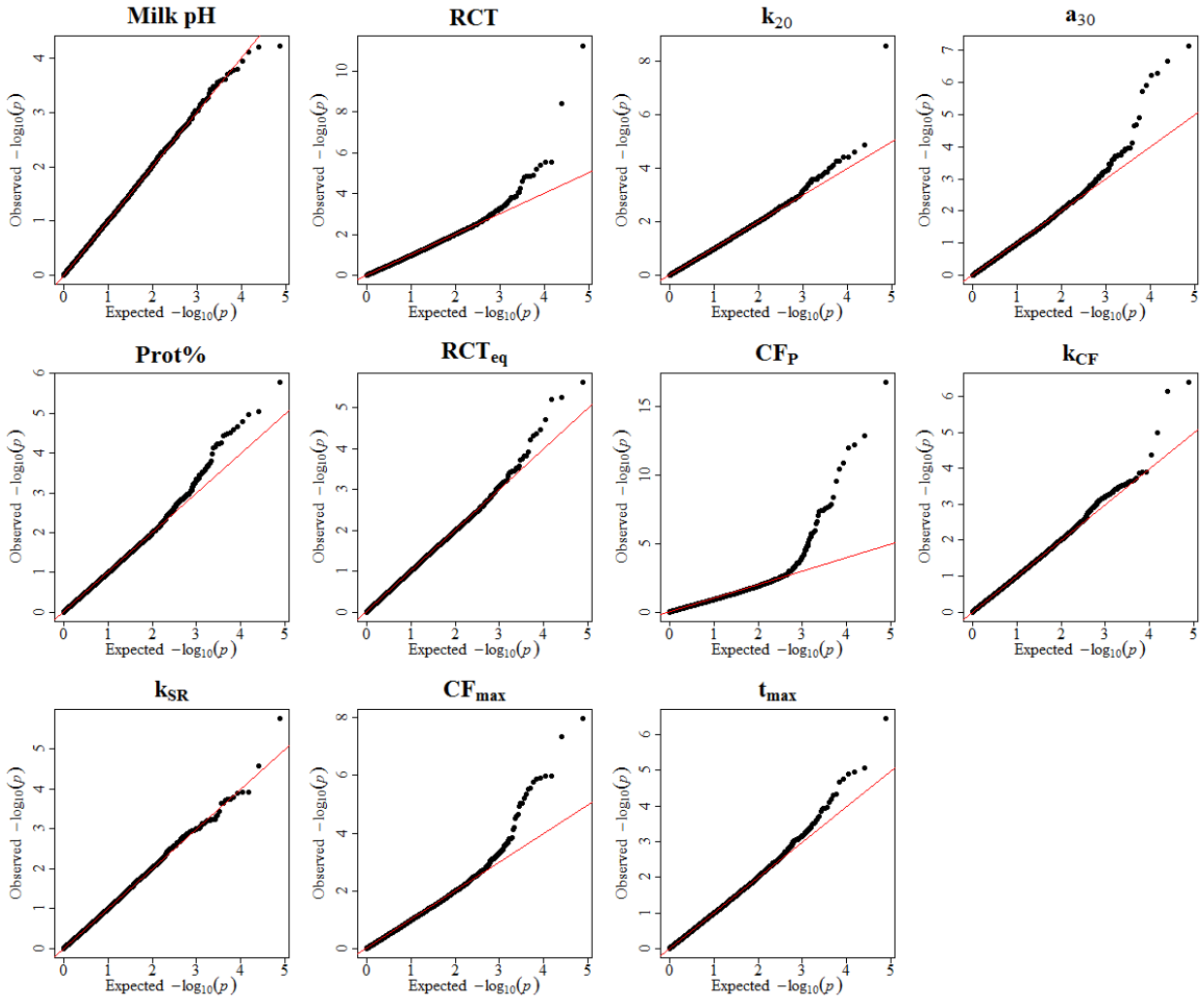
90	BTA-110431-no-rs	11	87692024	1.79E-05	4.75	-1.95	2%	12%	0.40	t_{max}
91	BTA-120876-no-rs	11	85935590	1.31E-05	4.88	-1.98	2%	13%	0.42	t_{max}
92	BTA-110429-no-rs	11	87670344	1.11E-05	4.95	-2.00	3%	13%	0.42	t_{max}
93	ARS-BFGL-NGS-119913	11	86779385	8.75E-06	5.06	1.98	3%	13%	0.50	t_{max}
94	Hapmap31215-BTA-32775	13	47879982	1.45E-05	4.84	5.82	1%	4%	0.01	RCT
95	ARS-BFGL-NGS-68607	15	55488319	6.76E-06	5.17	6.31	1%	4%	0.01	RCT
96	ARS-BFGL-NGS-114291	15	14242668	2.77E-06	5.56	5.64	1%	5%	0.01	RCT
97	ARS-BFGL-NGS-117603	15	14272339	2.81E-06	5.55	5.64	1%	5%	0.01	RCT
98	ARS-BFGL-NGS-17574	16	76311292	2.74E-05	4.56	-1.43	2%	12%	0.29	CF_P
99	Hapmap27040-BTA-25119	19	2270929	2.30E-05	4.64	2.07	1%	5%	0.32	a_{30}
100	Hapmap39832-BTA-46468	19	2093500	2.08E-05	4.68	2.09	1%	5%	0.32	a_{30}
101	ARS-BFGL-NGS-1751	20	17412441	3.57E-05	4.45	0.08	0%	0%	0.17	Prot
102	Hapmap38418-BTA-57213	23	8819178	2.69E-05	4.57	0.16	0%	0%	0.11	k_{SR}
103	ARS-BFGL-NGS-99929	23	10631079	1.75E-06	5.76	0.22	4%	64%	0.07	k_{SR}
104	ARS-BFGL-NGS-23064	26	20365711	4.30E-05	4.37	1.03	2%	9%	0.47	k_{CF}
105	Hapmap48306-BTA-36540	28	38449335	2.64E-05	4.58	-0.08	0%	0%	0.13	Prot
106	ARS-BFGL-NGS-115508	28	33729338	2.10E-05	4.68	-3.14	2%	13%	0.11	t_{max}

SNP= the name of the single nucleotide polymorphism; BTA= *Bos taurus* autosome chromosome; Location= position of the SNP on the chromosome in base pairs on UMD3.1; Pc1df= p-values adjusted for genomic control; LOG= the $-\log_{10}$ of Pc1df; effB= effect of the minor allele (B allele); snp.V_P (%)= percentage of phenotypic variance explained by each SNP; snp.V_a (%)= percentage of additive genetic variance explained by each SNP; MAF= minor allele frequency;

[1]RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Prot % = protein percent; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

*Undefined chromosome and position on the genome

Figure S1 Quantile-quantile (Q-Q) plots of the observed test statistics of the genome wide association studies (GWAS)



RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; Prot % = protein percent; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant (% x min^{-1}); k_{SR} = syneresis rate constant (% x min^{-1}); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min).

CHAPTER 2

Table S1. List of all significant SNP identified in the GWAS analyses sorted by chromosome.

N	SNP	BTA	Location (bp)	effB	Pc1df	LOG	V_{SNP}	V_P	V_G	V_{Psnp} (%)	V_{Gsnp} (%)	Trait
1	Hapmap60596-rs29017365	2	128812635	-0.16	3.42E-05	4.47	0.01	0.52	0.15	2.19	7.55	$\%CY_{SOLIDS}$
2	Hapmap53172-rs29012675	6	82706745	0.85	8.57E-08	7.07	0.06	1.81	0.66	3.57	9.76	$\%CY_{SOLIDS}$
3	Hapmap26275-BTC-043486	6	82409949	0.49	4.99E-06	5.30	0.05	1.81	0.66	2.64	7.22	$\%CY_{SOLIDS}$
4	Hapmap23975-BTC-043815	6	102937469	0.38	9.59E-06	5.02	0.05	1.81	0.66	2.52	6.90	$\%CY_{SOLIDS}$
5	Hapmap50464-BTA-77021	6	82560990	0.44	1.12E-05	4.95	0.04	1.81	0.66	2.33	6.37	$\%CY_{SOLIDS}$
6	Hapmap38371-BTA-105598	6	87715723	0.35	2.56E-05	4.59	0.04	1.81	0.66	2.22	6.07	$\%CY_{SOLIDS}$
7	BTA-110240-no-rs	6	81652194	0.34	4.85E-05	4.31	0.04	1.81	0.66	2.07	5.67	$\%CY_{SOLIDS}$
8	Hapmap53172-rs29012675	6	82706745	0.42	8.03E-07	6.10	0.02	0.52	0.15	3.00	10.36	$\%CY_{SOLIDS}$
9	Hapmap50464-BTA-77021	6	82560990	0.26	1.05E-06	5.98	0.01	0.52	0.15	2.85	9.83	$\%CY_{SOLIDS}$
10	Hapmap26275-BTC-043486	6	82409949	0.26	7.09E-06	5.15	0.01	0.52	0.15	2.53	8.73	$\%CY_{SOLIDS}$
11	Hapmap52348-rs29024684	6	87396306	-0.84	3.93E-07	6.41	0.26	7.75	2.22	3.31	11.54	REC_{SOLIDS}
12	Hapmap50464-BTA-77021	6	82560990	0.98	1.46E-06	5.84	0.21	7.75	2.22	2.77	9.67	REC_{SOLIDS}
13	Hapmap26275-BTC-043486	6	82409949	1.02	3.64E-06	5.44	0.21	7.75	2.22	2.69	9.37	REC_{SOLIDS}
14	Hapmap53172-rs29012675	6	82706745	1.48	6.29E-06	5.20	0.19	7.75	2.22	2.52	8.77	REC_{SOLIDS}
15	ARS-BFGL-NGS-82008	6	87600892	0.87	2.27E-05	4.64	0.17	7.75	2.22	2.18	7.59	REC_{SOLIDS}
16	ARS-BFGL-NGS-109039	6	85142067	0.88	4.38E-05	4.36	0.16	7.75	2.22	2.09	7.29	REC_{SOLIDS}
17	Hapmap52348-rs29024684	6	87396306	-1.34	1.91E-15	14.72	0.65	7.71	1.11	8.45	58.77	REC_{FAT}
18	BTA-111108-no-rs	6	85424500	-1.44	9.72E-12	11.01	0.48	7.71	1.11	6.22	43.26	REC_{FAT}
19	Hapmap41098-BTA-86027	6	84889974	-1.43	1.16E-11	10.93	0.47	7.71	1.11	6.14	42.67	REC_{FAT}
20	ARS-BFGL-NGS-36707	6	86354888	-1.38	1.33E-10	9.88	0.44	7.71	1.11	5.72	39.74	REC_{FAT}
21	Hapmap23387-BTC-072905	6	82078166	-1.27	5.34E-10	9.27	0.40	7.71	1.11	5.17	35.94	REC_{FAT}
22	Hapmap32475-BTC-050530	6	82047313	-1.27	5.34E-10	9.27	0.40	7.71	1.11	5.17	35.94	REC_{FAT}
23	Hapmap46932-BTA-111719	6	84819700	-1.02	1.28E-08	7.89	0.35	7.71	1.11	4.52	31.42	REC_{FAT}

24	ARS-BFGL-NGS-63312	6	82965163	-1.02	1.96E-08	7.71	0.32	7.71	1.11	4.08	28.40	REC_{FAT}
25	ARS-BFGL-NGS-27643	6	78786848	-1.04	3.32E-08	7.48	0.31	7.71	1.11	4.06	28.24	REC_{FAT}
26	ARS-BFGL-NGS-114609	6	84713584	-0.93	5.26E-08	7.28	0.31	7.71	1.11	3.98	27.70	REC_{FAT}
27	ARS-BFGL-NGS-70112	6	84448550	-0.92	1.13E-07	6.95	0.28	7.71	1.11	3.68	25.60	REC_{FAT}
28	BTA-114800-no-rs	6	77520815	-1.01	1.51E-07	6.82	0.27	7.71	1.11	3.56	24.76	REC_{FAT}
29	ARS-BFGL-NGS-80068	6	77650126	-1.00	1.77E-07	6.75	0.27	7.71	1.11	3.51	24.42	REC_{FAT}
30	Hapmap52479-rs29018853	6	79203343	-0.87	1.84E-07	6.74	0.27	7.71	1.11	3.52	24.51	REC_{FAT}
31	Hapmap60030-rs29013992	6	77585276	-1.00	2.14E-07	6.67	0.27	7.71	1.11	3.44	23.92	REC_{FAT}
32	ARS-BFGL-NGS-42175	6	79544981	-0.92	2.18E-07	6.66	0.27	7.71	1.11	3.48	24.21	REC_{FAT}
33	Hapmap28023-BTC-060518	6	87201599	-0.95	3.36E-07	6.47	0.26	7.71	1.11	3.39	23.59	REC_{FAT}
34	BTB-00264815	6	81019581	-0.92	4.20E-07	6.38	0.25	7.71	1.11	3.30	22.97	REC_{FAT}
35	Hapmap31932-BTC-042947	6	82350917	-0.79	4.61E-07	6.34	0.26	7.71	1.11	3.37	23.42	REC_{FAT}
36	Hapmap24184-BTC-070077	6	87245049	-0.95	4.78E-07	6.32	0.26	7.71	1.11	3.31	23.04	REC_{FAT}
37	Hapmap23419-BTC-059652	6	82314634	-0.78	6.68E-07	6.17	0.25	7.71	1.11	3.28	22.84	REC_{FAT}
38	Hapmap51938-BTA-21491	6	81057816	-0.82	1.10E-06	5.96	0.24	7.71	1.11	3.10	21.58	REC_{FAT}
39	Hapmap26317-BTC-059618	6	82253271	-0.76	1.26E-06	5.90	0.24	7.71	1.11	3.10	21.59	REC_{FAT}
40	BTA-122637-no-rs	6	88442145	-1.38	1.42E-06	5.85	0.24	7.71	1.11	3.05	21.23	REC_{FAT}
41	Hapmap49297-BTA-76961	6	83147633	-0.78	1.46E-06	5.84	0.23	7.71	1.11	3.00	20.88	REC_{FAT}
42	BTB-01393607	6	80062406	-0.80	7.75E-06	5.11	0.20	7.71	1.11	2.58	17.91	REC_{FAT}
43	BTB-00264414	6	79843283	-0.66	9.28E-06	5.03	0.19	7.71	1.11	2.46	17.09	REC_{FAT}
44	Hapmap60182-rs29025531	6	74606760	-0.77	2.09E-05	4.68	0.19	7.71	1.11	2.41	16.73	REC_{FAT}
45	Hapmap32099-BTA-151095	6	83345994	-0.65	2.47E-05	4.61	0.18	7.71	1.11	2.28	15.87	REC_{FAT}
46	BTA-76959-no-rs	6	83290843	-0.65	2.47E-05	4.61	0.18	7.71	1.11	2.29	15.89	REC_{FAT}
47	ARS-BFGL-NGS-17026	6	73734080	-0.74	3.88E-05	4.41	0.17	7.71	1.11	2.15	14.96	REC_{FAT}
48	Hapmap52348-rs29024684	6	87396306	-0.90	3.62E-08	7.44	0.29	7.52	1.81	3.92	16.27	REC_{ENERGY}
49	Hapmap50464-BTA-77021	6	82560990	0.98	1.06E-06	5.98	0.21	7.52	1.81	2.84	11.79	REC_{ENERGY}
50	Hapmap26275-BTC-043486	6	82409949	0.97	8.01E-06	5.10	0.19	7.52	1.81	2.49	10.36	REC_{ENERGY}
51	Hapmap41098-BTA-86027	6	84889974	-0.90	1.02E-05	4.99	0.19	7.52	1.81	2.51	10.44	REC_{ENERGY}
52	ARS-BFGL-NGS-109039	6	85142067	0.91	1.77E-05	4.75	0.17	7.52	1.81	2.30	9.56	REC_{ENERGY}

53	BTA-111108-no-rs	6	85424500	-0.86	2.81E-05	4.55	0.17	7.52	1.81	2.27	9.43	REC_{ENERGY}
54	ARS-BFGL-NGS-42175	6	79544981	-0.72	2.84E-05	4.55	0.16	7.52	1.81	2.18	9.06	REC_{ENERGY}
55	ARS-BFGL-NGS-82008	6	87600892	0.84	3.12E-05	4.51	0.16	7.52	1.81	2.10	8.72	REC_{ENERGY}
56	Hapmap52479-rs29018853	6	79203343	-0.67	3.13E-05	4.50	0.16	7.52	1.81	2.17	9.01	REC_{ENERGY}
57	BTA-113303-no-rs	6	85954909	0.98	3.20E-05	4.50	0.16	7.52	1.81	2.12	8.81	REC_{ENERGY}
58	Hapmap53172-rs29012675	6	82706745	1.31	4.48E-05	4.35	0.15	7.52	1.81	2.04	8.46	REC_{ENERGY}
59	BTB-00403185	9	91058994	0.60	4.74E-05	4.32	0.16	7.52	1.81	2.11	8.78	REC_{ENERGY}
60	ARS-BFGL-NGS-104610	11	104293559	-1.30	6.07E-36	35.22	0.83	3.64	1.67	22.87	49.96	$REC_{PROTEIN}$
61	ARS-BFGL-NGS-77843	11	103856100	-1.04	7.71E-17	16.11	0.35	3.64	1.67	9.50	20.76	$REC_{PROTEIN}$
62	ARS-BFGL-NGS-115328	11	103110855	-0.87	1.62E-15	14.79	0.33	3.64	1.67	9.17	20.04	$REC_{PROTEIN}$
63	ARS-BFGL-NGS-111682	11	104633267	-0.90	1.93E-12	11.71	0.25	3.64	1.67	6.96	15.20	$REC_{PROTEIN}$
64	ARS-BFGL-NGS-32953	11	101301047	-0.73	6.39E-12	11.19	0.24	3.64	1.67	6.51	14.22	$REC_{PROTEIN}$
65	ARS-BFGL-NGS-83830	11	102790757	-0.90	2.09E-11	10.68	0.22	3.64	1.67	6.14	13.42	$REC_{PROTEIN}$
66	ARS-BFGL-NGS-14392	11	102237822	-0.90	2.48E-10	9.61	0.20	3.64	1.67	5.53	12.09	$REC_{PROTEIN}$
67	ARS-BFGL-NGS-118288	11	106613461	-0.79	2.57E-10	9.59	0.20	3.64	1.67	5.52	12.05	$REC_{PROTEIN}$
68	ARS-BFGL-NGS-7395	11	102178546	-0.98	3.71E-10	9.43	0.20	3.64	1.67	5.50	12.02	$REC_{PROTEIN}$
69	ARS-BFGL-NGS-82968	11	105182939	-0.68	1.19E-09	8.92	0.18	3.64	1.67	4.98	10.87	$REC_{PROTEIN}$
70	ARS-BFGL-NGS-115623	11	102153173	-0.93	1.82E-09	8.74	0.18	3.64	1.67	5.06	11.06	$REC_{PROTEIN}$
71	ARS-BFGL-NGS-119318	11	102974570	-0.66	2.37E-09	8.63	0.18	3.64	1.67	5.00	10.93	$REC_{PROTEIN}$
72	Hapmap51810-BTA-119667	11	102861577	0.60	2.67E-09	8.57	0.18	3.64	1.67	4.91	10.72	$REC_{PROTEIN}$
73	ARS-BFGL-NGS-6104	11	104456040	-1.05	5.76E-09	8.24	0.18	3.64	1.67	4.85	10.59	$REC_{PROTEIN}$
74	ARS-BFGL-NGS-4101	11	104325453	-1.04	8.78E-09	8.06	0.17	3.64	1.67	4.66	10.17	$REC_{PROTEIN}$
75	ARS-BFGL-NGS-42578	11	93241685	-0.67	1.26E-08	7.90	0.17	3.64	1.67	4.58	10.00	$REC_{PROTEIN}$
76	ARS-BFGL-NGS-25612	11	93193697	-0.63	6.37E-08	7.20	0.15	3.64	1.67	4.17	9.10	$REC_{PROTEIN}$
77	ARS-BFGL-NGS-20482	11	92990077	-0.75	8.33E-08	7.08	0.16	3.64	1.67	4.2	9.35	$REC_{PROTEIN}$
78	ARS-BFGL-NGS-21607	11	97059246	-1.02	1.33E-07	6.88	0.15	3.64	1.67	4.09	8.94	$REC_{PROTEIN}$
79	ARS-BFGL-NGS-11064	11	106510932	-0.54	3.25E-07	6.49	0.13	3.64	1.67	3.58	7.82	$REC_{PROTEIN}$
80	ARS-BFGL-NGS-98548	11	103458444	0.67	1.37E-06	5.86	0.12	3.64	1.67	3.24	7.08	$REC_{PROTEIN}$
81	Hapmap32029-BTA-127208	11	105532933	-0.47	2.18E-06	5.66	0.11	3.64	1.67	3.06	6.68	$REC_{PROTEIN}$

82	ARS-BFGL-NGS-116123	11	94686959	-0.54	3.73E-06	5.43	0.11	3.64	1.67	3.13	6.83	$REC_{PROTEIN}$
83	Hapmap43475-BTA-115549	11	94772650	-0.55	4.17E-06	5.38	0.11	3.64	1.67	3.12	6.82	$REC_{PROTEIN}$
84	ARS-BFGL-NGS-22584	11	105845271	-0.47	4.34E-06	5.36	0.11	3.64	1.67	3.04	6.63	$REC_{PROTEIN}$
85	ARS-BFGL-NGS-11629	11	92963716	-0.69	4.80E-06	5.32	0.11	3.64	1.67	3.02	6.60	$REC_{PROTEIN}$
86	ARS-BFGL-NGS-70509	11	98510712	0.46	5.91E-06	5.23	0.10	3.64	1.67	2.85	6.23	$REC_{PROTEIN}$
87	BTA-118663-no-rs	11	101384765	0.46	6.13E-06	5.21	0.10	3.64	1.67	2.83	6.19	$REC_{PROTEIN}$
88	ARS-BFGL-NGS-25833	11	106543262	-0.49	7.15E-06	5.15	0.10	3.64	1.67	2.82	6.15	$REC_{PROTEIN}$
89	BTA-118661-no-rs	11	101357473	0.46	8.21E-06	5.09	0.10	3.64	1.67	2.75	6.01	$REC_{PROTEIN}$
90	ARS-BFGL-NGS-13452	11	102673893	-0.45	8.69E-06	5.06	0.10	3.64	1.67	2.70	5.90	$REC_{PROTEIN}$
91	BTA-93319-no-rs	11	91639283	-0.55	8.80E-06	5.06	0.10	3.64	1.67	2.84	6.21	$REC_{PROTEIN}$
92	ARS-BFGL-NGS-35656	11	97289960	0.57	1.57E-05	4.80	0.09	3.64	1.67	2.60	5.69	$REC_{PROTEIN}$
93	ARS-BFGL-NGS-101698	11	96230130	-0.52	1.69E-05	4.77	0.10	3.64	1.67	2.62	5.72	$REC_{PROTEIN}$
94	ARS-BFGL-NGS-11867	11	106741315	-0.62	2.17E-05	4.66	0.09	3.64	1.67	2.39	5.23	$REC_{PROTEIN}$
95	BTA-04956-no-rs	11	94713144	-0.56	2.24E-05	4.65	0.09	3.64	1.67	2.59	5.67	$REC_{PROTEIN}$
96	Hapmap56906-rs29014970	11	97844929	-0.47	3.31E-05	4.48	0.09	3.64	1.67	2.54	5.54	$REC_{PROTEIN}$
97	ARS-BFGL-NGS-59502	11	100519715	-0.55	3.74E-05	4.43	0.09	3.64	1.67	2.43	5.31	$REC_{PROTEIN}$
98	BTB-00507211	12	85272488	-0.29	2.42E-05	4.62	0.02	0.78	0.20	2.34	8.92	$\%CY_{WATER}$
99	Hapmap25446-BTC-054694	14	26003598	0.15	4.58E-05	4.34	0.01	0.52	0.15	2.08	7.19	$\%CY_{SOLIDS}$
100	Hapmap51570-BTA-17962	18	65269893	0.21	4.32E-05	4.36	0.02	0.78	0.20	2.05	7.83	$\%CY_{WATER}$
101	ARS-BFGL-NGS-102974	19	1822133	-0.15	3.05E-05	4.52	0.01	0.52	0.15	2.05	7.09	$\%CY_{SOLIDS}$
102	ARS-BFGL-NGS-87845	27	42118037	0.50	1.53E-05	4.82	0.01	0.52	0.15	2.47	8.51	$\%CY_{SOLIDS}$
103	Hapmap48306-BTA-36540	28	38449335	-0.27	1.66E-05	4.78	0.02	0.78	0.20	2.22	8.48	$\%CY_{WATER}$

SNP= the name of the single nucleotide polymorphism; BTA= *Bos taurus* autosome chromosome; Location= position of the SNP on the chromosome in base pairs on UMD3.1; Pc1df= p-values adjusted for genomic control; LOG= the $-\log_{10}$ of Pc1df; effB= effect of the minor allele (B allele); V_{SNP} = variance explained by the SNP (calculated as $2pqa^2$, where p is the frequency of one allele, q=1-p is the frequency of the second allele and a is the additive genetic effect); V_P= phenotypic variance; V_G= additive genetic variance; V_{Psnp} (%)= percentage of phenotypic variance explained by each SNP; V_{Gsnp} (%)= percentage of additive genetic variance explained by each SNP; Trait= name of the trait

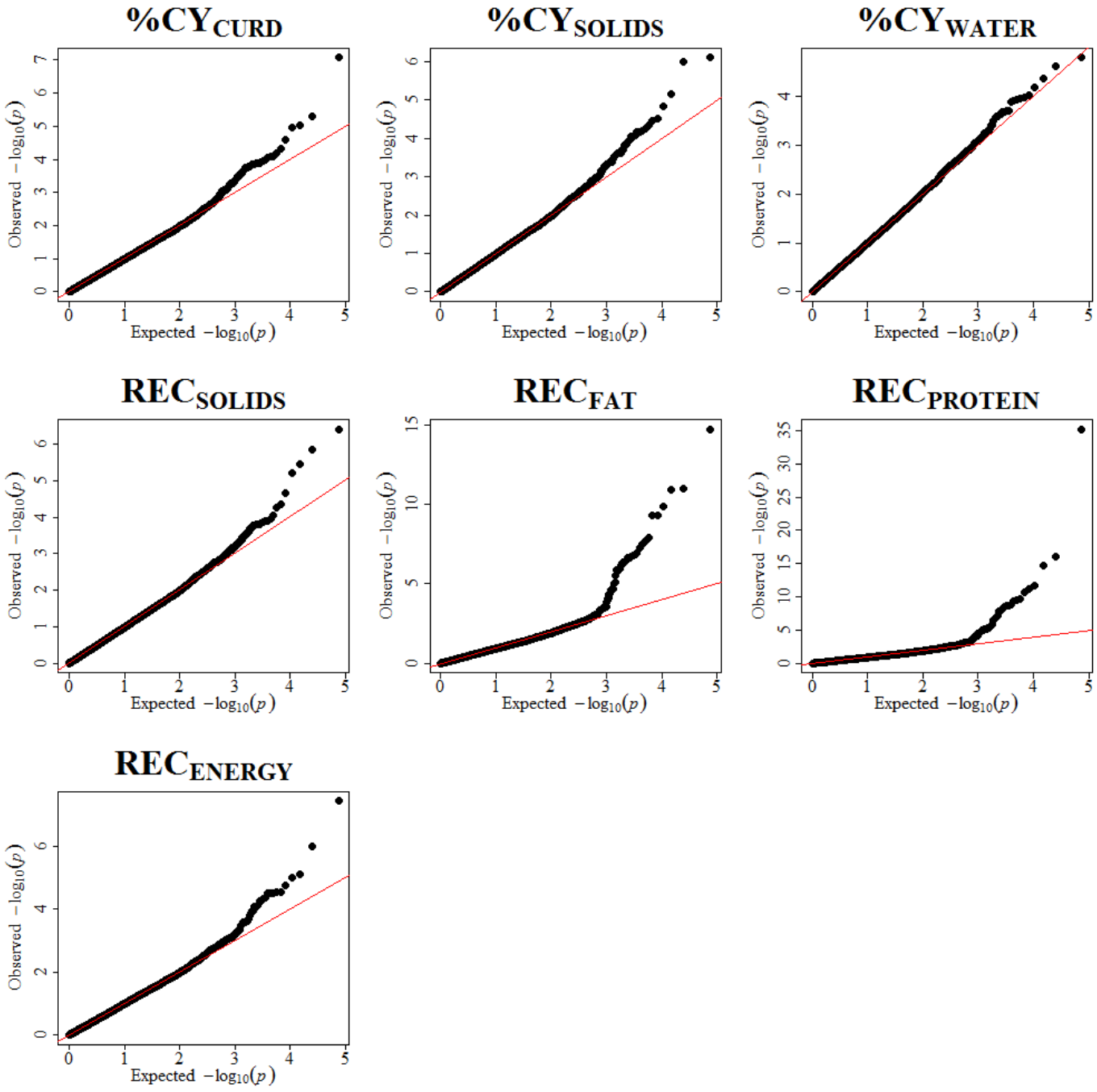
Table S2. List of all significant SNP identified in the GWAS analyses sorted by chromosome after fixing the marker Hapmap52348-rs29024684 at 87,396,306bp on BTA6 and the marker ARS-BFGL-NGS-104610 at 104,293,559 on BTA11.

N	SNP	BTA	Location (bp)	effB	Pc1df	LOG	V_{SNP}	V_P	V_G	V_{Psnp} (%)	V_{Gsnp} (%)	Trait
1	Hapmap26275-BTC-043486	6	82409949	0.45	2.22E-05	4.65	0.04	1.77	0.63	2.28	6.36	$\%CY_{CURD}$
2	Hapmap50464-BTA-77021	6	82560990	0.41	3.63E-05	4.44	0.04	1.77	0.63	2.07	5.76	$\%CY_{CURD}$
3	Hapmap27307-BTC-043200	6	82605943	0.35	2.19E-05	4.66	0.04	1.77	0.63	2.22	6.18	$\%CY_{CURD}$
4	Hapmap53172-rs29012675	6	82706745	0.88	2.30E-08	7.64	0.07	1.77	0.63	3.92	10.92	$\%CY_{CURD}$
5	Hapmap38371-BTA-105598	6	87715723	0.33	4.04E-05	4.39	0.04	1.77	0.63	2.11	5.89	$\%CY_{CURD}$
6	Hapmap32578-BTA-144239	6	88111170	0.34	1.05E-05	4.98	0.04	1.77	0.63	2.54	7.08	$\%CY_{CURD}$
7	Hapmap23975-BTC-043815	6	102937469	0.36	1.33E-05	4.87	0.04	1.77	0.63	2.44	6.81	$\%CY_{CURD}$
8	Hapmap26275-BTC-043486	6	82409949	0.23	4.28E-05	4.37	0.01	0.51	0.15	2.10	7.19	$\%CY_{SOLIDS}$
9	Hapmap50464-BTA-77021	6	82560990	0.24	4.85E-06	5.31	0.01	0.51	0.15	2.50	8.57	$\%CY_{SOLIDS}$
10	Hapmap53172-rs29012675	6	82706745	0.43	2.87E-07	6.54	0.02	0.51	0.15	3.27	11.22	$\%CY_{SOLIDS}$
11	Hapmap32578-BTA-144239	6	88111170	0.18	1.09E-05	4.96	0.01	0.51	0.15	2.51	8.59	$\%CY_{SOLIDS}$
12	BTB-02092741	6	114223059	0.62	4.29E-05	4.37	0.01	0.51	0.15	2.08	7.12	$\%CY_{SOLIDS}$
13	BTA-77034-no-rs	6	82540280	0.72	4.15E-05	4.38	0.15	7.10	1.45	2.09	10.20	REC_{ENERGY}
14	Hapmap50464-BTA-77021	6	82560990	0.87	6.34E-06	5.20	0.17	7.10	1.45	2.40	11.71	REC_{ENERGY}
15	Hapmap27307-BTC-043200	6	82605943	0.77	2.92E-06	5.53	0.19	7.10	1.45	2.62	12.79	REC_{ENERGY}
16	Hapmap53172-rs29012675	6	82706745	1.39	7.48E-06	5.13	0.17	7.10	1.45	2.43	11.90	REC_{ENERGY}
17	Hapmap60224-rs29001782	6	85178107	0.75	4.10E-05	4.39	0.15	7.10	1.45	2.14	10.46	REC_{ENERGY}
18	Hapmap32578-BTA-144239	6	88111170	0.65	1.81E-05	4.74	0.17	7.10	1.45	2.34	11.43	REC_{ENERGY}
19	BTA-77034-no-rs	6	82540280	0.76	9.09E-06	5.04	0.16	7.06	0.64	2.34	25.73	REC_{FAT}
20	Hapmap27307-BTC-043200	6	82605943	0.77	1.33E-06	5.88	0.19	7.06	0.64	2.66	29.32	REC_{FAT}
21	Hapmap23522-BTC-047379	6	87222751	1.17	2.32E-05	4.63	0.15	7.06	0.64	2.11	23.23	REC_{FAT}
22	Hapmap30192-BTC-072881	6	87281196	1.16	2.43E-05	4.61	0.15	7.06	0.64	2.10	23.18	REC_{FAT}
23	Hapmap54015-rs29022799	6	88421804	1.04	1.63E-05	4.79	0.15	7.06	0.64	2.12	23.36	REC_{FAT}
24	ARS-BFGL-NGS-118182	6	88592295	0.62	3.65E-05	4.44	0.14	7.06	0.64	1.97	21.72	REC_{FAT}
25	BTA-113303-no-rs	6	85954909	0.58	4.77E-05	4.32	0.06	2.77	0.67	1.99	8.20	$REC_{PROTEIN}$
26	ARS-BFGL-NGS-112872	6	88069548	0.54	5.21E-06	5.28	0.07	2.77	0.67	2.61	10.78	$REC_{PROTEIN}$

27	Hapmap50464-BTA-77021	6	82560990	0.88	1.20E-05	4.92	0.17	7.10	1.45	2.40	11.72	REC_{SOLIDS}
28	Hapmap27307-BTC-043200	6	82605943	0.74	1.33E-05	4.88	0.17	7.10	1.45	2.42	11.82	REC_{SOLIDS}
29	Hapmap53172-rs29012675	6	82706745	1.59	7.79E-07	6.11	0.23	7.10	1.45	3.17	15.51	REC_{SOLIDS}
30	Hapmap32578-BTA-144239	6	88111170	0.69	8.90E-06	5.05	0.19	7.10	1.45	2.67	13.04	REC_{SOLIDS}
31	BTB-01249963	9	37442535	-0.18	4.64E-05	4.33	0.02	0.77	0.20	1.96	7.70	$\%CY_{WATER}$
32	ARS-BFGL-NGS-26919	11	103352220	1.48	7.92E-12	11.10	0.16	2.77	0.67	5.73	23.61	$REC_{PROTEIN}$
33	ARS-BFGL-NGS-40995	11	103742782	0.58	1.26E-05	4.90	0.07	2.77	0.67	2.45	10.12	$REC_{PROTEIN}$
34	BTB-00507211	12	85272488	-0.29	2.51E-05	4.60	0.02	0.77	0.20	2.33	9.13	$\%CY_{WATER}$
35	ARS-BFGL-NGS-102974	19	1822133	-0.28	3.04E-05	4.52	0.04	1.77	0.63	2.05	5.72	$\%CY_{CURD}$
36	ARS-BFGL-NGS-24753	19	3024589	-0.29	3.50E-05	4.46	0.04	1.77	0.63	2.16	6.01	$\%CY_{CURD}$
37	ARS-BFGL-NGS-102974	19	1822133	-0.16	7.54E-06	5.12	0.01	0.51	0.15	2.37	8.12	$\%CY_{SOLIDS}$
38	ARS-BFGL-NGS-24753	19	3024589	-0.16	2.91E-05	4.54	0.01	0.51	0.15	2.20	7.55	$\%CY_{SOLIDS}$
39	ARS-BFGL-NGS-43028	19	1706305	-0.15	3.59E-05	4.45	0.01	0.51	0.15	2.09	7.15	$\%CY_{SOLIDS}$
40	ARS-BFGL-NGS-102974	19	1822133	-0.58	3.03E-05	4.52	0.15	7.10	1.45	2.14	10.47	REC_{SOLIDS}
41	ARS-BFGL-NGS-4481	22	38782041	2.05	1.92E-05	4.72	0.06	2.77	0.67	2.23	9.18	$REC_{PROTEIN}$
42	ARS-BFGL-NGS-87845	27	42118037	0.48	2.79E-05	4.55	0.01	0.51	0.15	2.32	7.94	$\%CY_{SOLIDS}$
43	Hapmap48306-BTA-36540	28	38449335	-0.26	2.63E-05	4.58	0.02	0.77	0.20	2.11	8.29	$\%CY_{WATER}$

SNP= the name of the single nucleotide polymorphism; BTA= *Bos taurus* autosome chromosome; Location= position of the SNP on the chromosome in base pairs on UMD3.1 (http://www.ensembl.org/index.html); Pc1df= *P*-values adjusted for genomic control; LOG= the $-\log_{10}$ of Pc1df; effB= effect of the minor allele (B allele); V_{SNP} = variance explained by the SNP (calculated as $2pqa^2$, where p is the frequency of one allele, q=1-p is the frequency of the second allele and a is the additive genetic effect); V_P= phenotypic variance; V_G= additive genetic variance; V_{Psnp} (%)= percentage of phenotypic variance explained by each SNP; V_{Gsnp} (%)= percentage of additive genetic variance explained by each SNP; Trait= name of the trait

Figure S1. Quantile-quantile (Q-Q) plots of the observed test statistics of the genome wide association studies (GWAS).



CHAPTER 3

Table S1. Significant genes involved in the enriched categories.

SNP	BTA	Location(bp)	Gene_Ensembl_ID	Gene_SYMBOL	Category	Trait
ARS-BFGL-NGS-72483	19	63501603	ENSBTAG00000001061	PRKCA	KEGG:bta04070	RCT
Hapmap36127-SCAFFOLD59715_196	5	83445048	ENSBTAG00000002313	ITPR2	KEGG:bta04070	RCT
Hapmap27855-BTA-123415	13	806792	ENSBTAG00000008338	PLCB1	KEGG:bta04070	RCT
ARS-BFGL-NGS-19964	29	43189370	ENSBTAG00000012510	PLCB3	KEGG:bta04070	RCT
ARS-BFGL-BAC-11281	13	2217706	ENSBTAG00000013116	PLCB4	KEGG:bta04070	RCT
Hapmap58728-rs29027441	12	12375872	ENSBTAG00000013879	DGKH	KEGG:bta04070	RCT
BTB-01082458	17	16032820	ENSBTAG00000014111	INPP4B	KEGG:bta04070	RCT
ARS-BFGL-NGS-112820	17	72084838	ENSBTAG00000017023	INPP5J	KEGG:bta04070	RCT
UA-IFASA-6532	22	21770149	ENSBTAG00000020455	ITPR1	KEGG:bta04070	RCT
ARS-BFGL-NGS-111260	5	91947526	ENSBTAG00000020715	PIK3C2G	KEGG:bta04070	RCT
ARS-BFGL-NGS-72483	19	63501603	ENSBTAG00000001061	PRKCA	KEGG:bta04730	RCT
Hapmap36127-SCAFFOLD59715_196	5	83445048	ENSBTAG00000002313	ITPR2	KEGG:bta04730	RCT
Hapmap27855-BTA-123415	13	806792	ENSBTAG00000008338	PLCB1	KEGG:bta04730	RCT
ARS-BFGL-NGS-19964	29	43189370	ENSBTAG00000012510	PLCB3	KEGG:bta04730	RCT
ARS-BFGL-BAC-11281	13	2217706	ENSBTAG00000013116	PLCB4	KEGG:bta04730	RCT
ARS-BFGL-NGS-112679	7	13336301	ENSBTAG00000014828	CACNA1A	KEGG:bta04730	RCT
UA-IFASA-6532	22	21770149	ENSBTAG00000020455	ITPR1	KEGG:bta04730	RCT
ARS-BFGL-NGS-33216	15	16809081	ENSBTAG00000044144	GUCY1A2	KEGG:bta04730	RCT
ARS-BFGL-NGS-11874	10	14388574	ENSBTAG00000000218	MAP2K5	KEGG:bta04540	RCT
ARS-BFGL-NGS-72483	19	63501603	ENSBTAG00000001061	PRKCA	KEGG:bta04540	RCT
Hapmap36127-SCAFFOLD59715_196	5	83445048	ENSBTAG00000002313	ITPR2	KEGG:bta04540	RCT
Hapmap27855-BTA-123415	13	806792	ENSBTAG00000008338	PLCB1	KEGG:bta04540	RCT
ARS-BFGL-NGS-19964	29	43189370	ENSBTAG00000012510	PLCB3	KEGG:bta04540	RCT
ARS-BFGL-BAC-11281	13	2217706	ENSBTAG00000013116	PLCB4	KEGG:bta04540	RCT
UA-IFASA-6532	22	21770149	ENSBTAG00000020455	ITPR1	KEGG:bta04540	RCT

BTB-01430463	15	4586930	ENSBTAG00000034827	PDGFD	KEGG:bta04540	RCT
ARS-BFGL-NGS-33216	15	16809081	ENSBTAG00000044144	GUCY1A2	KEGG:bta04540	RCT
ARS-BFGL-NGS-72483	19	63501603	ENSBTAG00000001061	PRKCA	KEGG:bta04270	RCT
Hapmap36127-SCAFFOLD59715_196	5	83445048	ENSBTAG00000002313	ITPR2	KEGG:bta04270	RCT
UA-IFASA-9742	15	42657355	ENSBTAG00000007129	MRVI1	KEGG:bta04270	RCT
Hapmap27855-BTA-123415	13	806792	ENSBTAG00000008338	PLCB1	KEGG:bta04270	RCT
ARS-BFGL-NGS-19964	29	43189370	ENSBTAG00000012510	PLCB3	KEGG:bta04270	RCT
ARS-BFGL-BAC-11281	13	2217706	ENSBTAG00000013116	PLCB4	KEGG:bta04270	RCT
BTA-22976-no-rs	4	99503031	ENSBTAG00000013953	CALD1	KEGG:bta04270	RCT
ARS-BFGL-NGS-114211	18	15099438	ENSBTAG00000014818	MYLK3	KEGG:bta04270	RCT
UA-IFASA-6532	22	21770149	ENSBTAG00000020455	ITPR1	KEGG:bta04270	RCT
ARS-BFGL-NGS-33216	15	16809081	ENSBTAG00000044144	GUCY1A2	KEGG:bta04270	RCT
ARS-BFGL-NGS-72483	19	63501603	ENSBTAG00000001061	PRKCA	KEGG:bta04020	RCT
Hapmap36127-SCAFFOLD59715_196	5	83445048	ENSBTAG00000002313	ITPR2	KEGG:bta04020	RCT
Hapmap27855-BTA-123415	13	806792	ENSBTAG00000008338	PLCB1	KEGG:bta04020	RCT
Hapmap49158-BTA-41145	17	56234762	ENSBTAG00000008948	P2RX7	KEGG:bta04020	RCT
ARS-BFGL-NGS-41667	17	56181953	ENSBTAG00000010812	P2RX4	KEGG:bta04020	RCT
BTA-97453-no-rs	2	14283570	ENSBTAG00000012100	PDE1A	KEGG:bta04020	RCT
ARS-BFGL-NGS-19964	29	43189370	ENSBTAG00000012510	PLCB3	KEGG:bta04020	RCT
ARS-BFGL-BAC-11281	13	2217706	ENSBTAG00000013116	PLCB4	KEGG:bta04020	RCT
ARS-BFGL-NGS-114211	18	15099438	ENSBTAG00000014818	MYLK3	KEGG:bta04020	RCT
ARS-BFGL-NGS-112679	7	13336301	ENSBTAG00000014828	CACNA1A	KEGG:bta04020	RCT
ARS-BFGL-NGS-83866	11	30851124	ENSBTAG00000016573	LHCGR	KEGG:bta04020	RCT
UA-IFASA-6532	22	21770149	ENSBTAG00000020455	ITPR1	KEGG:bta04020	RCT
Hapmap60761-rs29011573	28	10227920	ENSBTAG00000022886	RYR2	KEGG:bta04020	RCT
ARS-BFGL-NGS-72483	19	63501603	ENSBTAG00000001061	PRKCA	KEGG:bta04970	RCT
Hapmap36127-SCAFFOLD59715_196	5	83445048	ENSBTAG00000002313	ITPR2	KEGG:bta04970	RCT
Hapmap27855-BTA-123415	13	806792	ENSBTAG00000008338	PLCB1	KEGG:bta04970	RCT
ARS-BFGL-NGS-19964	29	43189370	ENSBTAG00000012510	PLCB3	KEGG:bta04970	RCT
ARS-BFGL-BAC-11281	13	2217706	ENSBTAG00000013116	PLCB4	KEGG:bta04970	RCT
UA-IFASA-6532	22	21770149	ENSBTAG00000020455	ITPR1	KEGG:bta04970	RCT

ARS-BFGL-NGS-33216	15	16809081	ENSBTAG00000044144	GUCY1A2	KEGG:bta04970	RCT
Hapmap28023-BTC-060518	6	87201599	ENSBTAG00000048250	HSTN	KEGG:bta04970	RCT
ARS-BFGL-NGS-31435	4	57296533	ENSBTAG00000004398	IMMP2L	GO_BP:0048511	RCTeq
Hapmap53648-rs29021240	11	86569656	ENSBTAG00000005847	ROCK2	GO_BP:0048511	RCTeq
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0048511	RCTeq
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_BP:0048511	RCTeq
ARS-BFGL-NGS-899	28	33328952	ENSBTAG00000013300	KCNMA1	GO_BP:0048511	RCTeq
ARS-BFGL-NGS-83866	11	30851124	ENSBTAG00000016573	LHCGR	GO_BP:0048511	RCTeq
Hapmap46359-BTA-26846	22	30044319	ENSBTAG00000019330	PROK2	GO_BP:0048511	RCTeq
ARS-BFGL-NGS-23358	11	31259588	ENSBTAG00000032424	FSHR	GO_BP:0048511	RCTeq
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0048511	RCTeq
ARS-BFGL-NGS-31435	4	57296533	ENSBTAG00000004398	IMMP2L	GO_BP:0008585	RCTeq
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0008585	RCTeq
ARS-BFGL-NGS-83866	11	30851124	ENSBTAG00000016573	LHCGR	GO_BP:0008585	RCTeq
ARS-BFGL-NGS-23358	11	31259588	ENSBTAG00000032424	FSHR	GO_BP:0008585	RCTeq
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0008585	RCTeq
ARS-BFGL-NGS-31435	4	57296533	ENSBTAG00000004398	IMMP2L	GO_BP:0022602	RCTeq
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0022602	RCTeq
ARS-BFGL-NGS-83866	11	30851124	ENSBTAG00000016573	LHCGR	GO_BP:0022602	RCTeq
ARS-BFGL-NGS-23358	11	31259588	ENSBTAG00000032424	FSHR	GO_BP:0022602	RCTeq
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0022602	RCTeq
ARS-BFGL-NGS-31435	4	57296533	ENSBTAG00000004398	IMMP2L	GO_BP:0042698	RCTeq
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0042698	RCTeq
ARS-BFGL-NGS-83866	11	30851124	ENSBTAG00000016573	LHCGR	GO_BP:0042698	RCTeq
ARS-BFGL-NGS-23358	11	31259588	ENSBTAG00000032424	FSHR	GO_BP:0042698	RCTeq
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0042698	RCTeq
ARS-BFGL-NGS-31435	4	57296533	ENSBTAG00000004398	IMMP2L	GO_BP:0046545	RCTeq
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0046545	RCTeq
ARS-BFGL-NGS-83866	11	30851124	ENSBTAG00000016573	LHCGR	GO_BP:0046545	RCTeq
ARS-BFGL-NGS-23358	11	31259588	ENSBTAG00000032424	FSHR	GO_BP:0046545	RCTeq
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0046545	RCTeq

ARS-BFGL-NGS-31435	4	57296533	ENSBTAG00000004398	IMMP2L	GO_BP:0046660	RCTeq
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0046660	RCTeq
ARS-BFGL-NGS-83866	11	30851124	ENSBTAG00000016573	LHCGR	GO_BP:0046660	RCTeq
ARS-BFGL-NGS-23358	11	31259588	ENSBTAG00000032424	FSHR	GO_BP:0046660	RCTeq
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0046660	RCTeq
ARS-BFGL-NGS-64156	18	24359982	ENSBTAG00000003794	GNAO1	GO_CC:0030425	RCTeq
ARS-BFGL-NGS-36950	28	6499231	ENSBTAG00000004515	KCNK1	GO_CC:0030425	RCTeq
ARS-BFGL-BAC-35294	16	36284584	ENSBTAG00000005410	RGS7	GO_CC:0030425	RCTeq
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_CC:0030425	RCTeq
Hapmap28049-BTA-148967	10	43956417	ENSBTAG00000010103	TRIM9	GO_CC:0030425	RCTeq
Hapmap50362-BTA-44583	19	17701476	ENSBTAG00000015527	MYO1D	GO_CC:0030425	RCTeq
ARS-BFGL-NGS-16742	3	43720374	ENSBTAG00000017655	PALMD	GO_CC:0030425	RCTeq
BTB-00358604	8	75109746	ENSBTAG00000018373	DPYSL2	GO_CC:0030425	RCTeq
ARS-BFGL-NGS-90129	26	18776144	ENSBTAG00000018564	ZFYVE27	GO_CC:0030425	RCTeq
BTA-77154-no-rs	6	93551941	ENSBTAG00000021372	SEPT11	GO_CC:0030425	RCTeq
Hapmap23274-BTA-154879	14	38155245	ENSBTAG00000040496	KCNB2	GO_CC:0030425	RCTeq
BTA-42522-no-rs	18	15039844	ENSBTAG00000002493	VPS35	GO_CC:0044456	RCTeq
ARS-BFGL-NGS-116126	3	94980177	ENSBTAG00000005337	RAB3B	GO_CC:0044456	RCTeq
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_CC:0044456	RCTeq
ARS-BFGL-NGS-34068	3	75737505	ENSBTAG00000009338	LRRC7	GO_CC:0044456	RCTeq
Hapmap28049-BTA-148967	10	43956417	ENSBTAG00000010103	TRIM9	GO_CC:0044456	RCTeq
Hapmap57288-ss46526584	12	90621015	ENSBTAG00000010242	LAMP1	GO_CC:0044456	RCTeq
ARS-BFGL-NGS-899	28	33328952	ENSBTAG00000013300	KCNMA1	GO_CC:0044456	RCTeq
ARS-BFGL-BAC-3741	25	34055277	ENSBTAG00000017075	STX1A	GO_CC:0044456	RCTeq
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_CC:0044456	RCTeq
ARS-BFGL-NGS-16742	3	43720374	ENSBTAG00000017655	PALMD	GO_CC:0044456	RCTeq
ARS-BFGL-NGS-12343	7	75305297	ENSBTAG00000018585	GABRB2	GO_CC:0044456	RCTeq
BTA-47828-no-rs	2	62977592	ENSBTAG00000018753	TMEM163	GO_CC:0044456	RCTeq
BTA-77154-no-rs	6	93551941	ENSBTAG00000021372	SEPT11	GO_CC:0044456	RCTeq
ARS-BFGL-NGS-28269	5	8926782	ENSBTAG00000034693	SYT1	GO_CC:0044456	RCTeq
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0044456	RCTeq

BTA-71636-no-rs	4	94085456	ENSBTAG00000001754	AHCYL2	GO_CC:0097458	RCTeq
BTA-42522-no-rs	18	15039844	ENSBTAG00000002493	VPS35	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-64156	18	24359982	ENSBTAG00000003794	GNAO1	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-36950	28	6499231	ENSBTAG00000004515	KCNK1	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-116126	3	94980177	ENSBTAG00000005337	RAB3B	GO_CC:0097458	RCTeq
ARS-BFGL-BAC-35294	16	36284584	ENSBTAG00000005410	RGS7	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-42578	11	93241685	ENSBTAG00000006716	PTGS1	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_CC:0097458	RCTeq
Hapmap28049-BTA-148967	10	43956417	ENSBTAG00000010103	TRIM9	GO_CC:0097458	RCTeq
Hapmap57288-ss46526584	12	90621015	ENSBTAG00000010242	LAMP1	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-899	28	33328952	ENSBTAG00000013300	KCNMA1	GO_CC:0097458	RCTeq
Hapmap50362-BTA-44583	19	17701476	ENSBTAG00000015527	MYO1D	GO_CC:0097458	RCTeq
ARS-BFGL-BAC-3741	25	34055277	ENSBTAG00000017075	STX1A	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-16742	3	43720374	ENSBTAG00000017655	PALMD	GO_CC:0097458	RCTeq
BTB-00358604	8	75109746	ENSBTAG00000018373	DPYSL2	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-90129	26	18776144	ENSBTAG00000018564	ZFYVE27	GO_CC:0097458	RCTeq
BTA-47828-no-rs	2	62977592	ENSBTAG00000018753	TMEM163	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-102000	7	16675279	ENSBTAG00000019220	SMARCA4	GO_CC:0097458	RCTeq
BTA-77154-no-rs	6	93551941	ENSBTAG00000021372	SEPT11	GO_CC:0097458	RCTeq
ARS-BFGL-NGS-28269	5	8926782	ENSBTAG00000034693	SYT1	GO_CC:0097458	RCTeq
Hapmap23274-BTA-154879	14	38155245	ENSBTAG00000040496	KCNB2	GO_CC:0097458	RCTeq
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0097458	RCTeq
BTA-42522-no-rs	18	15039844	ENSBTAG00000002493	VPS35	GO_CC:0045202	RCTeq
ARS-BFGL-NGS-36950	28	6499231	ENSBTAG00000004515	KCNK1	GO_CC:0045202	RCTeq
ARS-BFGL-NGS-116126	3	94980177	ENSBTAG00000005337	RAB3B	GO_CC:0045202	RCTeq
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_CC:0045202	RCTeq
ARS-BFGL-NGS-34068	3	75737505	ENSBTAG00000009338	LRRC7	GO_CC:0045202	RCTeq
Hapmap28049-BTA-148967	10	43956417	ENSBTAG00000010103	TRIM9	GO_CC:0045202	RCTeq
Hapmap57288-ss46526584	12	90621015	ENSBTAG00000010242	LAMP1	GO_CC:0045202	RCTeq
ARS-BFGL-NGS-899	28	33328952	ENSBTAG00000013300	KCNMA1	GO_CC:0045202	RCTeq
ARS-BFGL-BAC-3741	25	34055277	ENSBTAG00000017075	STX1A	GO_CC:0045202	RCTeq

ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_CC:0045202	RCTeq
ARS-BFGL-NGS-16742	3	43720374	ENSBTAG00000017655	PALMD	GO_CC:0045202	RCTeq
ARS-BFGL-NGS-12343	7	75305297	ENSBTAG00000018585	GABRB2	GO_CC:0045202	RCTeq
BTA-47828-no-rs	2	62977592	ENSBTAG00000018753	TMEM163	GO_CC:0045202	RCTeq
BTA-77154-no-rs	6	93551941	ENSBTAG00000021372	SEPT11	GO_CC:0045202	RCTeq
ARS-BFGL-NGS-28269	5	8926782	ENSBTAG00000034693	SYT1	GO_CC:0045202	RCTeq
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0045202	RCTeq
ARS-BFGL-NGS-64156	18	24359982	ENSBTAG00000003794	GNAO1	GO_CC:0036477	RCTeq
ARS-BFGL-NGS-36950	28	6499231	ENSBTAG00000004515	KCNK1	GO_CC:0036477	RCTeq
ARS-BFGL-BAC-35294	16	36284584	ENSBTAG00000005410	RGS7	GO_CC:0036477	RCTeq
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_CC:0036477	RCTeq
Hapmap28049-BTA-148967	10	43956417	ENSBTAG00000010103	TRIM9	GO_CC:0036477	RCTeq
Hapmap50362-BTA-44583	19	17701476	ENSBTAG00000015527	MYO1D	GO_CC:0036477	RCTeq
ARS-BFGL-NGS-16742	3	43720374	ENSBTAG00000017655	PALMD	GO_CC:0036477	RCTeq
BTB-00358604	8	75109746	ENSBTAG00000018373	DPYSL2	GO_CC:0036477	RCTeq
ARS-BFGL-NGS-90129	26	18776144	ENSBTAG00000018564	ZFYVE27	GO_CC:0036477	RCTeq
BTA-77154-no-rs	6	93551941	ENSBTAG00000021372	SEPT11	GO_CC:0036477	RCTeq
Hapmap23274-BTA-154879	14	38155245	ENSBTAG00000040496	KCNB2	GO_CC:0036477	RCTeq
BTA-71636-no-rs	4	94085456	ENSBTAG00000001754	AHCYL2	GO_CC:0043005	RCTeq
ARS-BFGL-NGS-64156	18	24359982	ENSBTAG00000003794	GNAO1	GO_CC:0043005	RCTeq
ARS-BFGL-NGS-36950	28	6499231	ENSBTAG00000004515	KCNK1	GO_CC:0043005	RCTeq
ARS-BFGL-BAC-35294	16	36284584	ENSBTAG00000005410	RGS7	GO_CC:0043005	RCTeq
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_CC:0043005	RCTeq
Hapmap28049-BTA-148967	10	43956417	ENSBTAG00000010103	TRIM9	GO_CC:0043005	RCTeq
ARS-BFGL-NGS-899	28	33328952	ENSBTAG00000013300	KCNMA1	GO_CC:0043005	RCTeq
Hapmap50362-BTA-44583	19	17701476	ENSBTAG00000015527	MYO1D	GO_CC:0043005	RCTeq
ARS-BFGL-BAC-3741	25	34055277	ENSBTAG00000017075	STX1A	GO_CC:0043005	RCTeq
ARS-BFGL-NGS-16742	3	43720374	ENSBTAG00000017655	PALMD	GO_CC:0043005	RCTeq
BTB-00358604	8	75109746	ENSBTAG00000018373	DPYSL2	GO_CC:0043005	RCTeq
ARS-BFGL-NGS-90129	26	18776144	ENSBTAG00000018564	ZFYVE27	GO_CC:0043005	RCTeq
BTA-77154-no-rs	6	93551941	ENSBTAG00000021372	SEPT11	GO_CC:0043005	RCTeq

ARS-BFGL-NGS-28269	5	8926782	ENSBTAG00000034693	SYT1	GO_CC:0043005	RCTeq
Hapmap23274-BTA-154879	14	38155245	ENSBTAG00000040496	KCNB2	GO_CC:0043005	RCTeq
ARS-BFGL-NGS-36950	28	6499231	ENSBTAG00000004515	KCNK1	GO_CC:0008076	RCTeq
ARS-BFGL-NGS-15423	16	74158269	ENSBTAG00000008710	KCNH1	GO_CC:0008076	RCTeq
ARS-BFGL-NGS-899	28	33328952	ENSBTAG00000013300	KCNMA1	GO_CC:0008076	RCTeq
ARS-BFGL-BAC-3741	25	34055277	ENSBTAG00000017075	STX1A	GO_CC:0008076	RCTeq
Hapmap23274-BTA-154879	14	38155245	ENSBTAG00000040496	KCNB2	GO_CC:0008076	RCTeq
ARS-BFGL-NGS-36950	28	6499231	ENSBTAG00000004515	KCNK1	GO_CC:0034705	RCTeq
ARS-BFGL-NGS-15423	16	74158269	ENSBTAG00000008710	KCNH1	GO_CC:0034705	RCTeq
ARS-BFGL-NGS-899	28	33328952	ENSBTAG00000013300	KCNMA1	GO_CC:0034705	RCTeq
ARS-BFGL-BAC-3741	25	34055277	ENSBTAG00000017075	STX1A	GO_CC:0034705	RCTeq
Hapmap23274-BTA-154879	14	38155245	ENSBTAG00000040496	KCNB2	GO_CC:0034705	RCTeq
ARS-BFGL-NGS-116126	3	94980177	ENSBTAG00000005337	RAB3B	GO_CC:0008021	RCTeq
Hapmap28049-BTA-148967	10	43956417	ENSBTAG00000010103	TRIM9	GO_CC:0008021	RCTeq
Hapmap57288-ss46526584	12	90621015	ENSBTAG00000010242	LAMP1	GO_CC:0008021	RCTeq
ARS-BFGL-BAC-3741	25	34055277	ENSBTAG00000017075	STX1A	GO_CC:0008021	RCTeq
BTA-47828-no-rs	2	62977592	ENSBTAG00000018753	TMEM163	GO_CC:0008021	RCTeq
ARS-BFGL-NGS-28269	5	8926782	ENSBTAG00000034693	SYT1	GO_CC:0008021	RCTeq
ARS-BFGL-NGS-116126	3	94980177	ENSBTAG00000005337	RAB3B	GO_CC:0098793	RCTeq
Hapmap28049-BTA-148967	10	43956417	ENSBTAG00000010103	TRIM9	GO_CC:0098793	RCTeq
Hapmap57288-ss46526584	12	90621015	ENSBTAG00000010242	LAMP1	GO_CC:0098793	RCTeq
ARS-BFGL-BAC-3741	25	34055277	ENSBTAG00000017075	STX1A	GO_CC:0098793	RCTeq
BTA-47828-no-rs	2	62977592	ENSBTAG00000018753	TMEM163	GO_CC:0098793	RCTeq
ARS-BFGL-NGS-28269	5	8926782	ENSBTAG00000034693	SYT1	GO_CC:0098793	RCTeq
ARS-BFGL-NGS-116126	3	94980177	ENSBTAG00000005337	RAB3B	GO_CC:0044456	Kcf
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_CC:0044456	Kcf
BTB-00135420	3	75690381	ENSBTAG00000009338	LRRC7	GO_CC:0044456	Kcf
BTB-01474679	6	66549547	ENSBTAG00000011817	GABRA2	GO_CC:0044456	Kcf
UA-IFASA-6955	28	29913288	ENSBTAG00000012667	CAMK2G	GO_CC:0044456	Kcf
ARS-BFGL-NGS-106765	28	33405497	ENSBTAG00000013300	KCNMA1	GO_CC:0044456	Kcf
ARS-BFGL-NGS-548	1	9900095	ENSBTAG00000017753	APP	GO_CC:0044456	Kcf

BTB-01068662	6	67643584	ENSBTAG00000017837	GABRB1	GO_CC:0044456	Kcf
BTA-47828-no-rs	2	62977592	ENSBTAG00000018753	TMEM163	GO_CC:0044456	Kcf
BTB-00745347	19	28768067	ENSBTAG00000021151	MYH10	GO_CC:0044456	Kcf
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT11	GO_CC:0044456	Kcf
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0044456	Kcf
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0044456	Kcf
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_CC:0098794	Kcf
BTB-00135420	3	75690381	ENSBTAG00000009338	LRRC7	GO_CC:0098794	Kcf
BTB-01474679	6	66549547	ENSBTAG00000011817	GABRA2	GO_CC:0098794	Kcf
UA-IFASA-6955	28	29913288	ENSBTAG00000012667	CAMK2G	GO_CC:0098794	Kcf
ARS-BFGL-NGS-106765	28	33405497	ENSBTAG00000013300	KCNMA1	GO_CC:0098794	Kcf
ARS-BFGL-NGS-548	1	9900095	ENSBTAG00000017753	APP	GO_CC:0098794	Kcf
BTB-01068662	6	67643584	ENSBTAG00000017837	GABRB1	GO_CC:0098794	Kcf
BTB-00745347	19	28768067	ENSBTAG00000021151	MYH10	GO_CC:0098794	Kcf
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT11	GO_CC:0098794	Kcf
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0098794	Kcf
BTA-32792-no-rs	5	77371444	ENSBTAG00000002651	PKP2	KEGG:bta05412	%Cysolids
BTB-01084639	2	15130630	ENSBTAG00000009256	ITGA4	KEGG:bta05412	%Cysolids
ARS-BFGL-NGS-5170	22	47762087	ENSBTAG00000010026	CACNA1D	KEGG:bta05412	%Cysolids
ARS-BFGL-NGS-100053	5	109167093	ENSBTAG00000010660	CACNA1C	KEGG:bta05412	%Cysolids
ARS-BFGL-NGS-44080	5	75361846	ENSBTAG00000010865	CACNG2	KEGG:bta05412	%Cysolids
ARS-BFGL-NGS-26419	22	46607237	ENSBTAG00000013117	CACNA2D3	KEGG:bta05412	%Cysolids
BTB-00437757	10	81069177	ENSBTAG00000018255	ACTN1	KEGG:bta05412	%Cysolids
ARS-BFGL-NGS-18294	5	108746845	ENSBTAG00000021994	CACNA2D4	KEGG:bta05412	%Cysolids
ARS-BFGL-NGS-25890	28	10044965	ENSBTAG00000022886	RYR2	KEGG:bta05412	%Cysolids
BTB-00944226	28	22989068	ENSBTAG00000045699	CTNNA3	KEGG:bta05412	%Cysolids
ARS-BFGL-NGS-40683	3	100402643	ENSBTAG00000009603	UQCRH	KEGG:bta04260	%Cysolids
ARS-BFGL-NGS-5170	22	47762087	ENSBTAG00000010026	CACNA1D	KEGG:bta04260	%Cysolids
ARS-BFGL-NGS-100053	5	109167093	ENSBTAG00000010660	CACNA1C	KEGG:bta04260	%Cysolids
ARS-BFGL-NGS-44080	5	75361846	ENSBTAG00000010865	CACNG2	KEGG:bta04260	%Cysolids
ARS-BFGL-NGS-26419	22	46607237	ENSBTAG00000013117	CACNA2D3	KEGG:bta04260	%Cysolids

ARS-BFGL-NGS-117685	18	51592949	ENSBTAG00000018635	ATP1A3	KEGG:bta04260	%Cysolids
ARS-BFGL-NGS-18294	5	108746845	ENSBTAG00000021994	CACNA2D4	KEGG:bta04260	%Cysolids
ARS-BFGL-NGS-25890	28	10044965	ENSBTAG00000022886	RYR2	KEGG:bta04260	%Cysolids
UA-IFASA-7801	5	83492863	ENSBTAG00000002313	ITPR2	KEGG:bta04020	RECfat
ARS-BFGL-NGS-71458	13	1053585	ENSBTAG00000008338	PLCB1	KEGG:bta04020	RECfat
BTA-90299-no-rs	26	12463857	ENSBTAG00000008783	HTR7	KEGG:bta04020	RECfat
ARS-BFGL-NGS-100053	5	109167093	ENSBTAG00000010660	CACNA1C	KEGG:bta04020	RECfat
ARS-BFGL-NGS-98581	2	14403663	ENSBTAG00000012100	PDE1A	KEGG:bta04020	RECfat
ARS-BFGL-NGS-114211	18	15099438	ENSBTAG00000014818	MYLK3	KEGG:bta04020	RECfat
UA-IFASA-1667	7	13489100	ENSBTAG00000014828	CACNA1A	KEGG:bta04020	RECfat
ARS-BFGL-NGS-26976	19	24995142	ENSBTAG00000015258	P2RX5	KEGG:bta04020	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	KEGG:bta04020	RECfat
BTA-18314-no-rs	26	15383866	ENSBTAG00000018966	PLCE1	KEGG:bta04020	RECfat
BTB-00631715	8	53743552	ENSBTAG00000021127	GNA14	KEGG:bta04020	RECfat
ARS-BFGL-NGS-35163	2	125798507	ENSBTAG00000027051	PTAFR	KEGG:bta04020	RECfat
ARS-BFGL-NGS-110110	10	20693677	ENSBTAG00000040056	LTB4R2	KEGG:bta04020	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0008585	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0008585	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0008585	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0008585	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0008585	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0008585	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0022602	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0022602	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0022602	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0022602	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0022602	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0022602	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0042698	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0042698	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0042698	RECfat

BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0042698	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0042698	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0042698	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0046545	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0046545	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0046545	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0046545	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0046545	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0046545	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0046660	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0046660	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0046660	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0046660	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0046660	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0046660	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0001541	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0001541	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0001541	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0001541	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0001541	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0008406	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0008406	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0008406	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0008406	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0008406	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0008406	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0045137	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0045137	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0045137	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0045137	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0045137	RECfat

ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0045137	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0007548	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0007548	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0007548	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0007548	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0007548	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0007548	RECfat
ARS-BFGL-NGS-112996	4	79996650	ENSBTAG00000002912	INHBA	GO_BP:0048511	RECfat
Hapmap54988-rs29020907	4	57259091	ENSBTAG00000004398	IMMP2L	GO_BP:0048511	RECfat
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0048511	RECfat
ARS-BFGL-NGS-100288	4	114431179	ENSBTAG00000007766	CDK5	GO_BP:0048511	RECfat
ARS-BFGL-NGS-15765	23	15005574	ENSBTAG00000009905	NFYA	GO_BP:0048511	RECfat
BTB-01944534	11	30826527	ENSBTAG00000016573	LHCGR	GO_BP:0048511	RECfat
BTB-00470654	11	31130270	ENSBTAG00000032424	FSHR	GO_BP:0048511	RECfat
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0048511	RECfat
BTA-32792-no-rs	5	77371444	ENSBTAG00000002651	PKP2	KEGG:bta05412	RECsolids
BTB-01084639	2	15130630	ENSBTAG00000009256	ITGA4	KEGG:bta05412	RECsolids
ARS-BFGL-NGS-5170	22	47762087	ENSBTAG00000010026	CACNA1D	KEGG:bta05412	RECsolids
ARS-BFGL-NGS-100053	5	109167093	ENSBTAG00000010660	CACNA1C	KEGG:bta05412	RECsolids
ARS-BFGL-NGS-44080	5	75361846	ENSBTAG00000010865	CACNG2	KEGG:bta05412	RECsolids
ARS-BFGL-NGS-26419	22	46607237	ENSBTAG00000013117	CACNA2D3	KEGG:bta05412	RECsolids
BTB-00437757	10	81069177	ENSBTAG00000018255	ACTN1	KEGG:bta05412	RECsolids
ARS-BFGL-NGS-25890	28	10044965	ENSBTAG00000022886	RYR2	KEGG:bta05412	RECsolids
BTB-00944226	28	22989068	ENSBTAG00000045699	CTNNA3	KEGG:bta05412	RECsolids
BTA-20822-no-rs	3	100584312	ENSBTAG00000002979	PIK3R3	KEGG:bta05218	RECsolids
ARS-BFGL-NGS-27069	7	55586641	ENSBTAG00000005198	FGF1	KEGG:bta05218	RECsolids
ARS-BFGL-NGS-81552	19	28984588	ENSBTAG00000005978	PIK3R5	KEGG:bta05218	RECsolids
Hapmap24263-BTA-161141	4	39120348	ENSBTAG00000017664	HGF	KEGG:bta05218	RECsolids
BTB-00632811	16	34181806	ENSBTAG00000017788	AKT3	KEGG:bta05218	RECsolids
ARS-BFGL-NGS-15268	7	44879216	ENSBTAG00000027357	FGF22	KEGG:bta05218	RECsolids
ARS-BFGL-NGS-29703	15	4692182	ENSBTAG00000034827	PDGFD	KEGG:bta05218	RECsolids

ARS-BFGL-NGS-112129	4	9837443	ENSBTAG00000044023	CDK6	KEGG:bta05218	RECsolids
ARS-BFGL-NGS-108629	19	63530174	ENSBTAG00000001061	PRKCA	KEGG:bta05200	RECsolids
BTB-01244791	4	8495236	ENSBTAG00000002107	FZD1	KEGG:bta05200	RECsolids
BTA-20822-no-rs	3	100584312	ENSBTAG00000002979	PIK3R3	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-107257	1	148756389	ENSBTAG00000004742	RUNX1	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-27069	7	55586641	ENSBTAG00000005198	FGF1	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-81552	19	28984588	ENSBTAG00000005978	PIK3R5	KEGG:bta05200	RECsolids
BTB-00180381	4	49318278	ENSBTAG00000011412	LAMB1	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-16738	5	109655733	ENSBTAG00000013988	BID	KEGG:bta05200	RECsolids
BTA-113935-no-rs	17	13686681	ENSBTAG00000016071	HHIP	KEGG:bta05200	RECsolids
Hapmap47325-BTA-55227	22	10563792	ENSBTAG00000016758	MLH1	KEGG:bta05200	RECsolids
Hapmap24263-BTA-161141	4	39120348	ENSBTAG00000017664	HGF	KEGG:bta05200	RECsolids
BTB-00632811	16	34181806	ENSBTAG00000017788	AKT3	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-112953	10	1206871	ENSBTAG00000018852	APC	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-34712	11	46794995	ENSBTAG00000019354	PAX8	KEGG:bta05200	RECsolids
Hapmap23975-BTC-043815	6	102937469	ENSBTAG00000020048	MAPK10	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-15268	7	44879216	ENSBTAG00000027357	FGF22	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-112129	4	9837443	ENSBTAG00000044023	CDK6	KEGG:bta05200	RECsolids
BTB-00944226	28	22989068	ENSBTAG00000045699	CTNNA3	KEGG:bta05200	RECsolids
ARS-BFGL-NGS-87906	29	9064567	ENSBTAG00000000246	ME3	KEGG:bta01100	RECsolids
BTA-98453-no-rs	5	88436433	ENSBTAG00000000593	ST8SIA1	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-81463	3	105819329	ENSBTAG00000001626	CTPS1	KEGG:bta01100	RECsolids
BTB-00635473	16	37748202	ENSBTAG00000002689	NME7	KEGG:bta01100	RECsolids
BTA-116788-no-rs	15	41971248	ENSBTAG00000002914	GALNT18	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-96503	3	107006890	ENSBTAG00000002983	NT5C1A	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-15769	4	15160644	ENSBTAG00000003222	ASNS	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-79934	3	48692833	ENSBTAG00000004422	ALG14	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-107813	18	34379769	ENSBTAG00000005152	TK2	KEGG:bta01100	RECsolids
BTB-01648093	14	26352020	ENSBTAG00000005287	CYP7A1	KEGG:bta01100	RECsolids
Hapmap50154-BTA-91586	2	41874101	ENSBTAG00000005562	GALNT13	KEGG:bta01100	RECsolids
Hapmap35881-SCAFFOLD20653_10639	14	48380429	ENSBTAG00000006209	EXT1	KEGG:bta01100	RECsolids

Hapmap56942-ss46526051	26	33234819	ENSBTAG00000006707	ACSL5	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-107510	7	19352717	ENSBTAG00000008095	ACER1	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-112419	8	49389656	ENSBTAG00000008103	ALDH1A1	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-71458	13	1053585	ENSBTAG00000008338	PLCB1	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-101697	10	9900839	ENSBTAG00000008341	ARSB	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-79824	21	5821878	ENSBTAG00000009125	ALDH1A3	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-40683	3	100402643	ENSBTAG00000009603	UQCRH	KEGG:bta01100	RECsolids
BTB-00424333	10	52372286	ENSBTAG00000010119	ALDH1A2	KEGG:bta01100	RECsolids
Hapmap38797-BTA-99366	24	21743914	ENSBTAG00000011206	GALNT1	KEGG:bta01100	RECsolids
BTA-82062-no-rs	8	85665183	ENSBTAG00000011859	IPPK	KEGG:bta01100	RECsolids
BTB-01381318	23	51118713	ENSBTAG00000012058	GMDS	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-112872	6	88069548	ENSBTAG00000012397	DCK	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-97709	18	1680104	ENSBTAG00000012985	FUK	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-115685	6	115813300	ENSBTAG00000013569	CD38	KEGG:bta01100	RECsolids
ARS-BFGL-BAC-34398	18	53393358	ENSBTAG00000013921	CKM	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-26880	1	134090244	ENSBTAG00000015221	PCCB	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-86323	13	74478509	ENSBTAG00000016122	PIGT	KEGG:bta01100	RECsolids
Hapmap47835-BTA-112028	23	37534897	ENSBTAG00000016519	MBOAT1	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-6067	2	98764258	ENSBTAG00000016662	CPS1	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-95132	3	15724767	ENSBTAG00000017819	PMVK	KEGG:bta01100	RECsolids
Hapmap49978-BTA-37376	15	66235937	ENSBTAG00000018261	PDHX	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-115485	2	63181292	ENSBTAG00000018744	MGAT5	KEGG:bta01100	RECsolids
BTA-107196-no-rs	1	82468782	ENSBTAG00000019625	EHHADH	KEGG:bta01100	RECsolids
Hapmap47202-BTA-121802	7	45196738	ENSBTAG00000020776	POLR2E	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-116949	14	78949314	ENSBTAG00000021092	ATP6V0D2	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-72285	4	22618815	ENSBTAG00000021905	DGKB	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-5848	17	66164570	ENSBTAG00000022058	ACACB	KEGG:bta01100	RECsolids
Hapmap24162-BTA-160429	5	76501658	ENSBTAG00000030632	ALG10	KEGG:bta01100	RECsolids
UA-IFASA-8370	3	45826189	ENSBTAG00000031358	DPYD	KEGG:bta01100	RECsolids
BTB-01276504	8	4098134	ENSBTAG00000035007	GALNTL6	KEGG:bta01100	RECsolids
ARS-BFGL-NGS-40683	3	100402643	ENSBTAG00000009603	UQCRH	KEGG:bta04260	RECsolids

ARS-BFGL-NGS-5170	22	47762087	ENSBTAG00000010026	CACNA1D	KEGG:bta04260	RECsolids
ARS-BFGL-NGS-100053	5	109167093	ENSBTAG00000010660	CACNA1C	KEGG:bta04260	RECsolids
ARS-BFGL-NGS-44080	5	75361846	ENSBTAG00000010865	CACNG2	KEGG:bta04260	RECsolids
ARS-BFGL-NGS-26419	22	46607237	ENSBTAG00000013117	CACNA2D3	KEGG:bta04260	RECsolids
ARS-BFGL-NGS-117685	18	51592949	ENSBTAG00000018635	ATP1A3	KEGG:bta04260	RECsolids
ARS-BFGL-NGS-25890	28	10044965	ENSBTAG00000022886	RYR2	KEGG:bta04260	RECsolids
Hapmap44892-BTA-50337	1	10102749	ENSBTAG00000000603	JAM2	KEGG:bta04670	RECsolids
ARS-BFGL-NGS-108629	19	63530174	ENSBTAG00000001061	PRKCA	KEGG:bta04670	RECsolids
Hapmap24163-BTA-160726	17	64011938	ENSBTAG00000002048	PTPN11	KEGG:bta04670	RECsolids
BTA-20822-no-rs	3	100584312	ENSBTAG00000002979	PIK3R3	KEGG:bta04670	RECsolids
ARS-BFGL-NGS-81552	19	28984588	ENSBTAG00000005978	PIK3R5	KEGG:bta04670	RECsolids
Hapmap47394-BTA-74024	5	75660485	ENSBTAG00000007531	NCF4	KEGG:bta04670	RECsolids
BTB-01084639	2	15130630	ENSBTAG00000009256	ITGA4	KEGG:bta04670	RECsolids
BTB-00437757	10	81069177	ENSBTAG00000018255	ACTN1	KEGG:bta04670	RECsolids
BTB-00944226	28	22989068	ENSBTAG00000045699	CTNNA3	KEGG:bta04670	RECsolids
BTA-20822-no-rs	3	100584312	ENSBTAG00000002979	PIK3R3	KEGG:bta05213	RECsolids
ARS-BFGL-NGS-81552	19	28984588	ENSBTAG00000005978	PIK3R5	KEGG:bta05213	RECsolids
Hapmap47325-BTA-55227	22	10563792	ENSBTAG00000016758	MLH1	KEGG:bta05213	RECsolids
BTB-00632811	16	34181806	ENSBTAG00000017788	AKT3	KEGG:bta05213	RECsolids
ARS-BFGL-NGS-112953	10	1206871	ENSBTAG00000018852	APC	KEGG:bta05213	RECsolids
BTB-00944226	28	22989068	ENSBTAG00000045699	CTNNA3	KEGG:bta05213	RECsolids
BTA-20822-no-rs	3	100584312	ENSBTAG00000002979	PIK3R3	KEGG:bta04930	RECsolids
ARS-BFGL-NGS-81552	19	28984588	ENSBTAG00000005978	PIK3R5	KEGG:bta04930	RECsolids
ARS-BFGL-NGS-5170	22	47762087	ENSBTAG00000010026	CACNA1D	KEGG:bta04930	RECsolids
ARS-BFGL-NGS-100053	5	109167093	ENSBTAG00000010660	CACNA1C	KEGG:bta04930	RECsolids
UA-IFASA-1667	7	13489100	ENSBTAG00000014828	CACNA1A	KEGG:bta04930	RECsolids
Hapmap23975-BTC-043815	6	102937469	ENSBTAG00000020048	MAPK10	KEGG:bta04930	RECsolids
BTA-20822-no-rs	3	100584312	ENSBTAG00000002979	PIK3R3	KEGG:bta04973	RECsolids
ARS-BFGL-NGS-81552	19	28984588	ENSBTAG00000005978	PIK3R5	KEGG:bta04973	RECsolids
ARS-BFGL-NGS-5170	22	47762087	ENSBTAG00000010026	CACNA1D	KEGG:bta04973	RECsolids
BTB-00632811	16	34181806	ENSBTAG00000017788	AKT3	KEGG:bta04973	RECsolids

ARS-BFGL-NGS-117685	18	51592949	ENSBTAG00000018635	ATP1A3	KEGG:bta04973	RECsolids
Hapmap44892-BTA-50337	1	10102749	ENSBTAG00000000603	JAM2	KEGG:bta04670	RECenergy
Hapmap24163-BTA-160726	17	64011938	ENSBTAG00000002048	PTPN11	KEGG:bta04670	RECenergy
BTA-20822-no-rs	3	100584312	ENSBTAG00000002979	PIK3R3	KEGG:bta04670	RECenergy
ARS-BFGL-NGS-24769	29	33568217	ENSBTAG00000003176	JAM3	KEGG:bta04670	RECenergy
ARS-BFGL-NGS-103383	12	76787101	ENSBTAG00000003568	CLDN10	KEGG:bta04670	RECenergy
Hapmap47394-BTA-74024	5	75660485	ENSBTAG00000007531	NCF4	KEGG:bta04670	RECenergy
BTB-01084639	2	15130630	ENSBTAG00000009256	ITGA4	KEGG:bta04670	RECenergy
ARS-BFGL-NGS-33248	14	3885798	ENSBTAG00000009578	PTK2	KEGG:bta04670	RECenergy
BTB-00437757	10	81069177	ENSBTAG00000018255	ACTN1	KEGG:bta04670	RECenergy
BTA-61748-no-rs	28	22911596	ENSBTAG00000045699	CTNNA3	KEGG:bta04670	RECenergy
Hapmap44892-BTA-50337	1	10102749	ENSBTAG00000000603	JAM2	KEGG:bta04514	RECenergy
BTB-01138108	15	26138126	ENSBTAG00000000977	CADM1	KEGG:bta04514	RECenergy
ARS-BFGL-NGS-24769	29	33568217	ENSBTAG00000003176	JAM3	KEGG:bta04514	RECenergy
ARS-BFGL-NGS-103383	12	76787101	ENSBTAG00000003568	CLDN10	KEGG:bta04514	RECenergy
BTB-01084639	2	15130630	ENSBTAG00000009256	ITGA4	KEGG:bta04514	RECenergy
Hapmap60475-rs29022896	23	7191371	ENSBTAG00000012451	BOLA-DMB	KEGG:bta04514	RECenergy
ARS-BFGL-NGS-4157	13	74391936	ENSBTAG00000015127	SDC4	KEGG:bta04514	RECenergy
Hapmap47139-BTA-95567	16	79553324	ENSBTAG00000023144	PTPRC	KEGG:bta04514	RECenergy
Hapmap25798-BTA-126388	11	32731961	ENSBTAG00000024021		KEGG:bta04514	RECenergy
BTB-01755781	4	112317788	ENSBTAG00000027832		KEGG:bta04514	RECenergy
BTB-01060598	11	32153221	ENSBTAG00000046199		KEGG:bta04514	RECenergy
Hapmap44892-BTA-50337	1	10102749	ENSBTAG00000000603	JAM2	KEGG:bta04530	RECenergy
ARS-BFGL-NGS-65466	13	66009719	ENSBTAG00000001640	EPB41L1	KEGG:bta04530	RECenergy
BTB-00067270	1	141351312	ENSBTAG00000002029	IGSF5	KEGG:bta04530	RECenergy
ARS-BFGL-NGS-24769	29	33568217	ENSBTAG00000003176	JAM3	KEGG:bta04530	RECenergy
ARS-BFGL-NGS-103383	12	76787101	ENSBTAG00000003568	CLDN10	KEGG:bta04530	RECenergy
ARS-BFGL-NGS-116126	3	94980177	ENSBTAG00000005337	RAB3B	KEGG:bta04530	RECenergy
BTA-121254-no-rs	2	125098723	ENSBTAG00000006667	EPB41	KEGG:bta04530	RECenergy
ARS-BFGL-NGS-102915	8	45696078	ENSBTAG00000011770	TJP2	KEGG:bta04530	RECenergy
BTB-00632811	16	34181806	ENSBTAG00000017788	AKT3	KEGG:bta04530	RECenergy

BTB-00437757	10	81069177	ENSBTAG00000018255	ACTN1	KEGG:bta04530	RECenergy
BTA-61748-no-rs	28	22911596	ENSBTAG00000045699	CTNNA3	KEGG:bta04530	RECenergy

SNP= the name of the single nucleotide polymorphism; BTA= *Bos taurus* autosome chromosome; Location= position of the SNP on the chromosome in base pairs on UMD3.1 (http://www.ensembl.org/index.html); RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; RCT_{eq} = Rennet coagulation time (min) estimated using the CF_t equation; k_{CF} = curd-firming rate constant (% x min^{-1}); $\%CY_{SOLIDS}$ = weight of curd solids as percentage of weight of milk processed; REC = fat, solids, and energy of the curd as percentage of the fat, solids, and energy of the milk processed.

KEGG: KEGG pathway; GO_BP.GO biological process; GO_CC: GO cellular component.

The Ensembl *Bos Taurus* UMD3.1 database (http://www.ensembl.org/index.html) was used as reference for SNP location and for mapping the significant SNP to genes.

CHAPTER 5

Figure S1. Heat map of Pearson (above diagonal) and partial (below diagonal) correlations among the 26 variables used in the factor analysis.

Description: The heat map was generated using the *psych* package in R. Partial correlation coefficients were calculated using the *corpcor* R package.

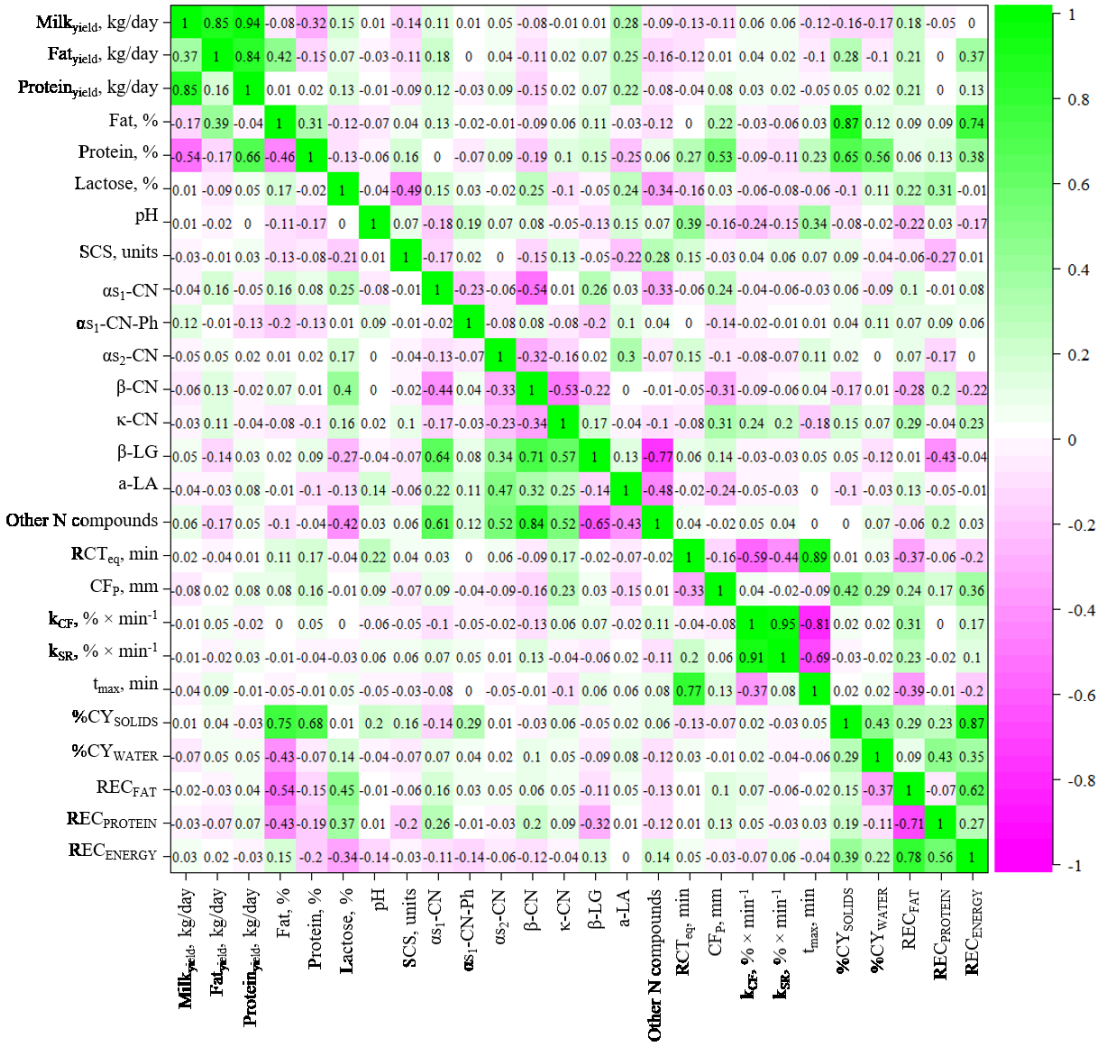


Table S1. Table List of all significant SNP identified in the GWAS analyses sorted by chromosome

N	SNP	BTA	Location (bp)	Pc1df	LOG	effB	V_{SNP}	V_P	V_G	V_{Psnp} (%)	V_{Gsnp} (%)	Trait
1	Hapmap53172-rs29012675	6	82706745	1.39E-06	5.86	0.45	0.02	0.62	0.15	2.94%	12.00%	F1_%CY
2	Hapmap50464-BTA-77021	6	82560990	1.87E-05	4.73	0.25	0.01	0.62	0.15	2.27%	9.25%	F1_%CY
3	ARS-BFGL-NGS-87845	27	42118037	2.38E-05	4.62	0.55	0.02	0.62	0.15	2.53%	10.35%	F1_%CY
4	ARS-BFGL-NGS-102974	19	1822133	4.32E-05	4.36	-0.17	0.01	0.62	0.15	2.07%	8.44%	F1_%CY
5	ARS-BFGL-NGS-101039	2	122509616	4.43E-05	4.35	0.17	0.01	0.62	0.15	2.17%	8.87%	F1_%CY
6	Hapmap60224-rs29001782	6	85178107	4.62E-05	4.34	0.22	0.01	0.62	0.15	2.23%	9.10%	F1_%CY
7	BTB-02092741	6	114223059	4.85E-05	4.31	0.71	0.01	0.62	0.15	2.23%	9.10%	F1_%CY
8	BTA-110429-no-rs	11	87670344	6.13E-07	6.21	0.21	0.02	0.69	0.15	3.20%	14.82%	F2_CFt
9	BTA-110431-no-rs	11	87692024	1.43E-06	5.84	0.21	0.02	0.69	0.15	2.94%	13.60%	F2_CFt
10	BTA-122637-no-rs	6	88442145	6.91E-06	5.16	-0.40	0.02	0.69	0.15	2.85%	13.16%	F2_CFt
11	Hapmap23226-BTA-159656	6	46599570	8.57E-06	5.07	0.22	0.02	0.69	0.15	2.60%	12.02%	F2_CFt
12	ARS-BFGL-NGS-105624	11	86711869	1.06E-05	4.97	-0.19	0.02	0.69	0.15	2.47%	11.40%	F2_CFt
13	ARS-BFGL-NGS-29506	11	87603011	1.07E-05	4.97	-0.19	0.02	0.69	0.15	2.52%	11.64%	F2_CFt
14	ARS-BFGL-NGS-116951	11	87793629	1.41E-05	4.85	0.19	0.02	0.69	0.15	2.41%	11.14%	F2_CFt
15	Hapmap30207-BTA-126940	11	88213919	1.50E-05	4.82	-0.19	0.02	0.69	0.15	2.43%	11.24%	F2_CFt
16	ARS-BFGL-NGS-119913	11	86779385	1.93E-05	4.71	-0.18	0.02	0.69	0.15	2.34%	10.80%	F2_CFt
17	ARS-BFGL-NGS-37074	11	88028793	2.03E-05	4.69	0.18	0.02	0.69	0.15	2.33%	10.76%	F2_CFt
18	ARS-BFGL-NGS-110235	11	85367073	2.47E-05	4.61	0.18	0.02	0.69	0.15	2.25%	10.40%	F2_CFt
19	BTA-120876-no-rs	11	85935590	2.59E-05	4.59	0.18	0.02	0.69	0.15	2.27%	10.48%	F2_CFt
20	ARS-BFGL-NGS-113560	11	88074583	2.64E-05	4.58	-0.18	0.02	0.69	0.15	2.31%	10.70%	F2_CFt
21	Hapmap46930-BTA-111382	11	88186796	3.73E-05	4.43	-0.18	0.02	0.69	0.15	2.22%	10.24%	F2_CFt
22	ARS-BFGL-NGS-104610	11	104293559	2.08E-26	25.68	-0.42	0.09	0.49	0.27	17.34%	31.33%	F4_Cheese N
23	ARS-BFGL-NGS-115328	11	103110855	6.88E-20	19.16	-0.37	0.06	0.49	0.27	12.60%	22.77%	F4_Cheese N
24	ARS-BFGL-NGS-119318	11	102974570	1.96E-15	14.71	-0.33	0.05	0.49	0.27	9.17%	16.56%	F4_Cheese N
25	ARS-BFGL-NGS-32953	11	101301047	9.02E-12	11.04	-0.27	0.03	0.49	0.27	6.71%	12.11%	F4_Cheese N
26	ARS-USMARC-Parent-AY851163-rs17871661	11	103047474	1.49E-09	8.83	-0.30	0.03	0.49	0.27	5.40%	9.76%	F4_Cheese N
27	ARS-BFGL-NGS-117078	11	104812200	3.64E-09	8.44	-0.30	0.03	0.49	0.27	5.16%	9.33%	F4_Cheese N
28	ARS-BFGL-NGS-40995	11	103742782	8.95E-08	7.05	-0.32	0.02	0.49	0.27	4.31%	7.78%	F4_Cheese N

29	BTA-118685-no-rs	11	101706587	1.66E-07	6.78	-0.33	0.02	0.49	0.27	4.11%	7.42%	F4_Cheese N
30	Hapmap56906-rs29014970	11	97844929	2.63E-07	6.58	-0.22	0.02	0.49	0.27	4.03%	7.28%	F4_Cheese N
31	ARS-BFGL-NGS-105689	11	102285689	4.60E-07	6.34	0.19	0.02	0.49	0.27	3.81%	6.89%	F4_Cheese N
32	ARS-BFGL-NGS-14392	11	102237822	1.27E-06	5.90	-0.26	0.02	0.49	0.27	3.35%	6.05%	F4_Cheese N
33	ARS-BFGL-NGS-116123	11	94686959	1.48E-06	5.83	-0.21	0.02	0.49	0.27	3.49%	6.31%	F4_Cheese N
34	ARS-BFGL-NGS-25833	11	106543262	2.08E-06	5.68	-0.19	0.02	0.49	0.27	3.25%	5.88%	F4_Cheese N
35	BTA-118663-no-rs	11	101384765	2.77E-06	5.56	0.18	0.02	0.49	0.27	3.12%	5.64%	F4_Cheese N
36	ARS-BFGL-NGS-45339	11	106067128	2.96E-06	5.53	-0.22	0.02	0.49	0.27	3.06%	5.54%	F4_Cheese N
37	BTA-118661-no-rs	11	101357473	3.45E-06	5.46	0.18	0.02	0.49	0.27	3.06%	5.52%	F4_Cheese N
38	ARS-BFGL-NGS-111682	11	104633267	4.00E-06	5.40	-0.22	0.02	0.49	0.27	3.16%	5.72%	F4_Cheese N
39	ARS-BFGL-NGS-24522	6	87878364	4.40E-06	5.36	0.31	0.01	0.49	0.27	2.94%	5.32%	F4_Cheese N
40	ARS-BFGL-NGS-11064	11	106510932	8.45E-06	5.07	-0.18	0.01	0.49	0.27	2.83%	5.11%	F4_Cheese N
41	ARS-BFGL-NGS-77843	11	103856100	1.17E-05	4.93	-0.20	0.01	0.49	0.27	2.72%	4.91%	F4_Cheese N
42	ARS-BFGL-NGS-101698	11	96230130	1.22E-05	4.91	-0.20	0.01	0.49	0.27	2.77%	5.01%	F4_Cheese N
43	ARS-BFGL-NGS-2573	11	101602103	2.94E-05	4.53	-0.38	0.01	0.49	0.27	2.50%	4.51%	F4_Cheese N
44	ARS-BFGL-NGS-98548	11	103458444	3.04E-05	4.52	0.22	0.01	0.49	0.27	2.52%	4.54%	F4_Cheese N
45	ARS-BFGL-NGS-35656	11	97289960	4.13E-05	4.38	0.20	0.01	0.49	0.27	2.40%	4.34%	F4_Cheese N
46	Hapmap24798-BTA-127049	11	95837993	4.51E-05	4.35	-0.17	0.01	0.49	0.27	2.51%	4.53%	F4_Cheese N
47	Hapmap28023-BTC-060518	6	87201599	2.84E-47	46.55	-0.90	0.23	0.65	0.44	35.60%	52.82%	F5_αs1-β-CN
48	Hapmap24184-BTC-070077	6	87245049	7.00E-45	44.15	-0.88	0.22	0.65	0.44	33.76%	50.09%	F5_αs1-β-CN
49	ARS-BFGL-NGS-36707	6	86354888	1.22E-26	25.92	-0.75	0.13	0.65	0.44	19.95%	29.60%	F5_αs1-β-CN
50	Hapmap41098-BTA-86027	6	84889974	7.27E-24	23.14	-0.70	0.11	0.65	0.44	17.35%	25.74%	F5_αs1-β-CN
51	Hapmap46932-BTA-111719	6	84819700	1.99E-22	21.70	-0.57	0.11	0.65	0.44	16.78%	24.89%	F5_αs1-β-CN
52	BTA-111108-no-rs	6	85424500	1.88E-21	20.73	-0.66	0.10	0.65	0.44	15.68%	23.26%	F5_αs1-β-CN
53	ARS-BFGL-NGS-114609	6	84713584	6.87E-18	17.16	-0.49	0.09	0.65	0.44	13.07%	19.39%	F5_αs1-β-CN
54	Hapmap25708-BTC-043671	6	87113639	6.54E-15	14.18	-0.41	0.07	0.65	0.44	10.67%	15.83%	F5_αs1-β-CN
55	ARS-BFGL-NGS-70112	6	84448550	7.31E-13	12.14	-0.41	0.06	0.65	0.44	8.53%	12.65%	F5_αs1-β-CN
56	Hapmap23387-BTC-072905	6	82078166	1.26E-12	11.90	-0.48	0.06	0.65	0.44	8.54%	12.68%	F5_αs1-β-CN
57	Hapmap32475-BTC-050530	6	82047313	1.26E-12	11.90	-0.48	0.06	0.65	0.44	8.54%	12.68%	F5_αs1-β-CN
58	BTA-121763-no-rs	6	84968944	2.49E-11	10.60	0.39	0.05	0.65	0.44	7.38%	10.95%	F5_αs1-β-CN
59	ARS-BFGL-NGS-63312	6	82965163	1.01E-10	10.00	-0.38	0.04	0.65	0.44	6.84%	10.14%	F5_αs1-β-CN

60	Hapmap49297-BTA-76961	6	83147633	1.62E-10	9.79	-0.34	0.04	0.65	0.44	6.70%	9.94%	F5_αs1-β-CN
61	Hapmap26317-BTC-059618	6	82253271	9.19E-10	9.04	-0.31	0.04	0.65	0.44	6.28%	9.32%	F5_αs1-β-CN
62	Hapmap23419-BTC-059652	6	82314634	1.41E-09	8.85	-0.31	0.04	0.65	0.44	6.18%	9.17%	F5_αs1-β-CN
63	Hapmap31932-BTC-042947	6	82350917	3.94E-09	8.40	-0.30	0.04	0.65	0.44	5.82%	8.63%	F5_αs1-β-CN
64	BTA-121766-no-rs	6	83256899	1.83E-08	7.74	-0.28	0.03	0.65	0.44	5.28%	7.83%	F5_αs1-β-CN
65	ARS-BFGL-NGS-118182	6	88592295	2.46E-08	7.61	0.31	0.03	0.65	0.44	5.13%	7.61%	F5_αs1-β-CN
66	BTB-00264815	6	81019581	3.56E-08	7.45	-0.33	0.03	0.65	0.44	5.04%	7.48%	F5_αs1-β-CN
67	BTB-01393607	6	80062406	8.02E-08	7.10	-0.32	0.03	0.65	0.44	4.88%	7.25%	F5_αs1-β-CN
68	Hapmap30192-BTC-072881	6	87281196	1.31E-07	6.88	0.54	0.03	0.65	0.44	4.88%	7.24%	F5_αs1-β-CN
69	Hapmap27307-BTC-043200	6	82605943	1.35E-07	6.87	0.31	0.03	0.65	0.44	4.48%	6.65%	F5_αs1-β-CN
70	Hapmap23522-BTC-047379	6	87222751	2.53E-07	6.60	0.53	0.03	0.65	0.44	4.64%	6.88%	F5_αs1-β-CN
71	ARS-BFGL-NGS-60491	6	85703164	1.08E-06	5.97	-0.24	0.03	0.65	0.44	4.14%	6.14%	F5_αs1-β-CN
72	BTA-76907-no-rs	0	0	1.14E-06	5.94	-0.26	0.02	0.65	0.44	3.81%	5.65%	F5_αs1-β-CN
73	ARS-BFGL-NGS-24522	6	87878364	1.16E-06	5.93	-0.41	0.02	0.65	0.44	3.77%	5.60%	F5_αs1-β-CN
74	Hapmap31617-BTC-043368	6	82458774	1.51E-06	5.82	0.31	0.02	0.65	0.44	3.77%	5.60%	F5_αs1-β-CN
75	BTB-01654826	6	88891318	2.81E-06	5.55	0.29	0.02	0.65	0.44	3.66%	5.43%	F5_αs1-β-CN
76	ARS-BFGL-NGS-42175	6	79544981	2.85E-06	5.55	-0.27	0.02	0.65	0.44	3.67%	5.45%	F5_αs1-β-CN
77	BTA-77077-no-rs	6	85527109	3.28E-06	5.48	-0.82	0.02	0.65	0.44	3.58%	5.31%	F5_αs1-β-CN
78	ARS-BFGL-NGS-110734	0	0	3.37E-06	5.47	-0.23	0.02	0.65	0.44	3.52%	5.22%	F5_αs1-β-CN
79	Hapmap43767-BTA-113302	6	85646902	3.73E-06	5.43	-0.23	0.02	0.65	0.44	3.76%	5.58%	F5_αs1-β-CN
80	Hapmap32099-BTA-151095	6	83345994	4.00E-06	5.40	-0.23	0.02	0.65	0.44	3.46%	5.13%	F5_αs1-β-CN
81	BTA-76959-no-rs	6	83290843	4.35E-06	5.36	-0.23	0.02	0.65	0.44	3.43%	5.09%	F5_αs1-β-CN
82	Hapmap60030-rs29013992	6	77585276	4.85E-06	5.31	-0.29	0.02	0.65	0.44	3.38%	5.01%	F5_αs1-β-CN
83	BTA-114800-no-rs	6	77520815	7.11E-06	5.15	-0.28	0.02	0.65	0.44	3.29%	4.89%	F5_αs1-β-CN
84	ARS-BFGL-NGS-80068	6	77650126	7.58E-06	5.12	-0.28	0.02	0.65	0.44	3.26%	4.84%	F5_αs1-β-CN
85	Hapmap27109-BTC-060711	6	87152621	1.44E-05	4.84	-0.22	0.02	0.65	0.44	3.12%	4.63%	F5_αs1-β-CN
86	Hapmap33631-BTC-043555	6	87327708	1.58E-05	4.80	-0.22	0.02	0.65	0.44	3.33%	4.94%	F5_αs1-β-CN
87	Hapmap51938-BTA-21491	6	81057816	2.34E-05	4.63	-0.24	0.02	0.65	0.44	3.05%	4.52%	F5_αs1-β-CN
88	ARS-BFGL-NGS-2391	6	91187144	3.23E-05	4.49	0.44	0.02	0.65	0.44	2.93%	4.35%	F5_αs1-β-CN
89	BTA-21753-no-rs	9	36790663	4.24E-05	4.37	-1.02	0.02	0.65	0.44	2.67%	3.97%	F5_αs1-β-CN
90	ARS-BFGL-NGS-37271	6	85743833	4.29E-05	4.37	0.38	0.02	0.65	0.44	2.85%	4.23%	F5_αs1-β-CN

91	BTB-00264414	6	79843283	4.49E-05	4.35	-0.20	0.02	0.65	0.44	2.70%	4.00%	F5_αs1-β-CN
92	BTA-22106-no-rs	6	89104201	4.61E-05	4.34	0.41	0.02	0.65	0.44	2.92%	4.33%	F5_αs1-β-CN
93	Hapmap26618-BTC-070864	6	39597740	1.19E-06	5.93	0.47	0.01	0.46	0.07	3.03%	20.10%	F6_Udder health
94	Hapmap27849-BTC-071108	6	39503443	1.72E-06	5.76	0.47	0.01	0.46	0.07	2.95%	19.55%	F6_Udder health
95	Hapmap31994-BTC-065943	25	5385729	3.12E-06	5.51	0.23	0.01	0.46	0.07	2.71%	17.97%	F6_Udder health
96	Hapmap52348-rs29024684	6	87396306	5.84E-06	5.23	0.18	0.01	0.46	0.07	2.68%	17.77%	F6_Udder health
97	Hapmap28075-BTC-035688	6	40377515	7.89E-06	5.10	0.40	0.01	0.46	0.07	2.61%	17.32%	F6_Udder health
98	BTB-01723556	11	4419032	4.72E-05	4.33	0.29	0.01	0.46	0.07	2.14%	14.21%	F6_Udder health
99	Hapmap52348-rs29024684	6	87396306	9.81E-56	55.01	-1.01	0.37	0.77	0.50	48.03%	74.19%	F7_κ-β-CN
100	BTA-111108-no-rs	6	85424500	4.65E-22	21.33	-0.78	0.14	0.77	0.50	18.34%	28.34%	F7_κ-β-CN
101	ARS-BFGL-NGS-114609	6	84713584	3.71E-21	20.43	-0.62	0.14	0.77	0.50	17.62%	27.23%	F7_κ-β-CN
102	ARS-BFGL-NGS-63312	6	82965163	9.19E-21	20.04	-0.64	0.13	0.77	0.50	16.24%	25.09%	F7_κ-β-CN
103	Hapmap41098-BTA-86027	6	84889974	3.55E-19	18.45	-0.72	0.12	0.77	0.50	15.61%	24.12%	F7_κ-β-CN
104	ARS-BFGL-NGS-36707	6	86354888	1.11E-18	17.96	-0.72	0.12	0.77	0.50	15.50%	23.95%	F7_κ-β-CN
105	Hapmap23387-BTC-072905	6	82078166	1.33E-17	16.88	-0.66	0.11	0.77	0.50	14.07%	21.74%	F7_κ-β-CN
106	Hapmap32475-BTC-050530	6	82047313	1.33E-17	16.88	-0.66	0.11	0.77	0.50	14.07%	21.74%	F7_κ-β-CN
107	Hapmap52479-rs29018853	6	79203343	5.79E-17	16.24	-0.54	0.10	0.77	0.50	13.52%	20.89%	F7_κ-β-CN
108	BTA-76907-no-rs	0	0	9.47E-17	16.02	-0.51	0.10	0.77	0.50	12.63%	19.51%	F7_κ-β-CN
109	Hapmap28023-BTC-060518	6	87201599	1.13E-16	15.95	-0.60	0.10	0.77	0.50	13.34%	20.61%	F7_κ-β-CN
110	Hapmap24184-BTC-070077	6	87245049	2.41E-16	15.62	-0.60	0.10	0.77	0.50	13.17%	20.34%	F7_κ-β-CN
111	Hapmap31932-BTC-042947	6	82350917	2.85E-16	15.55	-0.49	0.10	0.77	0.50	12.78%	19.74%	F7_κ-β-CN
112	ARS-BFGL-NGS-42175	6	79544981	5.02E-16	15.30	-0.55	0.10	0.77	0.50	12.54%	19.37%	F7_κ-β-CN
113	Hapmap26317-BTC-059618	6	82253271	2.99E-15	14.52	-0.47	0.09	0.77	0.50	11.87%	18.34%	F7_κ-β-CN
114	Hapmap23419-BTC-059652	6	82314634	3.66E-15	14.44	-0.47	0.09	0.77	0.50	11.87%	18.34%	F7_κ-β-CN
115	Hapmap46932-BTA-111719	6	84819700	4.09E-15	14.39	-0.53	0.09	0.77	0.50	12.11%	18.70%	F7_κ-β-CN
116	Hapmap32099-BTA-151095	6	83345994	6.21E-15	14.21	-0.45	0.09	0.77	0.50	11.25%	17.38%	F7_κ-β-CN
117	BTA-76959-no-rs	6	83290843	6.44E-15	14.19	-0.45	0.09	0.77	0.50	11.25%	17.38%	F7_κ-β-CN
118	Hapmap51938-BTA-21491	6	81057816	9.67E-15	14.01	-0.50	0.09	0.77	0.50	11.63%	17.97%	F7_κ-β-CN
119	ARS-BFGL-NGS-110734	0	0	1.11E-14	13.96	-0.45	0.09	0.77	0.50	11.09%	17.14%	F7_κ-β-CN
120	BTB-00264414	6	79843283	1.85E-14	13.73	-0.44	0.08	0.77	0.50	10.82%	16.72%	F7_κ-β-CN
121	Hapmap33631-BTC-043555	6	87327708	6.89E-14	13.16	-0.44	0.09	0.77	0.50	11.43%	17.65%	F7_κ-β-CN

122	BTB-00264815	6	81019581	5.41E-13	12.27	-0.50	0.08	0.77	0.50	9.82%	15.17%	F7_κ-β-CN
123	BTB-01393607	6	80062406	7.92E-13	12.10	-0.50	0.08	0.77	0.50	9.89%	15.28%	F7_κ-β-CN
124	Hapmap27109-BTC-060711	6	87152621	2.08E-12	11.68	-0.41	0.07	0.77	0.50	9.33%	14.41%	F7_κ-β-CN
125	Hapmap43767-BTA-113302	6	85646902	3.39E-12	11.47	-0.40	0.07	0.77	0.50	9.69%	14.97%	F7_κ-β-CN
126	Hapmap49297-BTA-76961	6	83147633	6.42E-12	11.19	-0.42	0.07	0.77	0.50	8.79%	13.58%	F7_κ-β-CN
127	ARS-BFGL-NGS-27643	6	78786848	1.00E-11	11.00	-0.49	0.07	0.77	0.50	8.91%	13.77%	F7_κ-β-CN
128	BTB-01900612	6	79817258	2.44E-11	10.61	-0.37	0.06	0.77	0.50	8.10%	12.51%	F7_κ-β-CN
129	ARS-BFGL-NGS-70112	6	84448550	2.49E-11	10.60	-0.43	0.06	0.77	0.50	8.28%	12.78%	F7_κ-β-CN
130	ARS-BFGL-NGS-60491	6	85703164	6.00E-11	10.22	-0.37	0.07	0.77	0.50	8.48%	13.10%	F7_κ-β-CN
131	BTB-00264506	6	79732129	8.30E-11	10.08	-0.36	0.06	0.77	0.50	7.91%	12.21%	F7_κ-β-CN
132	BTB-00265288	6	81286240	1.04E-10	9.98	-0.37	0.06	0.77	0.50	7.71%	11.91%	F7_κ-β-CN
133	BTA-77011-no-rs	6	82773692	1.10E-10	9.96	-0.36	0.06	0.77	0.50	7.98%	12.32%	F7_κ-β-CN
134	Hapmap26729-BTA-154447	6	77391041	1.53E-10	9.82	-0.37	0.06	0.77	0.50	7.86%	12.15%	F7_κ-β-CN
135	BTA-122637-no-rs	6	88442145	2.46E-10	9.61	-0.71	0.06	0.77	0.50	7.96%	12.30%	F7_κ-β-CN
136	Hapmap38629-BTA-76891	6	78289101	3.35E-10	9.47	-0.35	0.06	0.77	0.50	7.34%	11.33%	F7_κ-β-CN
137	BTA-114800-no-rs	6	77520815	4.15E-10	9.38	-0.46	0.06	0.77	0.50	7.25%	11.20%	F7_κ-β-CN
138	BTB-00217058	6	79592589	4.81E-10	9.32	-0.34	0.05	0.77	0.50	7.05%	10.88%	F7_κ-β-CN
139	ARS-BFGL-NGS-80068	6	77650126	4.99E-10	9.30	-0.45	0.06	0.77	0.50	7.16%	11.07%	F7_κ-β-CN
140	BTA-05242-no-rs	6	81225864	5.50E-10	9.26	-0.34	0.05	0.77	0.50	6.98%	10.78%	F7_κ-β-CN
141	Hapmap60030-rs29013992	6	77585276	1.82E-09	8.74	-0.44	0.05	0.77	0.50	6.65%	10.27%	F7_κ-β-CN
142	ARS-BFGL-NGS-112473	6	79680793	8.91E-09	8.05	-0.32	0.05	0.77	0.50	5.96%	9.21%	F7_κ-β-CN
143	Hapmap39314-BTA-121985	6	79982917	9.96E-09	8.00	-0.32	0.05	0.77	0.50	6.28%	9.70%	F7_κ-β-CN
144	BTA-76988-no-rs	6	90876389	1.65E-08	7.78	-0.86	0.04	0.77	0.50	5.82%	8.99%	F7_κ-β-CN
145	Hapmap43045-BTA-76998	6	90730485	1.73E-08	7.76	-1.31	0.05	0.77	0.50	6.10%	9.43%	F7_κ-β-CN
146	BTB-00264877	6	81042351	1.84E-08	7.73	-0.31	0.04	0.77	0.50	5.80%	8.96%	F7_κ-β-CN
147	BTB-00264606	6	79632703	1.90E-08	7.72	-0.31	0.04	0.77	0.50	5.78%	8.92%	F7_κ-β-CN
148	Hapmap25708-BTC-043671	6	87113639	2.07E-08	7.68	-0.34	0.05	0.77	0.50	6.28%	9.70%	F7_κ-β-CN
149	BTB-01885835	6	92279760	3.00E-08	7.52	-0.62	0.05	0.77	0.50	5.89%	9.10%	F7_κ-β-CN
150	Hapmap25412-BTC-031860	6	82008221	5.05E-08	7.30	-0.31	0.04	0.77	0.50	5.45%	8.41%	F7_κ-β-CN
151	Hapmap47405-BTA-76965	6	83117772	8.60E-08	7.07	-0.76	0.04	0.77	0.50	5.35%	8.26%	F7_κ-β-CN
152	ARS-BFGL-NGS-12798	6	78561987	2.49E-07	6.60	-0.72	0.04	0.77	0.50	4.74%	7.31%	F7_κ-β-CN

153	ARS-BFGL-NGS-60568	6	92465869	3.50E-07	6.46	-0.41	0.04	0.77	0.50	4.96%	7.67%	F7_κ-β-CN
154	Hapmap60182-rs29025531	6	74606760	5.95E-07	6.23	-0.34	0.04	0.77	0.50	4.79%	7.40%	F7_κ-β-CN
155	BTA-121695-no-rs	6	92552055	8.32E-07	6.08	-0.40	0.04	0.77	0.50	4.64%	7.16%	F7_κ-β-CN
156	Hapmap35196-BES10_Contig207_566	6	92578562	8.32E-07	6.08	-0.40	0.04	0.77	0.50	4.64%	7.16%	F7_κ-β-CN
157	Hapmap50778-BTA-77229	6	94334151	9.27E-07	6.03	-0.91	0.03	0.77	0.50	4.36%	6.73%	F7_κ-β-CN
158	Hapmap57014-rs29019575	6	87801255	1.03E-06	5.99	-0.73	0.03	0.77	0.50	4.52%	6.98%	F7_κ-β-CN
159	BTA-121766-no-rs	6	83256899	1.36E-06	5.87	-0.28	0.03	0.77	0.50	4.43%	6.84%	F7_κ-β-CN
160	BTB-01687386	6	94360125	1.79E-06	5.75	-0.87	0.03	0.77	0.50	4.14%	6.39%	F7_κ-β-CN
161	BTB-01995503	6	92527885	2.15E-06	5.67	-0.39	0.03	0.77	0.50	4.36%	6.74%	F7_κ-β-CN
162	BTA-10392-no-rs	6	88547858	2.69E-06	5.57	-0.86	0.03	0.77	0.50	4.49%	6.94%	F7_κ-β-CN
163	ARS-BFGL-NGS-99026	6	72930338	2.98E-06	5.53	-0.29	0.03	0.77	0.50	3.87%	5.98%	F7_κ-β-CN
164	Hapmap49058-BTA-121772	6	94228599	3.77E-06	5.42	-0.68	0.03	0.77	0.50	4.07%	6.28%	F7_κ-β-CN
165	Hapmap44699-BTA-122320	6	77228984	5.73E-06	5.24	-0.28	0.03	0.77	0.50	3.86%	5.96%	F7_κ-β-CN
166	BTA-111809-no-rs	6	77186116	6.47E-06	5.19	-0.28	0.03	0.77	0.50	3.86%	5.96%	F7_κ-β-CN
167	ARS-BFGL-NGS-68326	6	72906612	8.34E-06	5.08	-0.28	0.03	0.77	0.50	3.54%	5.47%	F7_κ-β-CN
168	Hapmap29639-BTC-041962	6	71350048	1.21E-05	4.92	-0.88	0.03	0.77	0.50	3.69%	5.70%	F7_κ-β-CN
169	UA-IFASA-2111	6	84351311	1.63E-05	4.79	-0.78	0.03	0.77	0.50	3.36%	5.19%	F7_κ-β-CN
170	BTB-00262270	6	74546008	2.11E-05	4.67	-0.26	0.03	0.77	0.50	3.39%	5.24%	F7_κ-β-CN
171	BTB-00263209	6	72382208	2.87E-05	4.54	-0.28	0.02	0.77	0.50	3.21%	4.97%	F7_κ-β-CN
172	BTB-00041036	1	90156001	4.19E-05	4.38	-1.37	0.03	0.77	0.50	3.57%	5.51%	F7_κ-β-CN
173	Hapmap28023-BTC-060518	6	87201599	8.34E-11	10.08	-0.35	0.04	0.55	0.22	6.37%	16.01%	F8_αs2-CN
174	Hapmap24184-BTC-070077	6	87245049	1.67E-10	9.78	-0.35	0.03	0.55	0.22	6.24%	15.70%	F8_αs2-CN
175	Hapmap43045-BTA-76998	6	90730485	1.63E-09	8.79	1.05	0.03	0.55	0.22	5.43%	13.66%	F8_αs2-CN
176	UA-IFASA-2111	6	84351311	2.02E-09	8.69	0.81	0.03	0.55	0.22	5.06%	12.71%	F8_αs2-CN
177	BTB-01687386	6	94360125	1.36E-08	7.87	0.77	0.03	0.55	0.22	4.53%	11.38%	F8_αs2-CN
178	Hapmap50778-BTA-77229	6	94334151	1.77E-08	7.75	0.78	0.02	0.55	0.22	4.47%	11.23%	F8_αs2-CN
179	ARS-BFGL-NGS-36707	6	86354888	2.17E-08	7.66	-0.34	0.03	0.55	0.22	4.84%	12.17%	F8_αs2-CN
180	ARS-BFGL-NGS-70112	6	84448550	5.04E-08	7.30	-0.27	0.02	0.55	0.22	4.34%	10.91%	F8_αs2-CN
181	BTA-68275-no-rs	6	89001383	5.73E-08	7.24	0.52	0.02	0.55	0.22	4.16%	10.47%	F8_αs2-CN
182	Hapmap33631-BTC-043555	6	87327708	9.47E-08	7.02	-0.23	0.02	0.55	0.22	4.47%	11.25%	F8_αs2-CN
183	Hapmap29639-BTC-041962	6	71350048	1.12E-07	6.95	0.80	0.02	0.55	0.22	4.22%	10.60%	F8_αs2-CN

184	Hapmap46932-BTA-111719	6	84819700	1.13E-07	6.95	-0.27	0.02	0.55	0.22	4.46%	11.20%	F8_αs2-CN
185	Hapmap57014-rs29019575	6	87801255	1.43E-07	6.85	0.59	0.02	0.55	0.22	4.08%	10.25%	F8_αs2-CN
186	BTA-121766-no-rs	6	83256899	2.15E-07	6.67	-0.22	0.02	0.55	0.22	3.99%	10.03%	F8_αs2-CN
187	Hapmap25708-BTC-043671	6	87113639	3.59E-07	6.44	-0.23	0.02	0.55	0.22	4.03%	10.12%	F8_αs2-CN
188	Hapmap41098-BTA-86027	6	84889974	4.94E-07	6.31	-0.30	0.02	0.55	0.22	3.84%	9.65%	F8_αs2-CN
189	BTA-76988-no-rs	6	90876389	8.03E-07	6.10	0.56	0.02	0.55	0.22	3.47%	8.71%	F8_αs2-CN
190	Hapmap49297-BTA-76961	6	83147633	1.67E-06	5.78	-0.22	0.02	0.55	0.22	3.33%	8.37%	F8_αs2-CN
191	BTA-111108-no-rs	6	85424500	2.31E-06	5.64	-0.29	0.02	0.55	0.22	3.42%	8.60%	F8_αs2-CN
192	BTB-00264815	6	81019581	2.94E-06	5.53	-0.24	0.02	0.55	0.22	3.21%	8.06%	F8_αs2-CN
193	Hapmap44699-BTA-122320	6	77228984	3.26E-06	5.49	-0.22	0.02	0.55	0.22	3.16%	7.95%	F8_αs2-CN
194	BTB-00264565	6	79705920	3.72E-06	5.43	0.65	0.02	0.55	0.22	3.17%	7.98%	F8_αs2-CN
195	ARS-BFGL-NGS-60491	6	85703164	4.23E-06	5.37	-0.20	0.02	0.55	0.22	3.28%	8.25%	F8_αs2-CN
196	BTB-01393607	6	80062406	4.64E-06	5.33	-0.24	0.02	0.55	0.22	3.15%	7.91%	F8_αs2-CN
197	Hapmap25412-BTC-031860	6	82008221	6.32E-06	5.20	-0.20	0.02	0.55	0.22	2.98%	7.49%	F8_αs2-CN
198	Hapmap47405-BTA-76965	6	83117772	8.14E-06	5.09	0.47	0.02	0.55	0.22	2.89%	7.26%	F8_αs2-CN
199	ARS-BFGL-NGS-12798	6	78561987	1.14E-05	4.94	0.46	0.01	0.55	0.22	2.68%	6.73%	F8_αs2-CN
200	BTA-76623-no-rs	6	71154473	1.26E-05	4.90	0.51	0.02	0.55	0.22	2.91%	7.31%	F8_αs2-CN
201	Hapmap23387-BTC-072905	6	82078166	2.16E-05	4.66	-0.25	0.01	0.55	0.22	2.71%	6.80%	F8_αs2-CN
202	Hapmap32475-BTC-050530	6	82047313	2.16E-05	4.66	-0.25	0.01	0.55	0.22	2.71%	6.80%	F8_αs2-CN
203	ARS-BFGL-NGS-111636	6	68546212	2.18E-05	4.66	0.39	0.01	0.55	0.22	2.57%	6.47%	F8_αs2-CN
204	Hapmap60030-rs29013992	6	77585276	2.45E-05	4.61	-0.23	0.01	0.55	0.22	2.56%	6.43%	F8_αs2-CN
205	Hapmap41952-BTA-73370	10	10659761	3.22E-05	4.49	0.17	0.01	0.55	0.22	2.43%	6.11%	F8_αs2-CN
206	Hapmap51938-BTA-21491	6	81057816	4.11E-05	4.39	-0.20	0.01	0.55	0.22	2.54%	6.40%	F8_αs2-CN
207	Hapmap28023-BTC-060518	6	87201599	3.86E-11	10.41	-0.32	0.03	0.48	0.12	6.06%	24.69%	F9_αs1-CN-Ph
208	Hapmap24184-BTC-070077	6	87245049	7.80E-11	10.11	-0.32	0.03	0.48	0.12	5.94%	24.20%	F9_αs1-CN-Ph
209	ARS-BFGL-NGS-36707	6	86354888	1.96E-07	6.71	-0.28	0.02	0.48	0.12	3.84%	15.65%	F9_αs1-CN-Ph
210	Hapmap25708-BTC-043671	6	87113639	4.37E-07	6.36	-0.21	0.02	0.48	0.12	3.65%	14.85%	F9_αs1-CN-Ph
211	ARS-BFGL-NGS-114609	6	84713584	1.70E-06	5.77	-0.21	0.02	0.48	0.12	3.27%	13.31%	F9_αs1-CN-Ph
212	BTA-111108-no-rs	6	85424500	2.73E-06	5.56	-0.25	0.01	0.48	0.12	3.10%	12.64%	F9_αs1-CN-Ph
213	Hapmap41098-BTA-86027	6	84889974	7.81E-06	5.11	-0.24	0.01	0.48	0.12	2.79%	11.36%	F9_αs1-CN-Ph
214	Hapmap46932-BTA-111719	6	84819700	1.52E-05	4.82	-0.20	0.01	0.48	0.12	2.73%	11.11%	F9_αs1-CN-Ph

215	Hapmap51592-BTA-41521	20	46709345	1.84E-05	4.74	-0.16	0.01	0.48	0.12	2.45%	9.96%	F9_αs1-CN-Ph
216	ARS-BFGL-NGS-70112	6	84448550	2.15E-05	4.67	-0.19	0.01	0.48	0.12	2.44%	9.93%	F9_αs1-CN-Ph
217	ARS-BFGL-NGS-56195	11	77493775	3.40E-05	4.47	-0.38	0.01	0.48	0.12	2.42%	9.87%	F9_αs1-CN-Ph
218	ARS-BFGL-NGS-110734	0	0	4.66E-05	4.33	-0.16	0.01	0.48	0.12	2.22%	9.02%	F9_αs1-CN-Ph
219	Hapmap32099-BTA-151095	6	83345994	4.86E-05	4.31	-0.16	0.01	0.48	0.12	2.18%	8.89%	F9_αs1-CN-Ph
220	BTA-76959-no-rs	6	83290843	4.89E-05	4.31	-0.16	0.01	0.48	0.12	2.18%	8.89%	F9_αs1-CN-Ph
221	BTA-51080-no-rs	20	7881875	8.02E-06	5.10	-1.01	0.01	0.47	0.08	2.79%	15.90%	F10_α-LA
222	ARS-BFGL-NGS-6104	11	104456040	1.85E-05	4.73	-0.27	0.01	0.47	0.08	2.46%	14.01%	F10_α-LA
223	ARS-BFGL-NGS-4101	11	104325453	2.67E-05	4.57	-0.26	0.01	0.47	0.08	2.33%	13.29%	F10_α-LA
224	ARS-BFGL-NGS-24170	27	43459156	2.84E-05	4.55	0.14	0.01	0.47	0.08	2.17%	12.34%	F10_α-LA
225	BTB-01261619	27	43901931	4.73E-05	4.32	0.15	0.01	0.47	0.08	2.20%	12.54%	F10_α-LA
226	Hapmap41400-BTA-101218	27	43435991	4.84E-05	4.32	0.14	0.01	0.47	0.08	2.03%	11.59%	F10_α-LA

Description: SNP= the name of the single nucleotide polymorphism; BTA= *Bos taurus* autosome chromosome; Location= position of the SNP on the chromosome in base pairs on UMD3.1 (http://www.ensembl.org/index.html); Pc1df= *P*-values adjusted for genomic control; LOG= the $-\log_{10}$ of Pc1df; effB= effect of the minor allele (B allele); V_{SNP} = variance explained by the SNP (calculated as $2pqa^2$, where p is the frequency of one allele, q=1-p is the frequency of the second allele and a is the additive genetic effect); V_P= phenotypic variance; V_G= additive genetic variance; V_{Psnp} (%)= percentage of phenotypic variance explained by each SNP; V_{Gsnp} (%)= percentage of additive genetic variance explained by each SNP; Trait= name of the latent variable. $F1_{\%CY}$ = 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; $F5_{as1\text{-}\beta\text{-}CN}$ = Factor related to the as_1- 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; $F8_{as2\text{-}CN}$ = Factor related to the milk as_2-CN, expressed as relative content to the total milk nitrogen; $F9_{as1\text{-}CN\text{-}Ph}$ = Factor related to the milk as_1-CN-Ph expressed as content to the total milk nitrogen; $F10_{\alpha\text{-}LA}$ = Factor related to the milk α-LA.

Table S2. Number of significant[1] SNP identified from GWAS and number of genes mapped by trait.

Trait[2]	No. of significant SNP	No. of significant SNP assigned to genes	No. of significant mapped Genes[3]
$F1_{\%CY}$	1,849	626	550
$F2_{CFt}$	1,756	588	505
$F3_{Yield}$	1,845	620	526
$F4_{Cheese\ N}$	1,504	550	463
$F5_{as1\text{-}\beta\text{-}CN}$	995	314	267
$F6_{Udder\ health}$	1,853	612	528
$F7_{\kappa\text{-}\beta\text{-}CN}$	622	183	164
$F8_{as2\text{-}CN}$	1,504	543	455
$F9_{as1\text{-}CN\text{-}Ph}$	1,713	589	527
$F10_{\alpha\text{-}LA}$	1,859	664	555
Background[4]	37,568	10,094	13,269

[1] P-value<0.05

[2] $F1_{\%CY}$ = Factor underlying the percentage of individual cheese yield; $F2_{CFt}$ = Factor underlying the milk curd firmness; $F3_{Yield}$ = Factor underlying the milk yield; $F4_{Cheese\ N}$ = Factor underlying the protein in the cheese; $F5_{as1\text{-}\beta\text{-}CN}$ = Factor underlying the αs_1 and β caseins; $F6_{Udder\ health}$ = Factor underlying the udder health condition of a cow; $F7_{\kappa\text{-}\beta\text{-}CN}$ = Factor underlying the κ and β caseins; $F8_{as2\text{-}CN}$ = Factor underlying the αs_2-casein; $F9_{as1\text{-}CN\text{-}Ph}$ = Factor underlying the phosphorylated αs_1-casein; $F10_{\alpha\text{-}LA}$ = Factor underlying the α-LA.

[3]Ensembl *Bos taurus*UMD3.1 (http://www.ensembl.org/index.html); window: 15kb

[4]Background represents the total number of SNP used in the GWAS analyses and the genes mapped to those SNPs.

Table S3 Gene ontology (GO) terms and Kyoto encyclopedia of genes and genomes (KEGG) pathways significantly enriched using genes associated with $F1_{\%CY}$, $F4_{Cheese\ N}$, $F8_{as_2-CN}$ and $F10_{a-LA}$.

Trait[1]	Category[2]	Term	No. genes in the term	No. sig genes[3]	FDR[4]
$F1_{\%CY}$	KEGG	bta04530: Tight junction	83	12	1.50×10^{-4}
$F10_{a-LA}$				13	3.74×10^{-5}
$F4_{Cheese\ N}$	KEGG	bta05412: Arrhythmogenic right ventricular cardiomyopathy (ARVC)	50	9	4.97×10^{-5}
$F8_{as_2-CN}$	GO_BP	GO:0055076~transition metal ion homeostasis	27	7	2.60×10^{-5}
		GO:0051716~cellular response to stimulus	918	55	3.36×10^{-5}
		GO:0030001~metal ion transport	132	15	4.61×10^{-5}
		GO:0006811~ion transport	269	23	5.37×10^{-5}
		GO:0034220~ion transmembrane transport	178	17	1.34×10^{-4}
		GO:0023052~signaling	750	45	1.76×10^{-4}
		GO:0044700~single organism signaling	750	45	1.76×10^{-4}
		GO:0007165~signal transduction	712	43	2.14×10^{-4}
		GO:0007154~cell communication	785	46	2.55×10^{-4}
		GO:0098660~inorganic ion transmembrane transport	137	14	2.56×10^{-4}
		GO:0003008~system process	226	19	2.93×10^{-4}
		GO:0055065~metal ion homeostasis	83	10	5.27×10^{-4}
	GO_CC	GO:0097458~neuron part	162	21	1.59×10^{-7}
		GO:0044456~synapse part	76	13	1.63×10^{-6}
		GO:0098794~postsynapse	48	10	4.13×10^{-6}
		GO:0043005~neuron projection	110	15	5.01×10^{-6}
		GO:0045202~synapse	102	14	9.57×10^{-6}
		GO:0097060~synaptic membrane	39	8	4.34×10^{-5}
		GO:0030424~axon	50	9	4.34×10^{-5}
		GO:0042995~cell projection	269	22	1.49×10^{-4}
		GO:0098590~plasma membrane region	146	14	4.93×10^{-4}
		GO:0098589~membrane region	181	16	5.01×10^{-4}
		GO:0044463~cell projection part	114	12	5.32×10^{-4}

GP_MF	GO:0022836~gated channel activity	59	9	1.66×10^{-4}
	GO:0046873~metal ion transmembrane transporter activity	73	10	1.83×10^{-4}
	GO:0015075~ion transmembrane transporter activity	168	16	2.16×10^{-4}
	GO:0005215~transporter activity	241	20	2.43×10^{-4}
	GO:0005216~ion channel activity	76	10	2.56×10^{-4}
	GO:0022857~transmembrane transporter activity	188	17	2.58×10^{-4}
	GO:0022838~substrate-specific channel activity	79	10	3.52×10^{-4}
	GO:0015267~channel activity	81	10	4.32×10^{-4}
	GO:0022803~passive transmembrane transporter activity	81	10	4.32×10^{-4}
	GO:0022891~substrate-specific transmembrane transporter activity	182	16	5.32×10^{-4}
KEGG	bta04912: GnRH signaling pathway	64	10	5.86×10^{-5}
	bta04270: Vascular smooth muscle contraction	74	10	2.05×10^{-4}
	bta04720: Long-term potentiation	41	7	4.35×10^{-4}

[1]$F1_{\%CY}$= Factor underlying the percentage of individual cheese yield; $F4_{Cheese\ N}$ = Factor underlying the protein in the cheese; $F8_{as_2-CN}$ = Factor underlying the αs_2-casein; $F10_{a-LA}$ = Factor underlying the α-LA.

[2]KEGG: KEGG pathway; GO_BP: GO biological process; GO_CC: GO cellular component; GO_MF: GO molecular function.

[3]Significant genes after mapping to Ensembl *Bos Taurus* UMD3.1 (http://www.ensembl.org/index.html)

[4]False discovery rate (FDR) correction for multiple testing (controlled at 5%).

Table S4. Significant genes involved in the enriched KEGG and GO categories

SNP	BTA	Location (bp)	Gene_Ensembl_ID	Gene_SYMBOL	Category	Trait
ARS-BFGL-NGS-108629	19	63530174	ENSBTAG00000001061	PRKCA	KEGG:bta04530	F1_%CY
ARS-BFGL-NGS-116740	13	65965727	ENSBTAG00000001640	EPB41L1	KEGG:bta04530	F1_%CY
BTB-00067270	1	141351312	ENSBTAG00000002029	IGSF5	KEGG:bta04530	F1_%CY
ARS-BFGL-NGS-103383	12	76787101	ENSBTAG00000003568	CLDN10	KEGG:bta04530	F1_%CY
ARS-BFGL-NGS-116126	3	94980177	ENSBTAG00000005337	RAB3B	KEGG:bta04530	F1_%CY
BTA-121254-no-rs	2	125098723	ENSBTAG00000006667	EPB41	KEGG:bta04530	F1_%CY
ARS-BFGL-NGS-118891	5	75114559	ENSBTAG00000010402	MYH9	KEGG:bta04530	F1_%CY
ARS-BFGL-NGS-31906	25	14257075	ENSBTAG00000015988	MYH11	KEGG:bta04530	F1_%CY
BTB-00632811	16	34181806	ENSBTAG00000017788	AKT3	KEGG:bta04530	F1_%CY
BTB-00437757	10	81069177	ENSBTAG00000018255	ACTN1	KEGG:bta04530	F1_%CY
Hapmap54744-rs29013475	3	84455380	ENSBTAG00000021975	INADL	KEGG:bta04530	F1_%CY
BTB-01561529	28	23349468	ENSBTAG00000045699	CTNNA3	KEGG:bta04530	F1_%CY
ARS-BFGL-NGS-81228	19	63625695	ENSBTAG00000012311	CACNG5	KEGG:bta05412	F4_Cheese N
ARS-BFGL-NGS-2876	22	46810135	ENSBTAG00000013117	CACNA2D3	KEGG:bta05412	F4_Cheese N
ARS-BFGL-NGS-117484	19	56495427	ENSBTAG00000018169	ITGB4	KEGG:bta05412	F4_Cheese N
ARS-BFGL-NGS-35716	24	29013292	ENSBTAG00000021190	CDH2	KEGG:bta05412	F4_Cheese N
BTB-00939873	26	33918405	ENSBTAG00000021574	TCF7L2	KEGG:bta05412	F4_Cheese N
ARS-BFGL-NGS-1585	24	25921286	ENSBTAG00000021923	DSG2	KEGG:bta05412	F4_Cheese N
Hapmap38792-BTA-92186	28	10127554	ENSBTAG00000022886	RYR2	KEGG:bta05412	F4_Cheese N
ARS-BFGL-NGS-61492	7	70011889	ENSBTAG00000044046	SGCD	KEGG:bta05412	F4_Cheese N
ARS-BFGL-NGS-77933	28	22792613	ENSBTAG00000045699	CTNNA3	KEGG:bta05412	F4_Cheese N
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	GO_BP:0055076	F8_αs2-CN
ARS-BFGL-NGS-37969	24	57316936	ENSBTAG00000006393	FECH	GO_BP:0055076	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0055076	F8_αs2-CN
ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_BP:0055076	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_BP:0055076	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_BP:0055076	F8_αs2-CN

ARS-BFGL-NGS-104728	8	10680165	ENSBTAG00000019636	SCARA5	GO_BP:0055076	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-102138	3	103021431	ENSBTAG00000000253	PTPRF	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_BP:0051716	F8_αs2-CN
Hapmap22746-BTA-141712	3	106235104	ENSBTAG00000000532	EXO5	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-107390	1	136623809	ENSBTAG00000000905	RAB6B	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-108131	27	36764232	ENSBTAG00000001244	PLAT	GO_BP:0051716	F8_αs2-CN
BTB-01882599	9	32606924	ENSBTAG00000002059	MCM9	GO_BP:0051716	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0051716	F8_αs2-CN
Hapmap31616-BTC-042811	6	71873004	ENSBTAG00000002699	KIT	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-5136	7	44901489	ENSBTAG00000003018	FSTL3	GO_BP:0051716	F8_αs2-CN
Hapmap47844-BTA-115673	6	113538490	ENSBTAG00000003218	RAB28	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_BP:0051716	F8_αs2-CN
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-116538	2	105169858	ENSBTAG00000003843	SMARCAL1	GO_BP:0051716	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0051716	F8_αs2-CN
Hapmap44203-BTA-115893	3	49431468	ENSBTAG00000004190	ARHGAP29	GO_BP:0051716	F8_αs2-CN
Hapmap53694-rs29015972	4	58023230	ENSBTAG00000004398	IMMP2L	GO_BP:0051716	F8_αs2-CN
BTB-00019705	1	42721268	ENSBTAG00000004471	ST3GAL6	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-106082	10	76711436	ENSBTAG00000004498	ESR2	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-55064	22	57139923	ENSBTAG00000004514	RAF1	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-24364	17	35250836	ENSBTAG00000005691	FGF2	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-37969	24	57316936	ENSBTAG00000006393	FECH	GO_BP:0051716	F8_αs2-CN
Hapmap58351-rs29010900	18	24897738	ENSBTAG00000006611	NUP93	GO_BP:0051716	F8_αs2-CN
BTB-01072640	9	24369582	ENSBTAG00000008121	RSPO3	GO_BP:0051716	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-103651	19	14742543	ENSBTAG00000010155	LOC504773	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-102190	24	43638478	ENSBTAG00000010792	SEH1L	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-38516	18	56280702	ENSBTAG00000011421	CD37	GO_BP:0051716	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0051716	F8_αs2-CN
BTA-69065-no-rs	3	101995252	ENSBTAG00000012355	RNF220	GO_BP:0051716	F8_αs2-CN

ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_BP:0051716	F8_αs2-CN
Hapmap49164-BTA-42015	17	74123863	ENSBTAG00000013038	UBE2L3	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-112106	13	76371179	ENSBTAG00000013336	EYA2	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0051716	F8_αs2-CN
BTA-34485-no-rs	14	33474585	ENSBTAG00000014691	ARFGEF1	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-80025	10	12386269	ENSBTAG00000015155	HACD3	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-101279	5	110399710	ENSBTAG00000015290	BAIAP2L2	GO_BP:0051716	F8_αs2-CN
BTB-00411248	10	11484520	ENSBTAG00000015716	ERO1A	GO_BP:0051716	F8_αs2-CN
Hapmap34322-BES8_Contig446_1023	3	110104792	ENSBTAG00000015969	STK40	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-37939	11	85215807	ENSBTAG00000016045	TRIB2	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-BAC-21453	14	44105320	ENSBTAG00000016284	IL7	GO_BP:0051716	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-27742	19	43948803	ENSBTAG00000016847	ARL4D	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_BP:0051716	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_BP:0051716	F8_αs2-CN
Hapmap59849-rs29025576	6	69601707	ENSBTAG00000018106	SPATA18	GO_BP:0051716	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-104728	8	10680165	ENSBTAG00000019636	SCARA5	GO_BP:0051716	F8_αs2-CN
ARS-USMARC-Parent-AY849380-no-rs	6	90562665	ENSBTAG00000019716	CXCL8	GO_BP:0051716	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_BP:0051716	F8_αs2-CN
UA-IFASA-6208	28	28149081	ENSBTAG00000021499	PSAP	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0051716	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0051716	F8_αs2-CN
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0051716	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0030001	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_BP:0030001	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0030001	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0030001	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0030001	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_BP:0030001	F8_αs2-CN

ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_BP:0030001	F8_αs2-CN
BTB-00411248	10	11484520	ENSBTAG00000015716	ERO1A	GO_BP:0030001	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_BP:0030001	F8_αs2-CN
ARS-BFGL-NGS-104728	8	10680165	ENSBTAG00000019636	SCARA5	GO_BP:0030001	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_BP:0030001	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0030001	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0030001	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_BP:0030001	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_BP:0030001	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_BP:0006811	F8_αs2-CN
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	GO_BP:0006811	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_BP:0006811	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0006811	F8_αs2-CN
Hapmap52874-ss46526209	2	122817488	ENSBTAG00000011582	SERINC2	GO_BP:0006811	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-97095	1	71431023	ENSBTAG00000012347	SLC51A	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_BP:0006811	F8_αs2-CN
BTB-00411248	10	11484520	ENSBTAG00000015716	ERO1A	GO_BP:0006811	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0006811	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-104728	8	10680165	ENSBTAG00000019636	SCARA5	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0006811	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0006811	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_BP:0006811	F8_αs2-CN

Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_BP:0006811	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_BP:0034220	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0034220	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_BP:0034220	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0034220	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-104728	8	10680165	ENSBTAG00000019636	SCARA5	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0034220	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0034220	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_BP:0034220	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_BP:0034220	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-107390	1	136623809	ENSBTAG00000000905	RAB6B	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-108131	27	36764232	ENSBTAG00000001244	PLAT	GO_BP:0023052	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0023052	F8_αs2-CN
Hapmap31616-BTC-042811	6	71873004	ENSBTAG00000002699	KIT	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-5136	7	44901489	ENSBTAG00000003018	FSTL3	GO_BP:0023052	F8_αs2-CN
Hapmap47844-BTA-115673	6	113538490	ENSBTAG00000003218	RAB28	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_BP:0023052	F8_αs2-CN
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	GO_BP:0023052	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0023052	F8_αs2-CN
Hapmap44203-BTA-115893	3	49431468	ENSBTAG00000004190	ARHGAP29	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-106082	10	76711436	ENSBTAG00000004498	ESR2	GO_BP:0023052	F8_αs2-CN

ARS-BFGL-NGS-55064	22	57139923	ENSBTAG00000004514	RAF1	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-24364	17	35250836	ENSBTAG00000005691	FGF2	GO_BP:0023052	F8_αs2-CN
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0023052	F8_αs2-CN
BTB-01072640	9	24369582	ENSBTAG00000008121	RSPO3	GO_BP:0023052	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-103651	19	14742543	ENSBTAG00000010155	LOC504773	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-102190	24	43638478	ENSBTAG00000010792	SEH1L	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-38516	18	56280702	ENSBTAG00000011421	CD37	GO_BP:0023052	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0023052	F8_αs2-CN
BTA-69065-no-rs	3	101995252	ENSBTAG00000012355	RNF220	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-112106	13	76371179	ENSBTAG00000013336	EYA2	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0023052	F8_αs2-CN
BTA-34485-no-rs	14	33474585	ENSBTAG00000014691	ARFGEF1	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-80025	10	12386269	ENSBTAG00000015155	HACD3	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-101279	5	110399710	ENSBTAG00000015290	BAIAP2L2	GO_BP:0023052	F8_αs2-CN
BTB-00411248	10	11484520	ENSBTAG00000015716	ERO1A	GO_BP:0023052	F8_αs2-CN
Hapmap34322-BES8_Contig446_1023	3	110104792	ENSBTAG00000015969	STK40	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-37939	11	85215807	ENSBTAG00000016045	TRIB2	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-BAC-21453	14	44105320	ENSBTAG00000016284	IL7	GO_BP:0023052	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-27742	19	43948803	ENSBTAG00000016847	ARL4D	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_BP:0023052	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_BP:0023052	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_BP:0023052	F8_αs2-CN
ARS-USMARC-Parent-AY849380-no-rs	6	90562665	ENSBTAG00000019716	CXCL8	GO_BP:0023052	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_BP:0023052	F8_αs2-CN
UA-IFASA-6208	28	28149081	ENSBTAG00000021499	PSAP	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0023052	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0023052	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0023052	F8_αs2-CN

ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0023052	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-107390	1	136623809	ENSBTAG00000000905	RAB6B	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-108131	27	36764232	ENSBTAG00000001244	PLAT	GO_BP:0044700	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0044700	F8_αs2-CN
Hapmap31616-BTC-042811	6	71873004	ENSBTAG00000002699	KIT	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-5136	7	44901489	ENSBTAG00000003018	FSTL3	GO_BP:0044700	F8_αs2-CN
Hapmap47844-BTA-115673	6	113538490	ENSBTAG00000003218	RAB28	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_BP:0044700	F8_αs2-CN
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	GO_BP:0044700	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0044700	F8_αs2-CN
Hapmap44203-BTA-115893	3	49431468	ENSBTAG00000004190	ARHGAP29	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-106082	10	76711436	ENSBTAG00000004498	ESR2	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-55064	22	57139923	ENSBTAG00000004514	RAF1	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-24364	17	35250836	ENSBTAG00000005691	FGF2	GO_BP:0044700	F8_αs2-CN
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0044700	F8_αs2-CN
BTB-01072640	9	24369582	ENSBTAG00000008121	RSPO3	GO_BP:0044700	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-103651	19	14742543	ENSBTAG00000010155	LOC504773	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-102190	24	43638478	ENSBTAG00000010792	SEH1L	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-38516	18	56280702	ENSBTAG00000011421	CD37	GO_BP:0044700	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0044700	F8_αs2-CN
BTA-69065-no-rs	3	101995252	ENSBTAG00000012355	RNF220	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-112106	13	76371179	ENSBTAG00000013336	EYA2	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0044700	F8_αs2-CN
BTA-34485-no-rs	14	33474585	ENSBTAG00000014691	ARFGEF1	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-80025	10	12386269	ENSBTAG00000015155	HACD3	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-101279	5	110399710	ENSBTAG00000015290	BAIAP2L2	GO_BP:0044700	F8_αs2-CN
BTB-00411248	10	11484520	ENSBTAG00000015716	ERO1A	GO_BP:0044700	F8_αs2-CN

Hapmap34322-BES8_Contig446_1023	3	110104792	ENSBTAG00000015969	STK40	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-37939	11	85215807	ENSBTAG00000016045	TRIB2	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-BAC-21453	14	44105320	ENSBTAG00000016284	IL7	GO_BP:0044700	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-27742	19	43948803	ENSBTAG00000016847	ARL4D	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_BP:0044700	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_BP:0044700	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_BP:0044700	F8_αs2-CN
ARS-USMARC-Parent-AY849380-no-rs	6	90562665	ENSBTAG00000019716	CXCL8	GO_BP:0044700	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_BP:0044700	F8_αs2-CN
UA-IFASA-6208	28	28149081	ENSBTAG00000021499	PSAP	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0044700	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0044700	F8_αs2-CN
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0044700	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-107390	1	136623809	ENSBTAG00000000905	RAB6B	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-108131	27	36764232	ENSBTAG00000001244	PLAT	GO_BP:0007165	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0007165	F8_αs2-CN
Hapmap31616-BTC-042811	6	71873004	ENSBTAG00000002699	KIT	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-5136	7	44901489	ENSBTAG00000003018	FSTL3	GO_BP:0007165	F8_αs2-CN
Hapmap47844-BTA-115673	6	113538490	ENSBTAG00000003218	RAB28	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_BP:0007165	F8_αs2-CN
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	GO_BP:0007165	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0007165	F8_αs2-CN
Hapmap44203-BTA-115893	3	49431468	ENSBTAG00000004190	ARHGAP29	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-106082	10	76711436	ENSBTAG00000004498	ESR2	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-55064	22	57139923	ENSBTAG00000004514	RAF1	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-24364	17	35250836	ENSBTAG00000005691	FGF2	GO_BP:0007165	F8_αs2-CN
BTB-01072640	9	24369582	ENSBTAG00000008121	RSPO3	GO_BP:0007165	F8_αs2-CN

BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-103651	19	14742543	ENSBTAG00000010155	LOC504773	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-102190	24	43638478	ENSBTAG00000010792	SEH1L	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-38516	18	56280702	ENSBTAG00000011421	CD37	GO_BP:0007165	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0007165	F8_αs2-CN
BTA-69065-no-rs	3	101995252	ENSBTAG00000012355	RNF220	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-112106	13	76371179	ENSBTAG00000013336	EYA2	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0007165	F8_αs2-CN
BTA-34485-no-rs	14	33474585	ENSBTAG00000014691	ARFGEF1	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-80025	10	12386269	ENSBTAG00000015155	HACD3	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-101279	5	110399710	ENSBTAG00000015290	BAIAP2L2	GO_BP:0007165	F8_αs2-CN
BTB-00411248	10	11484520	ENSBTAG00000015716	ERO1A	GO_BP:0007165	F8_αs2-CN
Hapmap34322-BES8_Contig446_1023	3	110104792	ENSBTAG00000015969	STK40	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-37939	11	85215807	ENSBTAG00000016045	TRIB2	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-BAC-21453	14	44105320	ENSBTAG00000016284	IL7	GO_BP:0007165	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-27742	19	43948803	ENSBTAG00000016847	ARL4D	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_BP:0007165	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_BP:0007165	F8_αs2-CN
ARS-USMARC-Parent-AY849380-no-rs	6	90562665	ENSBTAG00000019716	CXCL8	GO_BP:0007165	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_BP:0007165	F8_αs2-CN
UA-IFASA-6208	28	28149081	ENSBTAG00000021499	PSAP	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0007165	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0007165	F8_αs2-CN
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0007165	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-107390	1	136623809	ENSBTAG00000000905	RAB6B	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-108131	27	36764232	ENSBTAG00000001244	PLAT	GO_BP:0007154	F8_αs2-CN

BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0007154	F8_αs2-CN
Hapmap31616-BTC-042811	6	71873004	ENSBTAG00000002699	KIT	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-5136	7	44901489	ENSBTAG00000003018	FSTL3	GO_BP:0007154	F8_αs2-CN
Hapmap47844-BTA-115673	6	113538490	ENSBTAG00000003218	RAB28	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_BP:0007154	F8_αs2-CN
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	GO_BP:0007154	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0007154	F8_αs2-CN
Hapmap44203-BTA-115893	3	49431468	ENSBTAG00000004190	ARHGAP29	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-106082	10	76711436	ENSBTAG00000004498	ESR2	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-55064	22	57139923	ENSBTAG00000004514	RAF1	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-24364	17	35250836	ENSBTAG00000005691	FGF2	GO_BP:0007154	F8_αs2-CN
Hapmap58351-rs29010900	18	24897738	ENSBTAG00000006611	NUP93	GO_BP:0007154	F8_αs2-CN
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0007154	F8_αs2-CN
BTB-01072640	9	24369582	ENSBTAG00000008121	RSPO3	GO_BP:0007154	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-103651	19	14742543	ENSBTAG00000010155	LOC504773	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-102190	24	43638478	ENSBTAG00000010792	SEH1L	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-38516	18	56280702	ENSBTAG00000011421	CD37	GO_BP:0007154	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0007154	F8_αs2-CN
BTA-69065-no-rs	3	101995252	ENSBTAG00000012355	RNF220	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-112106	13	76371179	ENSBTAG00000013336	EYA2	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0007154	F8_αs2-CN
BTA-34485-no-rs	14	33474585	ENSBTAG00000014691	ARFGEF1	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-80025	10	12386269	ENSBTAG00000015155	HACD3	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-101279	5	110399710	ENSBTAG00000015290	BAIAP2L2	GO_BP:0007154	F8_αs2-CN
BTB-00411248	10	11484520	ENSBTAG00000015716	ERO1A	GO_BP:0007154	F8_αs2-CN
Hapmap34322-BES8_Contig446_1023	3	110104792	ENSBTAG00000015969	STK40	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-37939	11	85215807	ENSBTAG00000016045	TRIB2	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-BAC-21453	14	44105320	ENSBTAG00000016284	IL7	GO_BP:0007154	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0007154	F8_αs2-CN

ARS-BFGL-NGS-27742	19	43948803	ENSBTAG00000016847	ARL4D	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_BP:0007154	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_BP:0007154	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_BP:0007154	F8_αs2-CN
ARS-USMARC-Parent-AY849380-no-rs	6	90562665	ENSBTAG00000019716	CXCL8	GO_BP:0007154	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_BP:0007154	F8_αs2-CN
UA-IFASA-6208	28	28149081	ENSBTAG00000021499	PSAP	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0007154	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0007154	F8_αs2-CN
ARS-BFGL-NGS-65615	25	41309842	ENSBTAG00000040361	LFNG	GO_BP:0007154	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_BP:0098660	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_BP:0098660	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_BP:0098660	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_BP:0098660	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0098660	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0098660	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0098660	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_BP:0098660	F8_αs2-CN
ARS-BFGL-NGS-104728	8	10680165	ENSBTAG00000019636	SCARA5	GO_BP:0098660	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_BP:0098660	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0098660	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0098660	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_BP:0098660	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_BP:0098660	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_BP:0003008	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_BP:0003008	F8_αs2-CN
ARS-BFGL-NGS-84901	15	35537347	ENSBTAG00000000986	USH1C	GO_BP:0003008	F8_αs2-CN
Hapmap52482-ss46526125	6	73639744	ENSBTAG00000002333	HOPX	GO_BP:0003008	F8_αs2-CN
Hapmap31616-BTC-042811	6	71873004	ENSBTAG00000002699	KIT	GO_BP:0003008	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0003008	F8_αs2-CN

Hapmap53694-rs29015972	4	58023230	ENSBTAG00000004398	IMMP2L	GO_BP:0003008	F8_αs2-CN
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_BP:0003008	F8_αs2-CN
Hapmap49307-BTA-78198	6	16085913	ENSBTAG00000008332	ENPEP	GO_BP:0003008	F8_αs2-CN
ARS-BFGL-BAC-38065	25	33314090	ENSBTAG00000009780	GTF2I	GO_BP:0003008	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_BP:0003008	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_BP:0003008	F8_αs2-CN
Hapmap34322-BES8_Contig446_1023	3	110104792	ENSBTAG00000015969	STK40	GO_BP:0003008	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_BP:0003008	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_BP:0003008	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_BP:0003008	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0003008	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_BP:0003008	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_BP:0003008	F8_αs2-CN
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	GO_BP:0055065	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_BP:0055065	F8_αs2-CN
ARS-BFGL-NGS-37969	24	57316936	ENSBTAG00000006393	FECH	GO_BP:0055065	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_BP:0055065	F8_αs2-CN
ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_BP:0055065	F8_αs2-CN
BTB-00411248	10	11484520	ENSBTAG00000015716	ERO1A	GO_BP:0055065	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_BP:0055065	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_BP:0055065	F8_αs2-CN
ARS-BFGL-NGS-104728	8	10680165	ENSBTAG00000019636	SCARA5	GO_BP:0055065	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_BP:0055065	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0097458	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-NGS-102138	3	103021431	ENSBTAG00000000253	PTPRF	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0097458	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_CC:0097458	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-BAC-38065	25	33314090	ENSBTAG00000009780	GTF2I	GO_CC:0097458	F8_αs2-CN

ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-NGS-66442	13	60094367	ENSBTAG00000013009	AURKA	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0097458	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0097458	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0097458	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_CC:0097458	F8_αs2-CN
Hapmap50702-BTA-47830	2	63065836	ENSBTAG00000018753	TMEM163	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_CC:0097458	F8_αs2-CN
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT_11	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0097458	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0097458	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_CC:0097458	F8_αs2-CN
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0097458	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0044456	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0044456	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0044456	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_CC:0044456	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0044456	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0044456	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_CC:0044456	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0044456	F8_αs2-CN
Hapmap50702-BTA-47830	2	63065836	ENSBTAG00000018753	TMEM163	GO_CC:0044456	F8_αs2-CN
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT_11	GO_CC:0044456	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0044456	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0044456	F8_αs2-CN
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0044456	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0098794	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0098794	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_CC:0098794	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0098794	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0098794	F8_αs2-CN

ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_CC:0098794	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0098794	F8_αs2-CN
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT_11	GO_CC:0098794	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0098794	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0098794	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0043005	F8_αs2-CN
ARS-BFGL-NGS-102138	3	103021431	ENSBTAG00000000253	PTPRF	GO_CC:0043005	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0043005	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_CC:0043005	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0043005	F8_αs2-CN
ARS-BFGL-NGS-66442	13	60094367	ENSBTAG00000013009	AURKA	GO_CC:0043005	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0043005	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0043005	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0043005	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_CC:0043005	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_CC:0043005	F8_αs2-CN
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT_11	GO_CC:0043005	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0043005	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0043005	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_CC:0043005	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0045202	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0045202	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0045202	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_CC:0045202	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0045202	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0045202	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_CC:0045202	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0045202	F8_αs2-CN
Hapmap50702-BTA-47830	2	63065836	ENSBTAG00000018753	TMEM163	GO_CC:0045202	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_CC:0045202	F8_αs2-CN
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT_11	GO_CC:0045202	F8_αs2-CN

ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0045202	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0045202	F8_αs2-CN
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0045202	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0097060	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0097060	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0097060	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0097060	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_CC:0097060	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0097060	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0097060	F8_αs2-CN
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0097060	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0030424	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0030424	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_CC:0030424	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0030424	F8_αs2-CN
ARS-BFGL-NGS-66442	13	60094367	ENSBTAG00000013009	AURKA	GO_CC:0030424	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0030424	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_CC:0030424	F8_αs2-CN
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT_11	GO_CC:0030424	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0030424	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0042995	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-NGS-102138	3	103021431	ENSBTAG00000000253	PTPRF	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_CC:0042995	F8_αs2-CN
Hapmap47844-BTA-115673	6	113538490	ENSBTAG00000003218	RAB28	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0042995	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_CC:0042995	F8_αs2-CN
BTB-00184917	4	51792757	ENSBTAG00000004072	CAPZA2	GO_CC:0042995	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-BAC-38065	25	33314090	ENSBTAG00000009780	GTF2I	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_CC:0042995	F8_αs2-CN

ARS-BFGL-NGS-66442	13	60094367	ENSBTAG00000013009	AURKA	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0042995	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0042995	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0042995	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_CC:0042995	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_CC:0042995	F8_αs2-CN
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT_11	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0042995	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_CC:0042995	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0098590	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_CC:0098590	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_CC:0098590	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_CC:0098590	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0098590	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0098590	F8_αs2-CN
ARS-BFGL-NGS-106737	3	110043222	ENSBTAG00000007994	OSCP1	GO_CC:0098590	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0098590	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0098590	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_CC:0098590	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_CC:0098590	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0098590	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0098590	F8_αs2-CN
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0098590	F8_αs2-CN
ARS-BFGL-NGS-102172	18	46826215	ENSBTAG00000000153	LRFN3	GO_CC:0098589	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_CC:0098589	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_CC:0098589	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_CC:0098589	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0098589	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0098589	F8_αs2-CN

ARS-BFGL-NGS-106737	3	110043222	ENSBTAG00000007994	OSCP1	GO_CC:0098589	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_CC:0098589	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_CC:0098589	F8_αs2-CN
ARS-BFGL-NGS-119788	5	30185840	ENSBTAG00000017504	FAIM2	GO_CC:0098589	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0098589	F8_αs2-CN
Hapmap41549-BTA-46518	19	15038933	ENSBTAG00000020316	AP2B1	GO_CC:0098589	F8_αs2-CN
BTB-00883964	24	29132144	ENSBTAG00000021190	CDH2	GO_CC:0098589	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0098589	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_CC:0098589	F8_αs2-CN
BTB-01060234	11	32409123	ENSBTAG00000046199		GO_CC:0098589	F8_αs2-CN
BTB-00646135	16	52141211	ENSBTAG00000000215	GNB1	GO_CC:0044463	F8_αs2-CN
ARS-BFGL-NGS-112313	19	51767413	ENSBTAG00000000354	PDE6G	GO_CC:0044463	F8_αs2-CN
Hapmap47844-BTA-115673	6	113538490	ENSBTAG00000003218	RAB28	GO_CC:0044463	F8_αs2-CN
ARS-BFGL-NGS-23871	7	12617924	ENSBTAG00000003675	ADGRL1	GO_CC:0044463	F8_αs2-CN
BTB-00026110	1	67331639	ENSBTAG00000003865	CASR	GO_CC:0044463	F8_αs2-CN
UA-IFASA-5504	1	60889674	ENSBTAG00000006451	GAP43	GO_CC:0044463	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	GO_CC:0044463	F8_αs2-CN
ARS-BFGL-NGS-66442	13	60094367	ENSBTAG00000013009	AURKA	GO_CC:0044463	F8_αs2-CN
Hapmap54975-rs29019203	1	9636803	ENSBTAG00000017753	APP	GO_CC:0044463	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_CC:0044463	F8_αs2-CN
BTB-00268267	6	93589032	ENSBTAG00000021372	SEPT_11	GO_CC:0044463	F8_αs2-CN
ARS-BFGL-NGS-41201	10	10385233	ENSBTAG00000025853	HOMER1	GO_CC:0044463	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0022836	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0022836	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0022836	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0022836	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0022836	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0022836	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0022836	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0022836	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0022836	F8_αs2-CN

BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0046873	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_MF:0046873	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0046873	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_MF:0046873	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_MF:0046873	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_MF:0046873	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0046873	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0046873	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0046873	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0046873	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0015075	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_MF:0015075	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0015075	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0015075	F8_αs2-CN
Hapmap52874-ss46526209	2	122817488	ENSBTAG00000011582	SERINC2	GO_MF:0015075	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_MF:0015075	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0015075	F8_αs2-CN
ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_MF:0015075	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0015075	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_MF:0015075	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_MF:0015075	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0015075	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0015075	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0015075	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0015075	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0015075	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0005215	F8_αs2-CN
Hapmap27109-BTC-060711	6	87152621	ENSBTAG00000007695	CSN1S1	GO_MF:0005215	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0005215	F8_αs2-CN

Hapmap52874-ss46526209	2	122817488	ENSBTAG00000011582	SERINC2	GO_MF:0005215	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-NGS-97095	1	71431023	ENSBTAG00000012347	SLC51A	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_MF:0005215	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0005215	F8_αs2-CN
BTA-22850-no-rs	6	37983812	ENSBTAG00000017704	ABCG2	GO_MF:0005215	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_MF:0005215	F8_αs2-CN
Hapmap41549-BTA-46518	19	15038933	ENSBTAG00000020316	AP2B1	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0005215	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0005215	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0005215	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0005215	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0005216	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0005216	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0005216	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0005216	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0005216	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0005216	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0005216	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0005216	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0005216	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0005216	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0022857	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_MF:0022857	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0022857	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0022857	F8_αs2-CN
Hapmap52874-ss46526209	2	122817488	ENSBTAG00000011582	SERINC2	GO_MF:0022857	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_MF:0022857	F8_αs2-CN

ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0022857	F8_αs2-CN
ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_MF:0022857	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0022857	F8_αs2-CN
BTA-22850-no-rs	6	37983812	ENSBTAG00000017704	ABCG2	GO_MF:0022857	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_MF:0022857	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_MF:0022857	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0022857	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0022857	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0022857	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0022857	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0022857	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0022838	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0022838	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0022838	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0022838	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0022838	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0022838	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0022838	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0022838	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0022838	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0022838	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0015267	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0015267	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0015267	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0015267	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0015267	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0015267	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0015267	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0015267	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0015267	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0015267	F8_αs2-CN

BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0022803	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0022803	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0022803	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0022803	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0022803	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0022803	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0022803	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0022803	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0022803	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0022803	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	GO_MF:0022891	F8_αs2-CN
ARS-BFGL-NGS-54753	6	88210451	ENSBTAG00000002348	SLC4A4	GO_MF:0022891	F8_αs2-CN
ARS-BFGL-NGS-24122	5	65228699	ENSBTAG00000004145	ANO4	GO_MF:0022891	F8_αs2-CN
BTB-00661682	16	74429897	ENSBTAG00000008710	KCNH1	GO_MF:0022891	F8_αs2-CN
Hapmap52874-ss46526209	2	122817488	ENSBTAG00000011582	SERINC2	GO_MF:0022891	F8_αs2-CN
Hapmap41906-BTA-57339	1	140476945	ENSBTAG00000011626	ATP2C1	GO_MF:0022891	F8_αs2-CN
ARS-BFGL-NGS-113181	7	65047126	ENSBTAG00000014395	GLRA1	GO_MF:0022891	F8_αs2-CN
ARS-BFGL-NGS-13562	10	5561417	ENSBTAG00000014536	SFXN1	GO_MF:0022891	F8_αs2-CN
BTA-110747-no-rs	6	67170035	ENSBTAG00000016645	GABRA4	GO_MF:0022891	F8_αs2-CN
Hapmap42572-BTA-50505	20	37630456	ENSBTAG00000018245	SLC1A3	GO_MF:0022891	F8_αs2-CN
ARS-BFGL-NGS-3781	19	58875702	ENSBTAG00000018837	SLC39A11	GO_MF:0022891	F8_αs2-CN
ARS-BFGL-BAC-2036	14	9906433	ENSBTAG00000020667	KCNQ3	GO_MF:0022891	F8_αs2-CN
ARS-BFGL-NGS-94704	10	12486527	ENSBTAG00000025826	SLC24A1	GO_MF:0022891	F8_αs2-CN
BTB-00319838	7	75728672	ENSBTAG00000030286	GABRA1	GO_MF:0022891	F8_αs2-CN
ARS-BFGL-NGS-75302	14	38418141	ENSBTAG00000040496	KCNB2	GO_MF:0022891	F8_αs2-CN
Hapmap48459-BTA-75920	6	42023749	ENSBTAG00000047743	KCNIP4	GO_MF:0022891	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	KEGG:bta04912	F8_αs2-CN
ARS-BFGL-NGS-55064	22	57139923	ENSBTAG00000004514	RAF1	KEGG:bta04912	F8_αs2-CN
UA-IFASA-6018	22	47721961	ENSBTAG00000010026	CACNA1D	KEGG:bta04912	F8_αs2-CN
ARS-BFGL-NGS-85264	5	109382111	ENSBTAG00000010660	CACNA1C	KEGG:bta04912	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	KEGG:bta04912	F8_αs2-CN

ARS-BFGL-NGS-8293	5	110435582	ENSBTAG00000014015	PLA2G6	KEGG:bta04912	F8_αs2-CN
Hapmap31246-BTC-009295	14	10884437	ENSBTAG00000014600	ADCY8	KEGG:bta04912	F8_αs2-CN
BTA-44254-no-rs	1	96553339	ENSBTAG00000017490	PLD1	KEGG:bta04912	F8_αs2-CN
ARS-BFGL-NGS-53579	20	65584868	ENSBTAG00000019210	ADCY2	KEGG:bta04912	F8_αs2-CN
ARS-BFGL-NGS-10138	10	13239142	ENSBTAG00000033983	MAP2K1	KEGG:bta04912	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	KEGG:bta04270	F8_αs2-CN
BTB-02001746	17	44532671	ENSBTAG00000003840	GUCY1B3	KEGG:bta04270	F8_αs2-CN
ARS-BFGL-NGS-55064	22	57139923	ENSBTAG00000004514	RAF1	KEGG:bta04270	F8_αs2-CN
ARS-BFGL-NGS-1951	15	42599781	ENSBTAG00000007129	MRVI1	KEGG:bta04270	F8_αs2-CN
UA-IFASA-6018	22	47721961	ENSBTAG00000010026	CACNA1D	KEGG:bta04270	F8_αs2-CN
ARS-BFGL-NGS-85264	5	109382111	ENSBTAG00000010660	CACNA1C	KEGG:bta04270	F8_αs2-CN
ARS-BFGL-NGS-8293	5	110435582	ENSBTAG00000014015	PLA2G6	KEGG:bta04270	F8_αs2-CN
Hapmap31246-BTC-009295	14	10884437	ENSBTAG00000014600	ADCY8	KEGG:bta04270	F8_αs2-CN
ARS-BFGL-NGS-53579	20	65584868	ENSBTAG00000019210	ADCY2	KEGG:bta04270	F8_αs2-CN
ARS-BFGL-NGS-10138	10	13239142	ENSBTAG00000033983	MAP2K1	KEGG:bta04270	F8_αs2-CN
Hapmap51915-BTA-74618	5	96479969	ENSBTAG00000000219	GRIN2B	KEGG:bta04720	F8_αs2-CN
BTA-27682-no-rs	5	83402906	ENSBTAG00000002313	ITPR2	KEGG:bta04720	F8_αs2-CN
ARS-BFGL-NGS-55064	22	57139923	ENSBTAG00000004514	RAF1	KEGG:bta04720	F8_αs2-CN
ARS-BFGL-NGS-85264	5	109382111	ENSBTAG00000010660	CACNA1C	KEGG:bta04720	F8_αs2-CN
ARS-BFGL-NGS-43041	4	77762822	ENSBTAG00000012653	CAMK2B	KEGG:bta04720	F8_αs2-CN
Hapmap31246-BTC-009295	14	10884437	ENSBTAG00000014600	ADCY8	KEGG:bta04720	F8_αs2-CN
ARS-BFGL-NGS-10138	10	13239142	ENSBTAG00000033983	MAP2K1	KEGG:bta04720	F8_αs2-CN
ARS-BFGL-NGS-14087	1	66721297	ENSBTAG00000001009	HCLS1	KEGG:bta04530	F10_α-LA
ARS-BFGL-NGS-37708	18	49928003	ENSBTAG00000001400	AKT2	KEGG:bta04530	F10_α-LA
Hapmap23996-BTA-145153	7	59974127	ENSBTAG00000001862	PPP2R2B	KEGG:bta04530	F10_α-LA
ARS-BFGL-NGS-4287	18	56894807	ENSBTAG00000002580	MYH14	KEGG:bta04530	F10_α-LA
ARS-BFGL-NGS-30004	23	17013312	ENSBTAG00000011189	TJAP1	KEGG:bta04530	F10_α-LA
ARS-BFGL-NGS-116613	13	19227633	ENSBTAG00000014991	PARD3	KEGG:bta04530	F10_α-LA
ARS-BFGL-NGS-99149	25	14216892	ENSBTAG00000015988	MYH11	KEGG:bta04530	F10_α-LA
Hapmap58028-ss46526484	19	27602402	ENSBTAG00000019448	CLDN7	KEGG:bta04530	F10_α-LA
BTB-01260337	11	28183453	ENSBTAG00000020614	PRKCE	KEGG:bta04530	F10_α-LA

BTB-00745347	19	28768067	ENSBTAG00000021151	MYH10	KEGG:bta04530	F10_α-LA
BTA-29968-no-rs	3	84123491	ENSBTAG00000021975	INADL	KEGG:bta04530	F10_α-LA
Hapmap33212-BTA-145669	8	31123907	ENSBTAG00000043961	MPDZ	KEGG:bta04530	F10_α-LA
ARS-BFGL-NGS-91878	28	23149412	ENSBTAG00000045699	CTNNA3	KEGG:bta04530	F10_α-LA

Description: SNP= the name of the single nucleotide polymorphism ; BTA = *Bos taurus* autosome chromosome; Ensembl *Bos taurus*UMD3.1 (http://www.ensembl.org/index.html); KEGG: KEGG pathway; GO_BP: GO biological process; GO_CC: GO cellular component; GO_MF: GO molecular function; $F1_{\%CY}$ = Factor underlying the percentage of individual cheese yield; $F4_{Cheese\ N}$ = Factor underlying the protein in the cheese; $F8_{as2\text{-}CN}$ = Factor underlying the αs_2-casein; $F10_{\alpha\text{-}LA}$ = Factor underlying the α-LA.

APPENDIX II: List of abbreviations and trait definitions

Milk coagulation properties (MCP)

RCT: rennet coagulation time (min)

k_{20} : interval from RCT to the time at which a curd firmness (CF) of 20 mm is attained (min)

a_{30}: measure of the extent of curd firmness 30 min after coagulant addition (mm)

Curd firming (CF_t) and syneresis traits

RCTeq: rennet coagulation time (min)

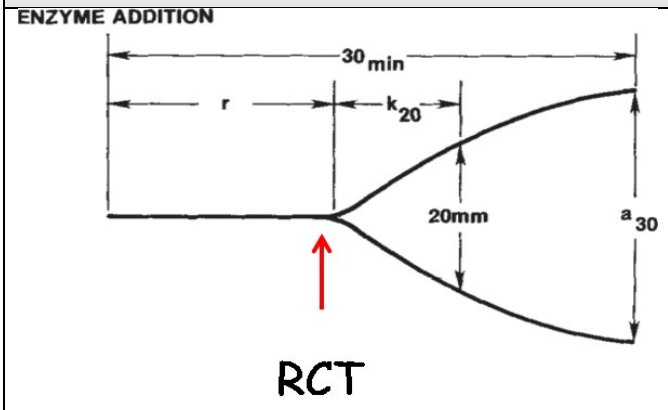
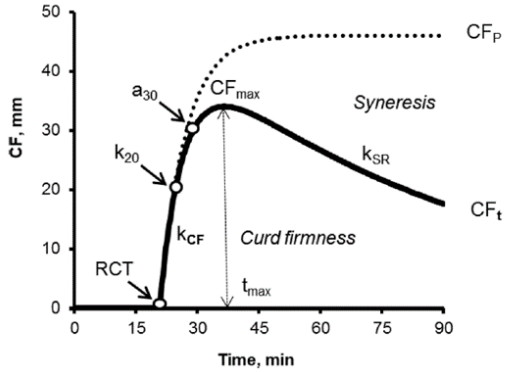
CF_P: asymptotic potential curd firmness in absence of syneresis(mm)

CF_{max}: the maximum curd firmness achieved within 45 min (mm)

k_{CF}: curd firming instant rate constant (%)

k_{SR}: syneresis instant rate constant (%)

t_{max}: the time at achievement of CF_{max}

MCP	CF_t
	
McMahon and Brown (1982). J. Dairy Sci. 65:1639-1642	Bittante et al. (2013). J. Dairy Sci. 96:7966-7979

Cheese Yield (CY) and nutrient recoveries (REC)

%CY$_{CURD}$: weight of fresh curd as percentage of weight of milk processed

%CY$_{SOLIDS}$: weight of curd solids as percentage of weight of milk processed

%CY $_{WATER}$: weight of water curd as percentage of weight of milk processed

REC$_{PROTEIN}$: protein of the curd as percentage of the protein of the milk processed

REC$_{FAT}$: fat of the curd as percentage of the fat of the milk processed

REC$_{SOLIDS}$: solids of the curd as percentage of the solids of the milk processed

REC$_{ENERGY}$: energy of the curd as percentage of energy of the milk processed

FACTORS

MFA = Multivariate factor analysis

FA = Factor analysis (the same as MFA)

F1$_{\%CY}$ (**F1: %CY**) = Factor underlying the percentage of individual cheese yield

F2$_{CFt}$ (**F2: CF$_t$**) = Factor underlying the milk curd firmness

F3$_{Yield}$ (**F3: Yield**) = Factor underlying milk yield

F4$_{Cheese\ N}$ (**F4: Cheese N**) = Factor underlying the protein in the cheese

F5 $_{\alpha s_1\text{-}\beta\text{-}CN}$ (**F5: as$_1$-β-CN**) = Factor underlying the αs_1 and β caseins

F6$_{Udder\ health}$ (**F6: Udder health**) = Factor underlying the udder health condition of a cow

F7$_{\beta\text{-}\kappa\text{-}CN}$ (**F7: κ-β-CN**) = Factor underlying the κ and β caseins

F8$_{\alpha s_2\text{-}CN}$ (**F8: as$_2$-CN**) = Factor underlying the αs_2-casein

F9$_{\alpha s_1\text{-}CN\text{-}Ph}$ (**F9: as$_1$-CN-Ph**) = Factor underlying the phosphorylated αs_1-casein

F10$_{\alpha\text{-}LA}$ (**F10: α-LA**) = Factor underlying the α-LA.

APPENDIX III: R scripts used in the research

This I believe in genetics: discovery can be a nuisance, ***replication is science***… (John. P.A. Ioanidis. Front Genet. 2013; 4: 33. doi: 10.3389/fgene.2013.00033).

This part of the PhD thesis contains *toy-examples* of the statistical analysis carried out throughout the 3 years of research. The coding requires some basic knowledge R, at minimum. Differences exist between the scripts presented here and the actual analysis, hence the name *toy-examples*. The scripts are far to be considered as elegant (which was not the goal). The reader is strongly recommended to read the official R packages!

Outline

- **Example 1**: Single marker GWAS with GenABEL (GRAMMAR-GC). An extended R script with basic information that can be extracted from the GenABEL package
- **Example 2**: Manhattan plots of the GWAS results. After uploading the results from the 1st Example (`my.gwa.csv`), Manhattan plots are constructed.
- **Example 3**: Single marker GWAS with GenABEL (GRAMMAR-GC). Summarizes basic GWAS results when a large number of phenotypes need to be analyzed.
- **Example 4**: Gene-set enrichment and pathways analysis. This R coding was provided from Prof. Dr. Francisco Peñagaricano (Department of Animal Sciences,University of Florida), who is fully acknowledged for this, as well as his help in setting up the gene-set enrichment and pathway analysis presented in the 3rd Chapter of the thesis. This Example is further splitted into 4 sub-parts:
    a) Mapping SNP to the genes (Note! This has to be done twice, once for the background SNP once for the significant SNP)
    b) Querying the GO and KEGG databases
    c) Identification of genes in the significant GO terms
    d) Identification of genes in the significant KEGG categories
- **Example 5**: Factor analysis

Example 1. Single marker GWAS with GenABEL (GRAMMAR-GC)

```
#######################################################
# Script for fast GWAS with GenABEL                   #
# R package: GenABEL                                  #
#     Method: GRAMMAR-GC                              #
#     GenABEL data: ge03d2                            #
#     Analysis: GWAS for height and weight            #
# Date: 22-03-2015                                    #
#######################################################


rm(list=ls(all=TRUE))
library("GenABEL")
library("LDheatmap")

## Upload GenABEL data
require(GenABEL.data)
data(ge03d2)
# keep subset of the data to make the example faster
my.data <- ge03d2[seq(from=1,to=nids(ge03d2),by=2),
                  seq(from=1,to=nsnps(ge03d2),by=2)]
#########################################################################
# In this example data are already uploaded.
# However, below a 3-lines code on how to import your own data
# dir <- "C:\\Users\\dadousis\\Documents\\"    # directory of the files

#convert.snp.tped(tped=file.path(dir,"mygenotypes.tped"),
#                 tfam=file.path(dir,"mygenotypes.tfam"),strand = "+",
#                 bcast = 10000,out="gwa.raw")

## Import genotypes in PLINK style (modificato e trasposto!)
#     convert.snp.tped: function to convert genotypic data in transposed-
#ped format (.tped and .tfam) to internal genotypic data formatted file
#     tped: Name of transposed-ped format (.tped) file to read
#     tfam: Name of individual data (.tfam) file to read
#     out: Name for output data file
#     strand: Specification of strand, one of "u" (unknown),
# "+", "-" or "file". In the latter case, extra column specifying the
#strand (again, one of "u", "+", or "-") should be included in the tpedfile
#     bcast: Reports progress every time this number of SNPs have been read
#     Note: The function does not check if "outfile" already exists, thus
#it is always over-written
#########################################################################

## summary statistics on phenotypes and genotypes of the raw data
head(my.data@phdata)  # view the first six rows of the phenos
dim(my.data@phdata)  # check dimensions of the phenos
summary(my.data[, 1:10])  # check first 10 SNP summary
summary(my.data@phdata)  # summary startistics of the phenos
descriptives.trait(my.data,digits = 3)    # basic descriptive statistics of
#the phenos
descriptives.marker(my.data)  # basic descriptive statistics of the
#genotypes
View(my.data@phdata$id) # check the id
my.data@gtdata@nsnps    # number of SNP
```

```
my.data@gtdata@nids           # number of individuals with genotypes
length(my.data@phdata$id)     # number of individuals with phenotypes
View(my.data@gtdata@snpnames)       # SNP names
View(my.data@gtdata@chromosome)           # Chromosomes
View(my.data@gtdata@map)                  # position of SNP on the
#chromosome
as.genotype(my.data@gtdata[1:5,1:3])      # A function to convert an object
of snp.data-class to "genotype" data frame
as.hsgeno(my.data@gtdata[1:5,1:3])        # Attempts to convert gwaa.data
#to "hsgeno"
as.double(my.data@gtdata[1:5,1:3])        # Attempts to convert snp.data to
#double


# number of SNP per chromosome
sort(table(my.data@gtdata@chromosome))# Chromosomes
sort(table(my.data@gtdata@chromosome)/sum(table(my.data@gtdata@chromosome))
)# Chromosomes

## HWE on the raw data
# extract the exact HWE test P-values into separate vector "Pexact"
# perform chi square test on the Pexact values and calculate ? (inflation
#factor).
# If ?=1.0 no inflation or diflation of test statistic (i.e. no
#stratification effect)

Pexactr <- summary(my.data@gtdata)[,"Pexact"]
estlambda(Pexactr, plot=TRUE)

## quality check
#############################################################################
#     callrate: cut-off SNP call rate
#     perid.call: cut-off individual call rate (maximum percent of missing
#genotypes in a person)
#     extr.call: SNPs with this low call rate are dropped prior to main
#analysis
#     extr.perid.call: people with this low call rate are dropped prior to
#main analysis
#     maf: cut-off Minor Allele Frequency. If not specified, the default
#value is 5 chromosomes 5/(2*nids(data))
#     ibs.mrk: How many random markers should be used to estimate IBS.
# When ibs.mrk < 1, IBS checks are turned off. When "all" all markers are
#used.                                                #
#     p.level: cut-off p-value in check for Hardy-Weinberg Equilibrium. If
#negative, FDR is applied
#############################################################################


mydata.qc <- check.marker(my.data,call=0.95,
                          perid.call=0.95,
                          extr.call = 0.1,
                          extr.perid.call = 0.1,
                          maf=0.005,
                          p.lev=0,
                          ibs.mrk =0 )

summary(mydata.qc)
```

```
names(mydata.qc)
View(mydata.qc$idok)
length(mydata.qc$idok)
length(mydata.qc$snpok)


## subset clean data
mydata.clean <- my.data[mydata.qc$idok,mydata.qc$snpok]


## summary statistics on phenotypes and genotypes of the filtered data
head(mydata.clean@phdata)  # view the first six rows of the phenos
dim(mydata.clean@phdata)  # check dimensions of the phenos
summary(mydata.clean[,1:10])  # check first 10 SNP summary
summary(mydata.clean@phdata)  # summary startistics of the phenos
descriptives.trait(mydata.clean,digits = 3)  # basic descriptive statistics
#of the phenos
descriptives.marker(mydata.clean)   # basic descriptive statistics of the
#genotypes
mydata.clean@gtdata@nsnps     # number of SNP
mydata.clean@gtdata@nids            # number of individuals
View(mydata.clean@gtdata@snpnames)        # SNP names
View(mydata.clean@gtdata@chromosome)            # Chromosomes
View(mydata.clean@gtdata@map)             # position of SNP on the
chromosome
as.genotype(mydata.clean@gtdata[1:5,1:3]) # A function to convert an object
#of snp.data-class to "genotype" data frame
as.hsgeno(mydata.clean@gtdata[1:5,1:3])         # Attempts to convert
#gwaa.data to "hsgeno"
as.double(mydata.clean@gtdata[1:5,1:3])         # Attempts to convert
#snp.data to double


# number of SNP per chromosome
sort(table(mydata.clean@gtdata@chromosome))# Chromosomes
sort(table(mydata.clean@gtdata@chromosome)/sum(table(mydata.clean@gtdata@ch
romosome)))# Chromosomes


## HWE on the filtered-cleaned data
# extract the exact HWE test P-values into separate vector "Pexact"
# perform chi square test on the Pexact values and calculate ? (inflation
#factor).
# If ?=1.0 no inflation or diflation of test statistic (i.e. no
#stratification effect)

Pexactc <- summary(mydata.clean@gtdata)[,"Pexact"]
estlambda(Pexactc, plot=TRUE)


##### Linkage Disequilibrium (LD)
# D' (dprime): A (Nsnps X Nsnps) matrix giving D' values between a pairs of
#SNPs above the diagonal
#     and number of SNP genotype measured for both SNPs below the diagonal
# R^2 (r2): A (Nsnps X Nsnps) matrix giving r2 values between a pairs of
#SNPs above the diagonal
#     and number of SNP genotype measured for both SNPs below the diagonal
```

```
dprime.ld <- dprfast(mydata.clean,snps=c(1:10))
r2.ld <- r2fast(mydata.clean,snps=c(1:10))
print(dprime.ld)
print(r2.ld)

## LD heatmap
# D'
dprime.heatmap <- LDheatmap(dprime.ld, color = grey.colors(20))
dprime.heatmap.names <- LDheatmap(dprime.heatmap,
SNP.name=rownames(dprime.ld)) # to include the SNP names in the graph

# R^2
r2.heatmap <- LDheatmap(r2.ld, color = grey.colors(20))
r2.heatmap.names <- LDheatmap(r2.heatmap, SNP.name=rownames(r2.ld)) # to
#include the SNP names in the graph


## genomic relationship
# The numbers below the diagonal show the genomic estimate of kinship
#('genome-wide IBD'),
# The numbers on the diagonal correspond to 0.5 plus the genomic
#homozigosity
# The numbers above the diagonal tell how many SNPs were typed successfully
#for both subjects

mydata.gkin <- ibs(mydata.clean[,autosomal(mydata.clean)],weight="freq")
dim(mydata.gkin)
mydata.gkin[1:6, 1:6]


## extract G matrix in full format (gkin is a lower triange)
library(matrixcalc)
G <- as.matrix(lower.triangle(as.matrix(mydata.gkin)))
G <- G + t(G) - diag(diag(G))
G[1:6, 1:6]


#########################################################################
#                                PERFORM GWAS                           #
# GRAMMAR-GC GWAS                                                       #
# Note: GenABEL accepts only one random effect (apart from residuals) in #
#regression, the polygenic term!!                                       #
#########################################################################

# GWAS for height                                                       #
height.poly <- polygenic(height ~ sex + age, kin=mydata.gkin,
data=mydata.clean)
height.poly$est
height.grammar <- grammar(height.poly,data=mydata.clean,method="gamma")

# GWAS for weight
weight.poly <- polygenic(weight ~ sex + age, kin=mydata.gkin,
data=mydata.clean)
weight.poly$est
weight.grammar <- grammar(weight.poly,data=mydata.clean,method="gamma")


```

```
226
227 ## all GWAS results
228 gwa.height <- descriptives.scan(height.grammar,
229                                 top= length(colnames(mydata.clean@gtdata)),
230                                  sort="Pc1df")
231 gwa.weight <- descriptives.scan(weight.grammar,
232                                  top= length(colnames(mydata.clean@gtdata)),
233                                  sort="Pc1df")
234
235
236 ## top10 hits (significant or not)
237 height.top <- descriptives.scan(height.grammar,sort="Pc1df")
238 height.top
239
240 weight.top <- descriptives.scan(weight.grammar,sort="Pc1df")
241 weight.top
242
243
244 ## keep only significant results
245 # assume level of significance at p < 0.00005
246 ## only significant results
247 height.sign <- gwa.height[which(gwa.height$Pc1df < 0.00005), ]
248 height.sign
249 dim(height.sign)
250
251 weight.sign <- gwa.weight[which(gwa.weight$Pc1df < 0.00005), ]
252 weight.sign
253 dim(weight.sign)
254
255
256 ##  create a vector with SNP names and calculate LOG of p-values
257 # create a vector with the names of the traits
258
259 height.name <- as.matrix(rep('height',
260                              length(colnames(mydata.clean@gtdata))))
261 weight.name <- as.matrix(rep('weight',
262                              length(colnames(mydata.clean@gtdata))))
263 trait <- rbind(height.name, weight.name)
264
265 # create a vector with the SNP names for each trait
266 snp.height <- as.matrix(rownames(gwa.height))
267 snp.weight <- as.matrix(rownames(gwa.height))
268 SNP <- rbind(snp.height, snp.weight)
269
270 my.gwa <- rbind(gwa.height, gwa.weight)
271 my.gwa <- cbind(my.gwa, trait)
272 my.gwa <- cbind(SNP, my.gwa)
273 head(my.gwa)
274 tail(my.gwa)
275
276
277 ## calculate -(log10(P-values))
278 for (i in 1: length(my.gwa)){
279       LOG <- -(log10(my.gwa$Pc1df))
280 }
281
282 my.gwa <- cbind(my.gwa, LOG)
```

```
head(my.gwa)


## save the results as a .csv file for further analysis
# manhattan plots can be done with the library(qqman)

write.csv(my.gwa, "my.gwa.csv", row.names=F, col.names=T)
#########################################################################


## check variances and heritability
# h2an: A list supplied by the nlm minimisation routine. Of particular
#interest are elements
# "estimate" containing parameter maximal likelihood estimates (MLEs)
# (order: mean, betas for covariates, heritability, (polygenic + residual
#variance)).
# The value of twice negative maximum log-likelihood is returned as
#h2an\$minimum.


# height
height.poly$h2an
height.poly$h2an$estimate
height.poly$h2an$estimate[4]  # heritability
height.poly$esth2  # heritability
height.poly$h2an$estimate[5]  # phenotypic variance: polygenic + residual


# example
h2.height <- height.poly$esth2  # heritability of height
h2.height
sigma2p.height <- height.poly$h2an$estimate[5]  # phenotypic variance of
#height
sigma2p.height
sigma2a.height <- h2.height * sigma2p.height
sigma2a.height
sigma2e.height <- sigma2p.height - sigma2a.height  # residual variance
sigma2e.height


# check the heritability
h2 <- sigma2a.height/(sigma2a.height+sigma2e.height)
height.poly$esth2  # heritability

# check the residual variance
sigma2e.height
var(height.poly$residualY, na.rm=TRUE)
var(height.poly$pgresidualY, na.rm=TRUE)
var(height.poly$grresidualY, na.rm=TRUE)


### calculate minor allele frequency (maf) in the whole dataset
snpSummary <- summary(mydata.clean@gtdata)
maf <- as.data.frame(pmin(snpSummary$Q.2,1.-snpSummary$Q.2))
summary(maf)
print(maf[1:10,])
rownames(maf) <- rownames(snpSummary)
```

```
colnames(maf) <- c("maf")
View(maf)

# save MAF in a txt file
write.table(maf, "maf.txt", col.names=TRUE, row.names=TRUE)
```

Example 2. Manhattan plots of the GWAS results

```
#############################################################
# Script for further analysis of GWAS results from GenABEL  #
# Results come from the "gwas_genabel_toy" R script          #
#  GWAS for height and weight was carried out                #
# R package: qqman                                          #
# Date: 22-03-2015                                          #
#############################################################

rm(list=ls(all=TRUE))
setwd("C:\\Users\\Documents")  # set your directory
library(qqman)

## upload the "my.gwa.csv"
snp <- read.table("my.gwa.csv", sep=",", header=T)
dim(snp)
head(snp)
unique(snp$trait) # View the names of the traits

# keep only the significant SNP of each trait
snp.sign <- snp[which(snp$ Pc1df < 0.00005), ]
dim(snp.sign)

colnames(snp)[1] <- "SNP"
colnames(snp)[2] <- "CHR"
colnames(snp)[3] <- "BP"
colnames(snp)[16] <- "P"
colnames(snp)


# separate the traits into different datasets
height <- snp[which(snp$trait=="height"), ]
weight <- snp[which(snp$trait=="weight"), ]


# keep only the significant SNP of each trait (assume here p < 0.00005)
height.sign <- height[which(height$P < 0.00005), ]
weight.sign <- weight[which(weight$P < 0.00005), ]
dim(height.sign)
dim(weight.sign)


height.sign[, c(1,2,3,7, 8, 9, 10, 16, 17, 18)]
weight.sign[, c(1,2,3,7, 8, 9, 10, 16, 17, 18)]

## keep only the columns of interest
sign.snp <- rbind(height.sign[, c(1,2,3,7, 8, 9, 10, 16, 17, 18)],
                  weight.sign[, c(1,2,3,7, 8, 9, 10, 16, 17, 18)])
dim(sign.snp)

## save the significant results in a txt file
write.table(sign.snp, "sign.snp.txt", col.names=T, row.names=F)

```

```
# check in which chromosomes the significant SNP are located for each
#trait
unique(height.sign$CHR)
unique(weight.sign$CHR)

as.data.frame(table(height.sign$CHR))
as.data.frame(table(weight.sign$CHR))


########################################
#                                      #
#     Q-Q plots of the p-values        #
#                                      #
########################################

## do it manually
# height
obs.height <- sort(height$P)
lobs.height <- -(log10(obs.height))
exp.height <- c(1:length(obs.height))
lexp.height <- -(log10(exp.height / (length(exp.height)+1)))


pdf("qqplot_height.pdf", width=6, height=6)
plot(c(0,7), c(0,7), col="red", lwd=3, type="l",
                     xlab="Expected (-logP)", ylab="Observed (-logP)",
                     xlim=c(0,7), ylim=c(0,7),
                     las=1, xaxs="i", yaxs="i", bty="l")
points(lexp.height, lobs.height, pch=23, cex=.4, bg="black")
dev.off()


# weight
obs.weight <- sort(weight$P)
lobs.weight <- -(log10(obs.weight))
exp.weight <- c(1:length(obs.weight))
lexp.weight <- -(log10(exp.weight / (length(exp.weight)+1)))


pdf("qqplot_weight.pdf", width=6, weight=6)
plot(c(0,7), c(0,7), col="red", lwd=3, type="l",
                     xlab="Expected (-logP)", ylab="Observed (-logP)",
                     xlim=c(0,7), ylim=c(0,7),
                     las=1, xaxs="i", yaxs="i", bty="l")
points(lexp.weight, lobs.weight, pch=23, cex=.4, bg="black")
dev.off()


## alternatively, using the "ggman" R package
library(qqman)
qq(height$P)
qq(weight$P)
```

```
pdf("QQplots_GWAS_height_weight.pdf")
par(mfrow=c(1, 2))
qq(height$P, main="height")
qq(weight$P, main="weight")
par(mfrow=c(1, 1))
dev.off()


########################################
#                                      #
#            Manhattan plots           #
#                                      #
########################################

-(log10(5e-05)) # my threshold
max(height$LOG)  # maximum y scale
# Note: chromosome X has been exluded from the plot
# CHR column should be numeric for "qqman" to work.
# Do you have 'X', 'Y', 'MT', etc? If so change to numbers

chr.height <- as.matrix(height$CHR)
chr.height[chr.height=="X"] <- 4
height$CHR <- as.numeric(chr.height)
height$CHR <- as.numeric(height$CHR)

chr.weight <- as.matrix(weight$CHR)
chr.weight[chr.weight=="X"] <- 4
weight$CHR <- as.numeric(chr.weight)
weight$CHR <- as.numeric(weight$CHR)

## manhattan plots
# manhattan plot for height
manhattan(height, col=c("black","#666666","#CC6600"),
          genomewideline=4.30103,
          suggestiveline=FALSE,
          highlight=height.sign$SNP,
          main="a) GWAS for height")

# manhattan plot for weight
manhattan(weight, col=c("black","#666666","#CC6600"),
          genomewideline=4.30103, suggestiveline=FALSE,
          highlight=weight.sign$SNP,
          main="b) GWAS for weight")


# plot both manhattan plots in one graph
par(mfrow=c(2, 1))
manhattan(height, col=c("black","#666666","#CC6600"),
          genomewideline=4.30103,
          suggestiveline=FALSE, highlight=height.sign$SNP,
          main="a) GWAS for height")
manhattan(weight, col=c("black","#666666","#CC6600"),
          genomewideline=4.30103,
          suggestiveline=FALSE,
          highlight=weight.sign$SNP,
          main="b) GWAS for weight")
par(mfrow=c(1, 1))
```

```
167
168 ## save manhattan plots as pdf
169 # change height and width appropriately to fit your demands
170 pdf("manhattan.pdf", height = 10, width = 20)
171 par(mfrow=c(2, 1))
172 manhattan(height, col=c("black","#666666","#CC6600"),
173           genomewideline=4.30103, suggestiveline=FALSE,
174           highlight=height.sign$SNP,
175           main="a) GWAS for height")
176 manhattan(weight, col=c("black","#666666","#CC6600"),
177           genomewideline=4.30103, suggestiveline=FALSE,
178           highlight=weight.sign$SNP,
179           main="b) GWAS for weight")
180 dev.off()
181
182
183 # zooming in into Chromosomes
184 #### Chromosome 6 ###
185 chr2.height <- subset(height, CHR==2)
186 manhattan(chr2.height, genomewideline=4.30103,
187           suggestiveline=FALSE,
188           highlight=height.sign$SNP,
189           main="GWAS for height")
190
191 # save as a pdf
192 pdf("Chr2_height.pdf", height = 8, width = 12)
193 manhattan(chr2.height, genomewideline=4.30103,
194           suggestiveline=FALSE,
195           highlight=height.sign$SNP,
196           main="GWAS for height")
197 dev.off()
```

Example 3. Single marker GWAS with GenABEL (GRAMMAR-GC). Summarizes basic GWAS results when a large number of phenotypes need to be analyzed.

```
#######################################################
# Script for fast GWAS with GenABEL                   #
# R package: GenABEL                                  #
#     Method: GRAMMAR-GC                              #
#     GenABEL data: ge03d2                            #
#     Analysis: GWAS for height, weight and bmi       #
#               using a loop                          #
# Date: 02-04-2016                                    #
#######################################################

rm(list=ls(all=TRUE))
library("GenABEL")

## Upload GenABEL data
require(GenABEL.data)
data(ge03d2)
# keep subset of the data to make the example faster
my.data <- ge03d2[seq(from=1,to=nids(ge03d2),by=2),
                  seq(from=1,to=nsnps(ge03d2),by=2)]

## quality check
mydata.qc <- check.marker(my.data,call=0.95,
                          perid.call=0.95,
                          extr.call = 0.1,
                          extr.perid.call = 0.1,
                          maf=0.005,
                          p.lev=0,
                          ibs.mrk =0 )

## subset clean data
mydata.clean <- my.data[mydata.qc$idok,mydata.qc$snpok]

## genomic relationship
mydata.gkin <- ibs(mydata.clean[,autosomal(mydata.clean)],weight="freq")




######################
# GRAMMAR-GC GWAS

my.list <- colnames(mydata.clean@phdata[, c(5,6,8)] )
nsign.snp <- NULL
nsign.chrom<- NULL
res <- NULL
r <- 0
my.h2 <- NULL
highest.sign <- NULL
```

```
## start of the loop
for (x in c(5,6, 8)){   # note that if replace x with i does not work!!
      r <- r+1
      results <- 0
      res.sign <- 0
      h2.prelim <- 0
      min.pvalue <- 0

      m1 <- polygenic(mydata.clean@phdata[, x] ~ sex + age,
                           kin=mydata.gkin, data=mydata.clean)
      m2 <- grammar(m1,data=mydata.clean, method="gamma")
      results <- data.frame(SNP = as.character(rownames(m2)),
                        CHR =  as.character(results(m2)$Chromosome),
                        BP = results(m2)$Position,
                        P = results(m2)$Pc1df,
                        LOG=-log10(results(m2)$Pc1df),
                        effB = results(m2)$effB,
                        Trait= my.list[r])
      results <- results[order(results$P), ]
      res.sign <- results[which(results$P < 0.005),]
      res <- rbind(res,res.sign)
      nsnp <- data.frame(no.sign.SNP = nrow(res.sign),
                  Trait= my.list[r])
      nsign.snp <- rbind(nsign.snp, nsnp)
      nchr <- data.frame(no.sign.CHR = length(unique(res.sign$CHR)),
                  Trait= my.list[r])
      nsign.chrom <- rbind(nsign.chrom, nchr)
      min.pvalue <- data.frame (min.Pvalue = min(results$P),
                                max.log = max(results$LOG),
                                CHR = results$CHR[which(results$LOG ==
max(results$LOG))],
                                BP = results$BP[which(results$LOG ==
max(results$LOG))],
                                Trait= my.list[r])
      highest.sign <- rbind(highest.sign, min.pvalue)
      h2.prelim <- data.frame(h2 = m1$esth2,
                        sigma2p = tail(m1$h2an$estimate, n=1),
                        sigma2a =  (m1$esth2) * (tail(m1$h2an$estimate,
n=1)),
                        Trait= my.list[r])
      my.h2 <- rbind(my.h2,h2.prelim)
}

## end of the loop


print(res)
print(nsign.snp)
print(nsign.chrom)
print(highest.sign)
print(my.h2)
```

```
### calculate SNP frequencies (p,q) - minor allele frequency (maf) and
high allele freq (haf)
snpSummary <- summary(mydata.clean@gtdata)
my.freq <- data.frame(
            SNP = rownames(snpSummary),
            p = snpSummary$Q.2,
            q = 1. - snpSummary$Q.2,
            maf = pmin(snpSummary$Q.2,1.-snpSummary$Q.2),
            haf = pmax(snpSummary$Q.2,1.-snpSummary$Q.2))
head(my.freq)
tail(my.freq)
summary(my.freq)


## add maf frequencies
my.gwas <- merge(res, my.freq, by="SNP")
my.gwas <- my.gwas[order(my.gwas$Trait, my.gwas$P), ]

# some checks
head(my.gwas)
head(res)
my.freq[which(my.freq$SNP== "rs3436694" |
              my.freq$SNP== "rs3175719" |
              my.freq$SNP== "rs1801282" |
              my.freq$SNP== "rs4277955" |
              my.freq$SNP== "rs1227627"), ]



## calculate snp variance

## start of the loop
for(l in 1:nrow(my.gwas)){
snp.var <- data.frame(
            SNP = my.gwas$SNP,
            snp.va = 2 * my.gwas$p * my.gwas$q * (my.gwas$effB)^2,
            Trait = my.gwas$Trait
      )
}
## end of the loop


## GWAS for height and weight manually - compare results

height.poly <- polygenic(height ~ sex + age,
                         kin=mydata.gkin,
                         data=mydata.clean)
height.grammar <- grammar(height.poly,data=mydata.clean,method="gamma")
weight.poly <- polygenic(weight ~ sex + age,
                         kin=mydata.gkin,
                         data=mydata.clean)
weight.grammar <- grammar(weight.poly,data=mydata.clean,method="gamma")
bmi.poly <- polygenic(bmi ~ sex + age,
                      kin=mydata.gkin,
                      data=mydata.clean)
bmi.grammar <- grammar(bmi.poly,data=mydata.clean,method="gamma")

```

```
# get gwas all results
gwa.height <- descriptives.scan(height.grammar,
                                top=length(colnames(mydata.clean@gtdata)),
                                sort="Pc1df")
gwa.weight <- descriptives.scan(weight.grammar,
                                top= length(colnames(mydata.clean@gtdata)),
                                sort="Pc1df")
gwa.bmi <- descriptives.scan(bmi.grammar,
                                top= length(colnames(mydata.clean@gtdata)),
                                sort="Pc1df")


## keep only significant results
# assume level of significance at p < 0.005
## only significant results
height.sign <- gwa.height[which(gwa.height$Pc1df < 0.005), ]
height.sign
dim(height.sign)

weight.sign <- gwa.weight[which(gwa.weight$Pc1df < 0.005), ]
weight.sign
dim(weight.sign)

bmi.sign <- gwa.bmi[which(gwa.bmi$Pc1df < 0.005), ]
bmi.sign
dim(bmi.sign)

## save the results
write.csv(my.gwas, "my.gwas.csv", row.names=F, col.names=T)
write.table(nsign.snp, "nsign_snp.txt", row.names=F, col.names=T)
write.table(nsign.chrom, "nsign_chrom.txt", row.names=F, col.names=T)
write.table(highest.sign, "highest_signals.txt",
            row.names=F, col.names=T)
write.table(my.h2, "heritability.txt", row.names=F, col.names=T)
write.table(my.freq, "frequencies.txt", col.names=TRUE, row.names=TRUE)
write.table(snp.var, "snp_variance.txt", col.names=TRUE, row.names=TRUE)
```

Example 4. Gene-set enrichment and pathway analysis

a) Mapping SNP to genes

```
rm(list=ls(all=TRUE))
setwd("C:\\Users\\Documents")  # set your directory

library(biomaRt)
database = useMart("ensembl")
genome = useDataset("btaurus_gene_ensembl", mart = database)
gene = getBM(c("ensembl_gene_id", "entrezgene",
               "start_position", "end_position",
               "chromosome_name", "hgnc_symbol"), mart = genome)


ndata <- read.table(myfile, header = TRUE)
sign_chr <- sort(unique(ndata$Chromosome))
length(sign_chr)
sign_chr


#3# from SNPs to Genes: create a database with those genes that have one
#or more SNPs in a 15kb window
plus = 1500

Name = character()
Chromosome = numeric()
Location = numeric()
Gene = numeric()
EntrezID = numeric()
SYMBOL = character()
m = 1

 for(k in 1:29) # check all 29 chromosomes
{

    SNP = subset(ndata, ndata$Chromosome == k)
    genes = subset(gene, gene$chromosome_name == k)

    for(i in 1:length(SNP$Chromosome))
    {
        for(j in 1:length(genes$chromosome_name))
        {

            if(genes$start_position[j] <= (plus + SNP$Location[i]) &
(SNP$Location[i] - plus) <= genes$end_position[j])
            {
                Name[m] = as.character(SNP$SNP[i])
                Chromosome[m] = SNP$Chromosome[i]
                Location[m] = SNP$Location[i]
                Gene[m] = genes$ensembl_gene_id[j]
                EntrezID[m] = genes$entrezgene[j]
                SYMBOL[m] = as.character(genes$hgnc_symbol[j])
                m = m+1
            }
        }
    }
}
```

```
55
56  SNPtoGENES = data.frame(Name,Chromosome,Location,Gene,EntrezID, SYMBOL)
57  dim(SNPtoGENES)
58
59  write.table(SNPtoGENES,”mygenes.txt” , col.names=TRUE, row.names=FALSE)
```

Example 4. Gene-set enrichment and pathway analysis

b) Querying the GO and KEGG databases

```
1  ######################################################################
2  # Gene enrichmnet for the MCP, CF, CY, REC                           #
3  # from the GWAS results                                              #
4  # GO and KEGG databases                                              #
5  # Rscript was based on the goseq:                                    #
6  # snp_to_genes: all genes in my SNP chip                             #
7  # snp_to_sign_genes: based on GWAS significant SNP                   #
8  # snp_to_sign_genes_pval: based on pval<=0.05 from GWAS              #
9  ######################################################################
10
11
12 rm(list=ls(all=TRUE))
13 setwd("C:\\Users\\Documents")  # set your directory
14 source("https://bioconductor.org/biocLite.R")
15 # biocLite("org.Bt.eg.db")
16 # biocLite("GOstats")
17 # biocLite("MeSH.Bta.eg.db")
18 # biocLite("meshr")
19 # biocLite("MeSH.db")
20 # biocLite("KEGG.db")
21 # biocLite("annotate")
22
23
24 library("org.Bt.eg.db")
25 library("GOstats")
26 library(meshr)
27 library(MeSH.db)
28 library(MeSH.Bta.eg.db)
29 library(goseq)
30 library(GO.db)
31 library(KEGG.db)
32 library(annotate)
33
34
35
36 ########################
37 #1# upload list of genes
38 all_genes <- read.table("snp_to_genes.txt", header=TRUE)
39 all_genes <- all_genes[, 4]
40 length(all_genes)
41 length(unique(all_genes))
42 head(all_genes)
43
44 #1a# remove duplicated values
45 total.genes <- all_genes[!duplicated(all_genes)]
46 length(total.genes)
47 head(total.genes)
48
49
50 #2# upload significant genes
51 sign_genes <- read.table("snp_to_sign_genes_pval.txt", header = TRUE)
52 sign_genes <- sign_genes[, 4]
53 length(sign_genes)
54 length(unique(sign_genes))
```

```
head(sign_genes)


#2a# remove duplicated values ensembl Gene ID
sig.genes <- sign_genes[!duplicated(sign_genes)]
length(sig.genes)
head(sig.genes)


################## END OF DATA PREPARATION ##########################


assayed.genes = array(total.genes) ## total.genes is a vector with ALL
#the genes evaluated

de.genes = array(sig.genes) ## sig.genes is a vector with significant
#genes


gene.vector = as.integer(assayed.genes%in%de.genes)
names(gene.vector) = assayed.genes

pwf = nullp(gene.vector, "bosTau3", "ensGene", plot.fit = FALSE)


#1# GO Biological Process - using Hypergeometric distribution
## Fisher exact #test (hypergeometric)
GO.hiper.bp <- goseq(pwf, "bosTau3", "ensGene",
                     method = "Hypergeometric",
                     test.cats="GO:BP",
                     use_genes_without_cat = TRUE)   dim(GO.hiper.bp)

## Consider only terms with genes > 5 and < 500
nGO.hiper.bp = GO.hiper.bp[(GO.hiper.bp$numInCat <= 1000 &
                            GO.hiper.bp$numInCat >= 10),]
dim(nGO.hiper.bp)

## Raw P-value < 0.01
enriched.GO.bp = nGO.hiper.bp[nGO.hiper.bp$over_represented_pvalue <=
0.01, ]

## fdr correction
enriched.GO.bp =
nGO.hiper.bp[p.adjust(nGO.hiper.bp$over_represented_pvalue,
                      method="fdr")<.05, ]
dim(enriched.GO.bp)
head(enriched.GO.bp)


#2# GO Molecular function - using Hypergeometric distribution
## Fisher exact test (hypergeometric)
GO.hiper.mf <- goseq(pwf, "bosTau3", "ensGene",
                     method = "Hypergeometric",
                     test.cats="GO:MF",
                     use_genes_without_cat = TRUE)
dim(GO.hiper.mf)

```

```
## Consider only terms with genes > 5 and < 500
nGO.hiper.mf = GO.hiper.mf[(GO.hiper.mf$numInCat <= 1000 &
                           GO.hiper.mf$numInCat >= 10),]
dim(nGO.hiper.mf)

## Raw P-value < 0.01
enriched.GO.mf = nGO.hiper.mf[nGO.hiper.mf$over_represented_pvalue <=
0.01, ]

## fdr correction
enriched.GO.mf =
nGO.hiper.mf[p.adjust(nGO.hiper.mf$over_represented_pvalue,
             method="fdr")<.05, ]
dim(enriched.GO.mf)
head(enriched.GO.mf)


#3# GO Cellular Component - using Hypergeometric distribution
## Fisher exact test (hypergeometric)
GO.hiper.cc <- goseq(pwf, "bosTau3", "ensGene",
                     method = "Hypergeometric",
                     test.cats="GO:CC",
                     use_genes_without_cat = TRUE)
dim(GO.hiper.cc)

## Consider only terms with genes > 5 and < 500
nGO.hiper.cc = GO.hiper.cc[(GO.hiper.cc$numInCat <= 1000 &
                           GO.hiper.cc$numInCat >= 10),]
dim(nGO.hiper.cc)

## Raw P-value < 0.01
enriched.GO.cc = nGO.hiper.cc[nGO.hiper.cc$over_represented_pvalue <=
0.01, ]

## fdr correction
enriched.GO.cc =
nGO.hiper.cc[p.adjust(nGO.hiper.cc$over_represented_pvalue,
method="fdr")<.05, ]

dim(enriched.GO.cc)
head(enriched.GO.cc)


#4# KEGG enrichment analysis - using Hypergeometric distribution
## Fisher exact test (hypergeometric)
GO.hiper.kegg <- goseq(pwf, "bosTau3", "ensGene",
                       method = "Hypergeometric",
                       test.cats="KEGG",
                       use_genes_without_cat = TRUE)
dim(GO.hiper.kegg)

## Consider only terms with genes > 5 and < 500
nGO.hiper.kegg = GO.hiper.kegg[(GO.hiper.kegg$numInCat <= 1000 &
GO.hiper.kegg$numInCat >= 10),]
dim(nGO.hiper.kegg)

## Raw P-value < 0.01
```

```
enriched.GO.kegg = nGO.hiper.kegg[nGO.hiper.kegg$over_represented_pvalue
<= 0.01, ]

## fdr correction
enriched.GO.kegg =
nGO.hiper.kegg[p.adjust(nGO.hiper.kegg$over_represented_pvalue,
method="fdr")<.05, ]
dim(enriched.GO.kegg)
head(enriched.GO.kegg)

############# END OF GO-KEGG gene enrichment analysis ##################
```

Example 4. Gene-set enrichment and pathway analysis

c) Identification of genes in the significant GO terms

```
#########################################################################
# Script for identifying the genes involved in each GO term             #
# Traits and GO categories have been previously identified               #
# 1. We re-run the pathway analysis, but only for the traits we found   #
# significance after False Discovery Rate (FDR) correction              #
# 2. We identify the genes                                              #
# Note: This script is for only 1 trait analysis!                       #
#########################################################################

rm(list=ls(all=TRUE))
setwd("C:\\Users\\Documents")  # set your directory
source("https://bioconductor.org/biocLite.R")
library("org.Bt.eg.db")
library(goseq)
library(GO.db)
# biocLite("KEGG.db")
library(KEGG.db)
library(annotate)

########################
#1# upload list of genes
all_genes <- read.table("C:\\Users\\Documents\\background_genes.txt",
header=TRUE)
all_genes <- all_genes[, 4]

#1a# remove duplicated values
total.genes <- all_genes[!duplicated(all_genes)]

#2# upload significant genes
sign_genes_pre <- read.table(mygenes.txt, header = TRUE)
sign_genes_pre1 <- sign_genes_pre[!duplicated(sign_genes_pre$Gene),]
sign_genes <- sign_genes_pre[, 4]

#2a# remove duplicated values ensembl Gene ID
sig.genes <- sign_genes[!duplicated(sign_genes)]


################# END OF DATA PREPARATION ##########################

assayed.genes = array(total.genes)     ## total.genes is a vector with
#ALL the genes evaluated
de.genes = array(sig.genes)  ## sig.genes is a vector with significant
#genes

gene.vector = as.integer(assayed.genes%in%de.genes)
names(gene.vector) = assayed.genes

pwf = nullp(gene.vector, "bosTau3", "ensGene", plot.fit = FALSE)
```

```
55
56  #3# GO Biological Process, Cellular Component, Molecular Function -
57  #using Hypergeometric distribution
58  ## Fisher exact test (hypergeometric)
59  GO.hiper <- goseq(pwf, "bosTau3", "ensGene",
60                    method = "Hypergeometric",
61                    test.cats=c("GO:BP", "GO:CC", "GO:MF"),
62                    use_genes_without_cat = TRUE)
63
64  #3a# Consider only terms with genes > 10 and < 1000
65  nGO.hiper = GO.hiper[(GO.hiper$numInCat <= 1000 & GO.hiper$numInCat >=
66  10),]
67
68  #3b# fdr correction
69  enriched.GO.fdr = nGO.hiper[p.adjust(nGO.hiper$over_represented_pvalue,
70  method="fdr")<.05, ]
71  enriched.GO.fdr <- enriched.GO.fdr[order(enriched.GO.fdr$ontology),]
72
73  ##########
74  #4# Identify the genes in the significant GO terms
75
76  go = getgo(names(gene.vector), "bosTau3", "ensGene")
77  my.list <- enriched.GO.fdr$category # NOTE: if we make it as a list,
78  #then the "category" column in the results table is not working
79
80  for (k in 1:length(enriched.GO.fdr$category)){
81        r <- r+1
82        genes <- 0
83        GOterm = enriched.GO.fdr$category[k] ## your favorite GO term
84        genes = as.character(); m = 1
85        sgenes = as.character(); n = 1
86
87        for(i in 1:length(go)){
88            if(length(go[[i]]) > 0){
89                for(j in 1:length(go[[i]])){
90                    if(go[[i]][j] == GOterm){
91                    genes[m] = names(go[i])
92                        m = m + 1
93                       }
94                  }
95             }
96        }
97  go.fdr.cat <- 0
98  go.fdr.cat <- as.data.frame(sort(genes[genes%in%sig.genes]))
99  write.table(go.fdr.cat, "mytrait_sign_go_genes.txt",
100                         col.names=TRUE, row.names=FALSE)
101
    ########### END ###########
```

Example 4. Gene-set enrichment and pathway analysis

d) Identification of genes in the significant KEGG categories

```
#########################################################################
# Script for identifying the genes involved in each KEGG Category      #
# Traits and KEGG categories have been previously identified           #
# 1. We re-run the pathway analysis, but only for the traits we found  #
# significance after False Discovery Rate (FDR) correction             #
# 2. We identify the genes                                             #
# Note: This script is for only 1 trait analysis!                      #
#########################################################################

rm(list=ls(all=TRUE))
setwd("C:\\Users\\Documents\\")
source("https://bioconductor.org/biocLite.R")
library("org.Bt.eg.db")
library(goseq)
library(GO.db)
# biocLite("KEGG.db")
library(KEGG.db)
library(annotate)


#1# upload list of genes
all_genes <- read.table("C:\\Users\\Documents\\background_genes.txt",
header=TRUE)
all_genes <- all_genes[, 4]

#1a# remove duplicated values
total.genes <- all_genes[!duplicated(all_genes)]

#2# upload significant genes
sign_genes_pre <- read.table(mygenes.txt, header = TRUE)
sign_genes_pre1 <- sign_genes_pre[!duplicated(sign_genes_pre$Gene),]
sign_genes <- sign_genes_pre[, 4]

#2a# remove duplicated values ensembl Gene ID
sig.genes <- sign_genes[!duplicated(sign_genes)]

################## END OF DATA PREPARATION ##########################


assayed.genes = array(total.genes)     ## total.genes is a vector with
#ALL the genes evaluated
de.genes = array(sig.genes)  ## sig.genes is a vector with significant
#genes

gene.vector = as.integer(assayed.genes%in%de.genes)
names(gene.vector) = assayed.genes

pwf = nullp(gene.vector, "bosTau3", "ensGene", plot.fit = FALSE)
```

```
55
56  #3# KEGG enrichment analysis - using Hypergeometric distribution
57  #3a# Fisher exact test (hypergeometric)
58  KEGG.hiper.kegg <- goseq(pwf, "bosTau3", "ensGene",
59                           method = "Hypergeometric",
60                           test.cats="KEGG",
61                           use_genes_without_cat = TRUE)
62
63  #3b# Consider only terms with genes > 10 and < 1000
64  nKEGG.hiper.kegg = KEGG.hiper.kegg[(KEGG.hiper.kegg$numInCat <= 1000 &
65  KEGG.hiper.kegg$numInCat >= 10),]
66
67  #3c# fdr correction
68  enriched.KEGG.fdr =
69  nKEGG.hiper.kegg[p.adjust(nKEGG.hiper.kegg$over_represented_pvalue,
70  method="fdr")<.05, ]
71
72  #################
73  #4# Identify the genes in the significant KEGG categories
74  kegg = getgo(names(gene.vector),
75                   "bosTau3",
76                   "ensGene",
77                    fetch.cats = c("KEGG"))
78
79  my.list <- enriched.KEGG.fdr$category # NOTE: if we make it as a list,
80  #then the "category" column in the results table is not working
81
82  for (k in 1:length(enriched.KEGG.fdr$category)){
83        r <- r+1
84        genes <- 0
85        KEGGterm = enriched.KEGG.fdr$category[k] ## your favorite KEGG
86  #pathway
87        genes = as.character(); m = 1
88        sgenes = as.character(); n = 1
89
90        for(i in 1:length(kegg)){
91            if(length(kegg[[i]]) > 0){
92                for(j in 1:length(kegg[[i]])){
93                    if(kegg[[i]][j] == KEGGterm){
94                        genes[m] = names(kegg[i])
95                        m = m + 1
96                    }
97                }
98            }
99        }
100 length(genes);genes
101 table(genes%in%sig.genes)
102 sort(genes[genes%in%sig.genes])
103
104 kegg.fdr.cat <- as.data.frame(sort(genes[genes%in%sig.genes]))
105 write.table(kegg.fdr.cat, "mytrait_sign_kegg_genes.txt",
106                           col.names=TRUE, row.names=FALSE)
107
108 ########### END ###########
```

Example 5. Factor analysis

```
#########################################################################
# Factor Analysis Models toy example                                    #
# Statistical Analysis:                                                 #
#     1. Exploratory Factor Analysis (EFA)- orthogonal factor rotation  #
#                                                                       #
# R package used: "psych" for EFA                                       #
# Author: Christos Dadousis                                             #
# Date: 11-01-2016                                                      #
#########################################################################


rm(list=ls())
require(graphics)
library(pastecs)
require(psych)
require(GPArotation)
require(lavaan)
library(semPlot)
library(calibrate)
library(MVN)
setwd("C:/Users/Documents/")


############################ Create the dataset ##########################
# number of individuals: 200                                            #
# number of exploratory variables: 11                                   #
# two variables (x1-x3) with low correlation: ~0.2                      #
# two variables (x2-x4)with high correlation: ~0.8                      #
# seven random variables                                                #
# All variables drawn from normal distribution                          #
#########################################################################

# create two variables with low correlation (~0.2) and
#        two with high correlation (~0.8)

correlatedValue = function(x, r){
  r2 = r**2
  ve = 1-r2
  SD = sqrt(ve)
  e  = rnorm(length(x), mean=0, sd=SD)
  y  = r*x + e
  return(y)
}

set.seed(4444)
x1 = rnorm(2000)
x2 = rnorm(2000)
x3 = correlatedValue(x=x1, r=.2)
x4 = correlatedValue(x=x2, r=.8)

# check correlations
cor(x1,x3)
cor(x2,x4)

# create seven random variables
```

```
my.x <- as.matrix(replicate(7, rnorm(200)))

# merge the data in one matrix

X <- cbind(x1, x2, x3, x4, my.x)
colnames(X) <- paste("x", 1:ncol(X), sep="")
rownames(X) <- paste("Sample_", 1:nrow(X), sep="")
dim(X)
head(X)
cor(X)

################## END of Data Preparation ######################


#1# Descriptive statistics and general checks

#1a# descr. stat
round(stat.desc(X),3)
describe(X)

#1b# correlation checks
pairs.panels(X)
cor.plot(cor(X))
corr.test(X)

#1c# error bar plots
error.bars(X)

#1d# potential outliers detection
outlier(X)

#1e# check of normality with Shapiro-Wilk normality test
shapiro.test(X[, 1])

#1f# Distributions
pdf("Distributions_rawdata.pdf", height = 12, width = 8)
par(mfrow=c(4,3))
hist(X[, 1], freq=F, ylab="Probability", xlab="x1", main="")
curve(dnorm(x, mean(X[, 1], na.rm = T), sd(X[, 1], na.rm = T)), add=T,
col="blue")
hist(X[, 2], freq=F, ylab="Probability", xlab="x2", main="")
curve(dnorm(x, mean(X[, 2], na.rm = T), sd(X[, 2], na.rm = T)), add=T,
col="blue")
hist(X[, 3], freq=F, ylab="Probability", xlab="x3", main="")
curve(dnorm(x, mean(X[, 3], na.rm = T), sd(X[, 3], na.rm = T)), add=T,
col="blue")
hist(X[, 4], freq=F, ylab="Probability", xlab="x4", main="")
curve(dnorm(x, mean(X[, 4], na.rm = T), sd(X[, 4], na.rm = T)), add=T,
col="blue")
hist(X[, 5], freq=F, ylab="Probability", xlab="x5", main="")
curve(dnorm(x, mean(X[, 5], na.rm = T), sd(X[, 5], na.rm = T)), add=T,
col="blue")
hist(X[, 6], freq=F, ylab="Probability", xlab="x6", main="")
curve(dnorm(x, mean(X[, 6], na.rm = T), sd(X[, 6], na.rm = T)), add=T,
col="blue")
hist(X[, 7], freq=F, ylab="Probability", xlab="x7", main="")

```

```
curve(dnorm(x, mean(X[, 7], na.rm = T), sd(X[, 7], na.rm = T)), add=T,
col="blue")
hist(X[, 8], freq=F, ylab="Probability", xlab="x8", main="")
curve(dnorm(x, mean(X[, 8], na.rm = T), sd(X[, 8], na.rm = T)), add=T,
col="blue")
hist(X[, 9], freq=F, ylab="Probability", xlab="x9", main="")
curve(dnorm(x, mean(X[, 9], na.rm = T), sd(X[, 9], na.rm = T)), add=T,
col="blue")
hist(X[, 10], freq=F, ylab="Probability", xlab="x10", main="")
curve(dnorm(x, mean(X[, 10], na.rm = T), sd(X[, 10], na.rm = T)), add=T,
col="blue")
hist(X[, 11], freq=F, ylab="Probability", xlab="x11", main="")
curve(dnorm(x, mean(X[, 11], na.rm = T), sd(X[, 11], na.rm = T)), add=T,
col="blue")
dev.off()


#1f# check multivariate normality

mvn <- mardiaTest(X, qqplot = TRUE)
print(mvn)


################## check for multicollinearity ######################
#2# check for multicollinearity
#2a# determinant of the correlation matrix
#(cut of value: 0.00001; problem with lower values)
det(cor(X))


#2b# VIF>=10 problematic. Corresponds to R^2>=0.90
my.vif <- NULL
for (i in 1:ncol(X)){
y <- X[, i]
x <- as.data.frame(X[, -c(i)])
my.model <- lm(y ~ . , data= x)
my.res <- 0
my.res <- data.frame(r2 = summary(my.model)$r.squared,
                     r2.adjusted = summary(my.model)$adj.r.squared,
                     vif = 1/(1-summary(my.model)$r.squared)
)
my.vif <- rbind(my.vif, my.res)
}
rownames(my.vif) <- colnames(X)
my.vif
my.vif$vif>10


# plot of vif values
png("multicollinearity.png")
par(mfrow=c(1, 1), mar=c(5.1, 5.1, 4.1, 2.1)+0.1)
plot(my.vif[, 1],
      col = ifelse(my.vif[, 1] == 1, "red",
            ifelse(my.vif[, 1] >0.9, "orange", "black")),
      pch = ifelse(my.vif[, 1] == 1, 16,
            ifelse(my.vif[, 1] >0.9, 16, 1)),
      main = "Multicolinearity check",
```

```
      xlab = "Trait",
      ylab = expression("R"^2),
      cex = 2,
      cex.main = 2,
      cex.axis= 1.5,
      cex.lab = 2,
      las=1
)
abline(h=0.9, col="red")
textxy(X = seq(1:nrow(my.vif)), Y=my.vif[, 1], labs=rownames(my.vif),
cex=1.3)
dev.off()


#2c#
#########################################################################
# based on eigenvalues                                                  #
# exat linear dependense: then one or more eigenvalues will be zero    #
# near linear dependence: some eigenvalues will be very small. Large    #
# variances                                                             #
# condition number k                                                    #
#     - k<100: no serious problem with multicollinearity                #
#     -100<=k<1000: moderate to strong problem                          #
#     k>1000: severe problem                                            #
#########################################################################


pca <- eigen(cor(X))
pca$values
write.table(pca$values, "eigenvalues.txt", col.names=FALSE,
row.names=FALSE) # save the eigenvalues
k <- max(pca$values)/min(pca$values)
k

################ END of multicollinearity check ########################


#3# calculation of Kaiser Measure of Sampling Adequacy (Kaiser MSA)
KMO(X)


#########################################################################
###  Exploratory Factor Analysis with ML Estimation and varimax rotation
# Factors are forced to be un-correlated
# We want:
#     Highest Prop. Var (debatable)
#     Lowest RMSR
#     Lowest MLE chi square
#     MLE chi square > 0.05
#     TLI > 0.95
#     RMSEA < 0.08. (if RMSEA > 0.1 is not good)
#     Lowest BIC
# For more details see here:
#http://www.inside-r.org/packages/cran/psych/docs/fa and
#https://cran.r-project.org/web/packages/psych/vignettes/overview.pdf


```

```
## create a loop for fast model check
## Start of the loop


nf=5 # number of factors to be tested
varimax.table <- NULL
r <- 0

for (i in seq(1,nf)){
  r<-r+1
  if (i%%10<1){
    cat(paste('Round: ',r,'\n'))
  }

MLvari <- fa(X,nfactors=i,rotate="varimax",fm="minres")
results <- 0
results <- data.frame(  Model=i,
                        No.Factors = MLvari$factors,
                        RMSR = round(MLvari$rms, 2),
                        cRMSR = round(MLvari$crms, 2),
                        TLI = round(MLvari$TLI, 3),
                        RMSEA= round(MLvari$RMSEA[1], 3),
                        BIC =  MLvari$BIC,
                        eBIC = MLvari$EBIC,
                        chisq = MLvari$STATISTIC,
                        Pval = MLvari$PVAL,
                        echisq = MLvari$chi
)

varimax.table <- rbind(varimax.table, results)
}
## END of the fast model check


## Test for the number of factors in your data using
# parallel analysis (fa.parallel)  or
# Very Simple Structure (vss)
varimax.table
fa.parallel(X)
vss.vari <- vss(X, n=nf, rotate="varimax",fm="minres")
vss.vari


## identify the models with best fitting manually
min(varimax.table$RMSR)
min(varimax.table$cRMSR)
min(varimax.table$RMSEA)
min(varimax.table$BIC)
min(varimax.table$eBIC)
max(varimax.table$TLI)
varimax.table[order(varimax.table$RMSR), ]
varimax.table[order(varimax.table$cRMSR), ]
varimax.table[order(varimax.table$RMSEA), ]
varimax.table[order(varimax.table$BIC), ]
varimax.table[order(varimax.table$eBIC), ]
varimax.table[order(-varimax.table$TLI), ]

```

```
284
285 ## identify the models with best fitting with graphs
286 # a png file will be saved in your directory
287 # after running all the lines below till
288 # the END of plots
289 png("ModelCheck_varimax.png", 1000, 1200)
290 par(mfrow=c(4, 3), mar=c(5.1, 5.1, 4.1, 2.1)+0.1)
291 plot(fa.parallel(X),
292       cex.main=2,
293       cex.axis=2,
294       cex.lab=2)
295 plot(vss.vari,
296       cex.main=2,
297       cex.axis=2,
298       cex.lab=2)
299 plot(varimax.table$RMSR,
300       col = ifelse(varimax.table$RMSR == min(varimax.table$RMSR),
301 "blue", "black"),
302       pch = ifelse(varimax.table$RMSR == min(varimax.table$RMSR), 8, 1),
303       cex = ifelse(varimax.table$RMSR == min(varimax.table$RMSR), 3, 2),
304       xlab = "Number of Factors",
305       ylab = "RMSR",
306       main= "RMSR",
307       cex.main=2,
308       cex.axis=2,
309       cex.lab=2)
310 textxy(X=varimax.table$Model, Y=varimax.table$RMSR,
311 labs=varimax.table$No.Factors, cex=1.5)
312 plot(varimax.table$cRMSR,
313       col = ifelse(varimax.table$cRMSR == min(varimax.table$cRMSR),
314 "blue", "black"),
315       pch = ifelse(varimax.table$cRMSR == min(varimax.table$cRMSR), 8,
316 1),
317       cex = ifelse(varimax.table$cRMSR == min(varimax.table$cRMSR), 3,
318 2),
319       xlab = "Number of Factors",
320       ylab = "cRMSR",
321       main= "cRMSR",
322       cex.main=2,
323       cex.axis=2,
324       cex.lab=2)
325 textxy(X=varimax.table$Model, Y=varimax.table$cRMSR,
326 labs=varimax.table$No.Factors, cex=1.5)
327 plot(varimax.table$RMSEA,
328       col = ifelse(varimax.table$RMSEA == min(varimax.table$RMSEA),
329 "blue", ifelse(varimax.table$RMSEA >.1, "red", "black")),
330       pch = ifelse(varimax.table$RMSEA == min(varimax.table$RMSEA), 8,
331 ifelse(varimax.table$RMSEA >.1, 16, 1)),
332       cex = ifelse(varimax.table$RMSEA == min(varimax.table$RMSEA), 3,
333 2),
334       xlab = "Number of Factors",
335       ylab = "RMSEA",
336       main= "RMSEA",
337       cex.main=2,
338       cex.axis=2,
339       cex.lab=2)
340
```

```
textxy(X=varimax.table$Model, Y=varimax.table$RMSEA,
labs=varimax.table$No.Factors, cex=1.5)
plot(varimax.table$TLI,
      col = ifelse(varimax.table$TLI >=0.95, "green",
ifelse(varimax.table$TLI == max(varimax.table$TLI), "blue", "black")),
      pch = ifelse(varimax.table$TLI == max(varimax.table$TLI), 8, 1),
      cex = ifelse(varimax.table$TLI == max(varimax.table$TLI), 3, 2),
      xlab = "Number of Factors",
      ylab = "TLI",
      main= "TLI",
      cex.main=2,
      cex.axis=2,
      cex.lab=2)
textxy(X=varimax.table$Model, Y=varimax.table$TLI,
labs=varimax.table$No.Factors, cex=1.5)
plot(varimax.table$BIC,
      col = ifelse(varimax.table$BIC == min(varimax.table$BIC), "blue",
"black"),
      pch = ifelse(varimax.table$BIC == min(varimax.table$BIC), 8, 1),
      cex = ifelse(varimax.table$BIC == min(varimax.table$BIC), 3, 2),
      xlab = "Number of Factors",
      ylab = "BIC",
      main= "BIC",
      cex.main=2,
      cex.axis=2,
      cex.lab=2)
textxy(X=varimax.table$Model, Y=varimax.table$BIC,
labs=varimax.table$No.Factors, cex=1.5)
plot(varimax.table$chisq,
      col = ifelse(varimax.table$chisq == min(varimax.table$chisq),
"blue", "black"),
      pch = ifelse(varimax.table$chisq == min(varimax.table$chisq), 8,
1),
      cex = ifelse(varimax.table$chisq == min(varimax.table$chisq), 3,
2),
      xlab = "Number of Factors",
      ylab = "chisq",
      main= "chisq",
      cex.main=2,
      cex.axis=2,
      cex.lab=2)
textxy(X=varimax.table$Model, Y=varimax.table$chisq,
labs=varimax.table$No.Factors, cex=1.5)
plot(varimax.table$echisq,
      col = ifelse(varimax.table$echisq == min(varimax.table$echisq),
"blue", "black"),
      pch = ifelse(varimax.table$echisq == min(varimax.table$echisq), 8,
1),
      cex = ifelse(varimax.table$echisq == min(varimax.table$echisq), 3,
2),
      xlab = "Number of Factors",
      ylab = "echisq",
      main= "echisq",
      cex.main=2,
      cex.axis=2,
      cex.lab=2)

```

```
textxy(X=varimax.table$Model, Y=varimax.table$echisq,
labs=varimax.table$No.Factors, cex=1.5)
plot(vss.vari$cfit.1,
      col = ifelse(vss.vari$cfit.1 == max(vss.vari$cfit.1), "blue",
"black"),
      pch = ifelse(vss.vari$cfit.1 == max(vss.vari$cfit.1), 8, 1),
      cex = ifelse(vss.vari$cfit.1 == max(vss.vari$cfit.1), 3, 2),
      xlab = "Number of Factors",
      ylab = "vss1",
      main= "VSS complexity 1",
      cex.main=2,
      cex.axis=2,
      cex.lab=2
)
textxy(X=varimax.table$Model, Y=vss.vari$cfit.1,
labs=varimax.table$No.Factors, cex=1.5)
plot(vss.vari$cfit.2,
      col = ifelse(vss.vari$cfit.2 == max(vss.vari$cfit.2), "blue",
"black"),
      pch = ifelse(vss.vari$cfit.2 == max(vss.vari$cfit.2), 8, 1),
      cex = ifelse(vss.vari$cfit.2 == max(vss.vari$cfit.2), 3, 2),
      xlab = "Number of Factors",
      ylab = "vss2",
      main= "VSS complexity 2",
      cex.main=2,
      cex.axis=2,
      cex.lab=2
)
textxy(X=varimax.table$Model, Y=vss.vari$cfit.2,
labs=varimax.table$No.Factors, cex=1.5)
plot(vss.vari$map,
      col = ifelse(vss.vari$map == min(vss.vari$map), "blue", "black"),
      pch = ifelse(vss.vari$map == min(vss.vari$map), 8, 1),
      cex = ifelse(vss.vari$map == min(vss.vari$map), 3, 2),
      xlab = "Number of Factors",
      ylab = "map",
      main= "Velicer MAP",
      cex.main=2,
      cex.axis=2,
      cex.lab=2
)
textxy(X=varimax.table$Model, Y=vss.vari$map,
labs=varimax.table$No.Factors, cex=1.5)
dev.off()
## END of plots



## check manually each model
MLvari1 <- fa(X,nfactors=1,rotate="varimax",fm="ml")
MLvari1
fa.diagram(MLvari1,cut=0,digits=2)
fa.diagram(MLvari1,cut=0.2,digits=2, simple=FALSE) # Only Loadings with
abs(loading) > 0.2 cut will be shown
fa.diagram(MLvari1,digits=2)

```

```
MLvari2 <- fa(X,nfactors=2,rotate="varimax",fm="ml")
MLvari2
fa.diagram(MLvari2,cut=0,digits=2)
fa.diagram(MLvari2,cut=0.2,digits=2, simple=FALSE) # Only Loadings with
abs(loading) > 0.2 cut will be shown
fa.diagram(MLvari2,digits=2)

MLvari3 <- fa(X,nfactors=3,rotate="varimax",fm="ml")
MLvari3
fa.diagram(MLvari3,cut=0,digits=2)
fa.diagram(MLvari3,cut=0.2,digits=2, simple=FALSE) # Only Loadings with
abs(loading) > 0.2 cut will be shown
fa.diagram(MLvari3,digits=2)


## An example on how to check the variances manually
# (for a better understanding)
# take the 3 factors model
v <- print(MLvari3)
v

#SS loadings
ss1 <- round(sum((MLvari3$loadings[, 1])^2), 3)
ss2 <- round(sum((MLvari3$loadings[, 2])^2), 3)
ss3 <- round(sum((MLvari3$loadings[, 3])^2), 3)
ss1
ss2
ss3

# Proportion Var
prop.ss1 <- ss1/ncol(X)
prop.ss2 <- ss2/ncol(X)
prop.ss3 <- ss3/ncol(X)
prop.ss1
prop.ss2
prop.ss3

# Cumulative Var
# cumulative variance captured by all 3 factors
cum.var <- prop.ss1+ prop.ss2+prop.ss3
cum.var

########### END ###########
```

ACKNOWLEDGMENTS

I have been taught, throughout my life, that nothing comes "*out of the blue*". Relationships is the keyword. Positive or negative, they always have an impact. Hence, what I have been, what I have done in my life, including the research in this Dissertation, is in strong relationship with my environment. Even a small discussion, accompanied with a glass of wine, can be enlightening; and I really enjoy discussions with wine, especially those starting with a Valdobbiadene prosecco.

During the three years of my Ph.D., I had the chance, honor and pleasure to meet with plenty of people in which I am grateful:

My supervisors: I want to thank my supervisor Prof. Alessio Cecchinato as well as my co-supervisor Prof. Giovanni Bittante: Thank you very much for giving me the opportunity to work on this project, with a unique – worldwide – dataset, and for your supervision.

Collaborators: I am particularly indebted to Prof. Daniel Gianola (University of Wisconsin-Madison): Firstly, for introducing me to my group. Also, I have to say a great thanks to him, as well as to Prof. Guilherme J. M. Rosa, for inviting me to UW-Madison. I am grateful to both for their collaboration. This collaboration was a great opportunity for me to study in a top research group in our field, but also to meet and interact with more people.

I would like to acknowledge Dr. Stefano Biffani (Italian National Research Council) and Dr. Ezequiel Nicollazi (Council on Dairy Cattle Breeding - CDCB) for their scientific support. Your knowledge and advice was of great help to me, and I am vey happy that I had the chance to collaborate with both of you! All of my best wishes to your new careers!

Agripolis: My time at the department was made enjoyable due to many friends. The group in Agripolis has been a source of friendships and collaboration. Firstly, I want to thank my direct

collaborators and office-colleagues (past and present): Claudio Cipolat-Gotet, Alessandro Ferragina, Franzesco Zendri, Tania Bobbo, Giorgia Stocco, Hugo O. Toledo Alvarado (thanks also for your feedback in my Dissertation Huginio!), Matteo Bergamaschi, Cinzia Ribeca, Maria A. Paragga, Erika Pellattiero and Sara Pegolo. Especially for the first two, I have to say the following: Cla and Ale, thank you very much for everything!

Moreover I wish to thank for their collaboration and friendship: Cristina Sartori, Eirini Filiou, Serena Mazza, Alessandro Dalla Riva, Giovanni Niero, Giulio Visentin, Paolo Gottardo, Wilson Jose Fernandes Lemos (Junior), Roberta Rostellato, Luisa Magrin, Paola Semenzato, Georgia Faccioni, Giacomo Cesaro, Miriam Scalzo, Sarah Curoh, Giuseppe Carco, Chiara Roveglia, Giovanni Battista Conselvan, Marco Berton, Alice Varotto, Elisabetta Viale, Mirko Dalla Bona, Mirko Cattani, Alice Maurmayr, Anna Concollato, Enrica Franceschin, Giulia Rossi, Alba Sturaro, Valentina Toffanin, Laura Maccarana, Enrico Zanetti, Gisele Felix and Miriam Villamuelas Revenga.

UW-Madison: I feel particularly indebted to Vera Cardoso Ferreira for the great help provided to me, before and during my stay in Madison, as well as her hospitality: Thank you very much Vera, and I hope to reciprocate in the future! Moreover, I would like to thank my colleagues at UW: Tiago Luciano Passafaro, Arthur F. A. Fernandes, Mariana Mamede, Flavia Martins de Souza, Flaviana Miranda Gonçalves Miranda, Tiago Bresolin, Andriana Somavilla, Prof. Angelina Fraga, Fernando Brito Lopes, Inoue Keiichi, Tom Murphy, Ashley Mikshowsky, Jenna Kropp, Ahmed Sallam, Emma King.

I am thankful to Dr. Bruno Valente, Dr. Yaodong Hu and Dr. Vivian Felipe for their help in my research activities! Bruno, I really enjoyed our discussions on genetics, physics and philosophy!

I am grateful to Prof. Gerson Barreto Mourão (Universidade de São Paulo - USP), as well as his entire family, for their kindness and friendship: Thank you very much for being next to me in the most difficult period of my life.

Great thanks to Dr. Francisco Peñagaricano (now at the Department of Animal Sciences, University of Florida) for his assistance in setting up the gene-set enrichment and pathways analysis.

Moreover, during my stay at UW, I had the chance to meet with the Professors: George Shook, Kent Weigel, Hasan Khatib, David L. Thomas, Brian Kirkpatrick. I feel really lucky that had the chance to interact with you and grab some of your knowledge and experience in our field.

ANARB: I would like to acknowledge the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy) and especially my direct collaborators Attilio Rossoni and Enrico Santus for their alliance.

Secretaries: I would like to acknowledge the secretaries that were of major help to me during these 3 years: Rosalba Moro and Katia Milan (University of Padova) and Sheila Pink (University of Wisconsin-Madison). Especially for Rosalba, I need to say that I have never met such a friendly and always eager to help secretary!

CARIPARO: During the 3 years of my Ph.D. studies I was benefitted from financial support of the CARIPARO (Cassa di Risparmio di Padova e Rovigo) Foundation. This scholarship, apart from the general PhD financial support at UNIPD, provided me with free boarding (gym included!)... What else should a PhD student request for, to be undistracted dedicated in his studies?

Apart from my friends in Agripolis another group of friends made my stay in Padova enjoyable those 3 years:

Copernico friends: I would like to say a big thanks to Evangelia Chatzidimitriou (Elina), Sara Gulotta, Katy Gavazzoli, Valeria Valleri, Angela Abasimi, Anna Velliou, Areti Skreki and Francesca Giglioni.

Physics fellas: During these 3 years I had the chance, and the great pleasure, to interact with friends from the department of physics, with most of the discussion being accompanied with a glass of wine or beer. We enjoyed discussions in cosmology, amplitudes, black holes, super gravity, string theory, statistics and genetics/genomics! Thank you all guys: Dionysis Karagiannis (Dionysaras), William Torres Bobadilla (Willy), Andrea Galliani (Gaggia), Giulia Colluci, Nicolo Cribiori, Sukruti Bansal, Praxitelis Ntokos (Praxitelaras) and Fotis Farakos (Fotaras).

Moreover, a great thanks to Prof. Marigel Ángeles Pérez Cabal (Departamento de Producción Animal Facultad de Veterinaria - UCM, Madrid), Dr. Panagiotis Kougias (Department of Environmental Engineering, Technical University of Denmark), Dr. Stefano Campanaro (Department of Biology, UNIPD) and Dr. Vasileios Vasileiadis (now at ANADIAG Hellas Ltd). I really enjoyed the interaction, our discussions and your friendship. Marigel, I am really thankful for your feedback on my little notes on genomic predictions!

Finally, I am in depth to Hugo, Tania, Cristina, Marigel, Elina, Willy, Panagiotis, as well as to Noemi Isacchi, Marta Vescovi and Clorinda Bello (Department of Biotechnology and Biosciences, University of Milano-Bicocca) for reading parts of the Dissertation and providing with valuable comments and suggestions.

Thanks all the readers for their patience and I hope to find something interesting in this work!

Thank you very much all!!

Grazie mille!!

Ευχαριστώ πολύ!!

I enjoyed myself!

Χρήστος Δαδούσης
(Christos Dadousis)

THE THESIS FOR THE GENERAL PUBLIC: THE BEAUTY OF GENOMICS AND INTERDISCIPLINARITY OF SCIENCE

This chapter was written for the general public that might be interested in this research but lacking of basic background in animal breeding and genetics/genomics, and terminology that is required to understand this work.

Based on own experience, there is much of confusion in the public on what is breeding. Hence, the objective was to i) shed some light in the field of animal breeding, ii) introduce to the reader the beauty of genomics and the interdisciplinarity of science, iii) present the backbone of this research.

This section is presented in the form of a dialogue. The two persons participating are: Chris (me; C) and in the role of the interviewer Anastasia (A). All of the questions are real and have been selected the past few years during discussions with various people, researchers and non-researchers.

Research topic and introduction to dairy cattle breeding

A: So Chris, what is the topic of your PhD?

C: The general subject deals with cheese breeding in dairy cattle. A large amount of bovine milk is used for cheese production; hence we are interested in selecting cows that produce more cheese, and not just more milk. The work was split in 2 parts: i) the genomics and ii) the phenomics. Particularly, I was interested in identifying regions on the DNA that may be related with a higher cheese yield in the cows. If such regions exist, it might then be possible to select the best cows solely by observing these DNA regions. However, since cheese is a quite complicated process and many factors are involved, e.g. the percentage of the protein and the fat in the milk, the type of

these proteins, the coagulation properties, e.g. how fast the milk is transformed from liquid to solid, we would like to suppress all this information into a minimum set of variables. Note, that already in some breeding programs more than 40 traits are included in their breeding goal, without taking into account the cheese-related traits. The case is getting more complicated due to different correlations among traits, both on sign and strength. For instance, the more the milk a cow is producing the less the percentage of the fat in the milk we will have. But we wish to have both, simultaneously.

A: So, working with animals means you are a vet?

C: No. I am not a veterinarian.

A: What is your research area?

C: My field is called *animal breeding and genetics*.

A: Genetics? So, you change the DNA of the cows? Like taking genes from the fish and import it to plants to be resistant in cold? Are you working with genetically modified organisms (GMO)?

C: NO! Animal breeding is not focused on GMO, although this technology can be used as a tool for breeding. Briefly, with GMO technology, we can take a small part of the DNA of one organism and plug it in to another organism. Doing this, the new organism who obtained this piece of the DNA will have phenotypic characteristics that are regulated from this specific DNA piece. Taking advantage of this question, I can say the following: A similar to GMO methodology has recently made its first steps in livestock species and in breeding. The differences are that i) the parts of the DNA are not transferred from one organism to another, but we change the DNA within the individual and ii) we call this technology *gene editing* or *genome editing*. More details on this topic can be found on the internet: http://nas-sites.org/ilar-roundtable/roundtable-activities/gene-

editing-to-modify-animal-genomes-for-research/?utm_source=NAP+Newsletter&utm_campaign=5e4a4f4867-Event_2015_12_1_Gene_Editing_Workshop&utm_medium=email&utm_term=0_96101de015-5e4a4f4867-101857429&mc_cid=5e4a4f4867&mc_eid=6e10e4a391. Moreover, especially in the case of beef cattle and sheep it might be interesting to read the recent paper of Proudfoot and colleagues (Proudfoot et al., 2015). The potential benefits of this technology in animal breeding have been summarized in another recent work (Jenko et al., 2015).

A: And what is your opinion about this method?

C: I will reply by re-quoting Richard Feynman: "*For a successful technology, reality must take precedence over public relations for nature cannot be fooled"*. Moreover, especially in the case of genetics/genomics the history has taught us, so far, that *the more we learn, the less we understand in genetics* (Prof. Steve Jones). However, I would also like to add the following, for those that are afraid of everything new coming: *With the same knife that you cut your bread every day you can also harm someone*. Hence, technology itself is neither good nor bad. The way we use the technology makes the difference, keeping always in mind that *Ignorance more frequently breeds confidence than does knowledge* (Charles Darwin).

Let us explain now what animal breeding is about, and more specific the dairy cattle breeding: Imagine that you have the whole population of dairy cows in Italy. We are interested in increasing in every generation the milk production in the whole population, as an average.

A: Ok.

C: How we do this: We just need to identify and select the best males and the best females and mate them. For e.g. let us define as "best" the cows that produce more milk and the bulls that have female relatives that produce high amounts of milk. We firstly identify and then select these

"best" males and females and we control the mating (this is called "assortative mating") to produce the next generation.

A: Ok. So you are a geneticist or biologist?

C: Not exactly. The cornerstone of animal breeding is the quantitative genetics theory. However, knowledge from more scientific disciplines is required, such as population genetics, statistics and molecular biology. Moreover, with the recent technological advances, knowledge in programming, bioinformatics and advanced statistics is also essential. It is an integration of a variety of scientific disciplines.

Genomics

C: This work is heavily based on genomics, so it would be wise to have a small overview on the basis of genomics, i.e. the DNA. We could say that the DNA contains all the alphabet that each organism needs to express itself, in a similar way that humans are using alphabet to express themselves and communicate to each other. In a similar way that a vast amount of words, expressions, sentences and grammars can be produced in languages, an enormous amount of combinations can be done among different parts of the DNA in each organism, to produce a variability of products necessary for life.

To the best of my knowledge, the history of DNA traces back to Erwin Schrodinger and his book *What is life? The Physical Aspects of the Living Cell* (Schrodinger, 1946). The book was based in his public lectures at Dublin Institute for Advanced Studies (Trinity College). Schrodinger talked about the *hereditary code script* and made the hypothesis that *we believe a gene – or perhaps the whole chromosome fibre – to be an aperiodic solid*. This book has been considered one of the most influential scientific writings. Indeed, it is has been reported that many scientists, including

famous biologists such as J.B.S. Haldane and Francis Crick, were influenced by *What is life?*. This is one of the reasons that I strongly recommend this book to any student in my field. Moreover, I strongly believe that if we really want to understand nature, physics and life sciences should interact. This view has been summarized in two reports of the National Research Council of US in 2010 (National Research Council (US) Committee on Research at the Intersection of the Physical and Life Sciences, 2010) and 2014 (National Research Council (US). Division on Earth and Life Studies. Committee on Key Challenge Areas for Convergence and Health, Board on Life Sciences, 2014). An interesting field has already been born, integrating physics (quantum mechanics) and biology, called Quantum Biology. For more details in quantum biology see: the book *Quantum aspects of life* (Abbott et al., 2008) and the lectures of the Workshop "Quantum Biology: Current Status and Opportunities" (http://www.ias.surrey.ac.uk/workshops/quantumbiology/).

Now, let us imagine this molecule: you can envisage the DNA as a skein that if we unwind it will be ~ 2m long. Note, however, that this is within a cell! How many cells do we have? According to a recent study (Bianconi et al., 2013), an adult human body contains 37.2×10^{12} cells. We know that the distance between earth and the moon is $\sim 384 \times 10^3$ km. If we do the math, we get that the total length of the DNA that an adult human body contains is approximately 10,000 times to the moon and back!

Another analogy, to think of the amount of the information enclosed in the DNA, is the following: Imagine the DNA as a very small string where upon it are attached small beads (like a kompoloi: for those that are familiar with) and each of those beads, has as a name a capital letter of the alphabet, but in this case we need only 4 of the letters (A, T, G, C). Now, within a cell, DNA comprises ~3billion letters. The famous book of Leo Tolstoy, *War and Peace*, contains 3 million

letters. To reach the 3 billion we need 1,000 copies of the *War and Peace*, and if we place the books one above the other, we will have the height of an 18-storey building (Oshlack, 2013)!

A: Wow! But each of the cells contains exactly the same DNA?

C: No, differences appear among cells. For many years, DNA was extracted based on thousands or millions of cells. However, recent technological advances allow for DNA extraction from a single cell (Owens, 2012). The question is on how to use the new information. Now, a good question could also be: why the DNA contains ~3 billion letters, why not more or less. Well, in this case I do not have any answer.

A: Does the DNA remains the same over the lifetime of an individual?

C: That is another extremely interesting question, thank you! I am aware of 1 study from John Hopkins Medical Institutions published in 2008. Researchers tried to address this question. What they found is that, indeed, our genome changes over lifetime. However, what changes is not the DNA letters, but what is known as the *epigenetic* mechanism. And this might contain part of the answer for many diseases.

A: We are discussing about genetics, and now you said epigenetics? What is epigenetics?

C: I think it is easy to answer, but difficult to understand. If we say that genetics is the sequence of the DNA letters, epigenetics is all the rest around the DNA! And while genetics represent something that is heritable, i.e. is transferred from generation to generation, the epigenetic part is strongly interacting with the environment (although recent studies have proposed a heritable mechanism of epigenetics, too! This means that the inheritance of phenotypic characteristics might not solely depend on the DNA).

A: And how you deal with all this complexity in your work?

C: It is easy: we make assumptions, simplify and approximate the solutions and, consequently, the new knowledge obtained. If the simplest version is working then we are ok. If not, then we need to increase the complexity.

A: Talking about genomics, what is a gene Chris?

C: *The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid* (Crick, 1970). Giving a definition for a gene, nowadays, is not so easy. Here is a more recent one: *A gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products* (Gerstein et al., 2007). As it has been stated by Prof. J. A. Stamatoyannopoulos - based on the results from The ENCODE Project Consortium (ENCODE) - *results are re-shaping many long-held beliefs… including the very definition of the gene* (Stamatoyannopoulos, 2012). More definitions for the gene can be found in the human genome project webtv (https://www.youtube.com/watch?v=1MTZwa_TE_o&list=PL1ay9ko4A8slDIOZvtYjTys_BTDc0klkS).

However, borrowing the idea from Prof Steve Jones, together with searching for a definition of the gene, I would challenge the reader of this Dissertation to have a google search on the internet, typing the phrase: ***scientists find gene for***. Results are surprising, both in terms of quantity and quality.

There are, though, even more interesting stuff in the genome.

A: Like?

C: We know, for e.g. that the human genome contains roughly 20,000 genes. Also, the difference in the genome between any 2 persons is less than ~2%! However, before making

conclusions, we need to say that the human genome and the genome of the murine differ only in 300 genes (Ivanitskii, 2010)!

A: Ups

C: Exactly! That's the reason we believe that thinking the genomic information as a blueprint is actually a fallacy. It is much more likely that the information included in the DNA is rather an algorithm, which we need to understand on how it works. This is where the epigenetics and quantum biology come into the game!

Data: phenotypes and genotypes

C: We collected 1.5L of milk from 1,264 Brown Swiss cows. The animals belonged to 85 herds. Each animal was sampled only once. Based on this milk, individual cow cheese was produced in the lab (Cipolat-Gotet et al., 2013). Moreover, the whole coagulation and curd firmness process was monitored. Hence, we had 1,264 cheeses, one from each cow. Blood was taken from 1,152 cows in order to extract the DNA and genotype the animals with the Illumina SNP (single nucleotide polymorphism) chip, containing ~50,000 SNP. Note that today the cost of genotyping one animal with this SNP chip is ~60 euros.

A: What is cheese?

C: According to World Health Organization (WHO) and Food and Agriculture Organization of the United Nations (FAO) standards cheese is *"the ripened or unripened soft, semi-hard, hard, or extra-hard product, which may be coated, and in which the whey protein/casein ratio does not exceed that of milk…"*(World Health Organization and Food and Agriculture Organization of the United Nations). Note, ricotta is not a cheese!

A: Why ricotta is not considered as a cheese?

C: Because ricotta is made from the whey, i.e. the liquid that remains after obtaining the cheese curd.

A: And what is coagulation and curd firmness?

C: Milk coagulation after rennet (or similar coagulation agents) is the first step to cheese production. Coagulation (also known as clotting) is the transformation of the milk from its liquid state to a gel.

A: You said before that the animals were genotyped with SNP. What is a SNP?

C: The SNP is a type of genetic marker in the DNA. It is a polymorphism, i.e. a change in the genome that appears in the population with a frequency greater than a – arbitrarily chosen – threshold (e.g. 1%). It is called *single nucleotide* because the change/substitution is made on only one letter (nucleotide) of the DNA, e.g. the letter A is changed into the letter C.

A: Why you are using SNP for the genotypes?

C: Imagine SNPs as flags in the genome. We are interested in the genes, however, it is practically difficult to work with all genes. Thus, instead of genes, we use the DNA markers. By identifying markers across the whole genome, we simply increase the probability each of these markers to be closely linked with at least one gene. This, in our terminology, is known as linkage disequilibrium (LD). So, the markers represent the genes in our analysis. Take the following analogy: Imagine that I am interested in a girl (her name is Ylenia). However, it is difficult to find Ylenia in Padova and she doesn't like social networks, e.g. facebook (fb). Nevertheless, her best friend, Camilla, is a fun of fb and always update her position, wherever she goes. So, since they are best friends, by tracking Camilla, through her fb, it is very likely to spot Ylenia as well, which is our target. How much likely? This depends on how close friends they are (the strength of "their

LD"). So, in this analogy, Ylenia is the gene, Camilla represents the marker linked to the gene and the LD is the relationship between them.

A: Are the SNPs the only genetic polymorphisms or there are more types?

C: There are also other types of genetic markers, such as copy number variation, microsatellites, minisattelites, etc. However, working with SNP is much easier and much cheaper. The rest of the genetic markers have much longer sequence in the genome compared to SNP, i.e. they are not just one letter changes, but can be hundreds or thousands.

A: And what is the reason to do this analysis on individual cow cheese with genomics?

C: That is a great question! Firstly, we have already seen that the cheese traits are heritable (Bittante et al., 2013). This means that cow cheese characteristics can pass from generation to generation, which implies that the information can be used in breeding programs. However, for breeding purposes, recording at a population level is required. At present, it is very difficult to produce cheese from tens, hundreds thousands or millions of cows, due to high costs, labor demand and lack of an appropriate technology. Instead, we can investigate in a small sample of our population if there are some specific genomic regions that are related to the cheese production. Then, perhaps it would be possible to select the best individuals based on their genomic background.

Methods: the five parts of the Thesis

C: The project was split in 5 parts. The first 3 parts were focused in genomic studies. The CHAPTERS 1 and 2 dealt with genome-wide association analysis (GWAS) and CHAPTER 3 with gene-set enrichment and pathway-based analysis. The first method is called *genome-wide* because the markers that are used, the SNP, are distributed along the whole genome, i.e. located on all the

chromosomes. The term *association* means that we link/associate the phenotypes with the markers, and thereby with specific regions on the genome. Further, to alleviate some of the problems related to GWAS analysis and to complement the GWAS analysis, we collected the results from the first 2 GWAS analyses and performed a gene-set enrichment and pathway-based analysis (CHAPTER 3). The aim was to identify biological pathways that might be related to the cheese-making traits. In this case, firstly we collected the most important SNPs. Then, we identified the genes that were located very close to each of these SNPs. Finally, we tried to identify the biological pathways that these genes are known to be involved in.

The 4th CHAPTER was focused on the cheese-related phenotypes. We tried to diminish a set of 26 traits involved in the cheese-making process into a small set of variables. The 26 measured phenotypes were i) milk quantity and quality traits, i.e. milk yield, milk fat and protein percentages, ii) coagulation and curd firmness properties, i.e. trait that are measured during the gelation process (from milk to cheese) and iii) the actual cheese traits, e.g. the percentage of cheese obtained from the 1.5L of milk, the percentage of fat and protein of milk that retained in the cheese. A brief description on the coagulation, curd firmness and the cheese traits is given in the Appendix I. Finally, the 5th CHAPTER was an integration of all the previous analysis, i.e. we used the small set of the Fs identified in the 4th CHAPTER as phenotypes and performed GWAS and gene-set enrichment and pathway-based analysis.

Results

C: The genomic analysis revealed two main chromosomic regions. The first one was on chromosome 6, in the region where the casein genes are located, and more precisely, close to kappa casein. The importance of kappa casein in cheese making is known. The second important

chromosomic region was at the tail of chromosome 11, where again an effect on the casein variants and the beta lactoglobulin (a protein that is present in the whey, not in the cheese) is known. However, we detected signals in many more chromosomes, albeit at a weaker strength. It is still in question the importance of these chromosomic regions in the cow's ability to produce cheese. Concerning the gene-set enrichment analysis, the pathways that have been detected were mostly linked to the mammary gland functionality, bovine reproduction and general metabolism.

Concerning the last 2 contributions: Firstly, we managed to transform a set of 26 cheese-making traits into 10 variables (factors; Fs). These Fs captured the basic concept of the cheese-making process. For e.g. we found 1 factor that was mainly related with the cheese yield traits, a second factor was describing the coagulation properties of the milk, a third one was linked to the mammary gland health status of the cow, and so on. Interestingly, previous analyses with similar, albeit smaller, datasets observed the same pattern. Moreover, our results using the factors were consistent with the given name of the factors. Hence, although more research is needed in this type of analysis, before final application in breeding programs, our results appear quite promising.

Conclusion

C: In this work we were primarily focused on the genomic background of a cow's ability to produce cheese. Results are available for breeding purposes and for further exploration. Since this is a pioneering work from the phenotypic point of view, replication of our results from independent studies remains crucial (Ioannidis, 2013).

Before we close, I would like to propose to the reader the small video of R. Feynman https://www.youtube.com/watch?v=EYPapE-3FRw on the scientific method, complemented by inspector Clouseau https://www.youtube.com/watch?v=9KsVu11CSrQ.

Do not stop imagine and practicing!

A: Thank you very much Chris!

C: Thank you all for taking the time to read those lines! I enjoyed myself!

References

Abbott, D., P. C. Davies and A. K. Pati. 2008. Quantum Aspects of Life. Imperial College Press, .

Bianconi, E., A. Piovesan, F. Facchin, A. Beraudi, R. Casadei, F. Frabetti, L. Vitale, M. C. Pelleri, S. Tassani and F. Piva. 2013. An estimation of the number of cells in the human body. Ann. Hum. Biol. 40:463-471.

Bittante, G., C. Cipolat-Gotet and A. Cecchinato. 2013. 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.

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.

Crick, F. 1970. Central dogma of molecular biology. Nature. 227:561-563.

Gerstein, M. B., C. Bruce, J. S. Rozowsky, D. Zheng, J. Du, J. O. Korbel, O. Emanuelsson, Z. D. Zhang, S. Weissman and M. Snyder. 2007. What is a gene, post-ENCODE? history and updated definition. Genome Res. 17:669-681.

Ioannidis, J. P. 2013. This I believe in genetics: Discovery can be a nuisance, replication is science, implementation matters. Frontiers in Genetics. 4:.

Ivanitskii, G. R. 2010. 21st century: What is life from the perspective of physics? Physics-Uspekhi. 53:327-356.

Jenko, J., G. Gorjanc, M. A. Cleveland, R. K. Varshney, C. B. A. Whitelaw, J. A. Woolliams and J. M. Hickey. 2015. Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. Genetics Selection Evolution. 47:1.

National Research Council (US) Committee on Research at the Intersection of the Physical and Life Sciences. 2010.

National Research Council (US). Division on Earth and Life Studies. Committee on Key Challenge Areas for Convergence and Health, Board on Life Sciences. 2014. Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond.

Oshlack, A. 2013. A 10-step guide to party conversation for bioinformaticians. Genome Biol. 14:1.

Owens, B. 2012. The single life. Nature. 491:27-29.

Proudfoot, C., D. F. Carlson, R. Huddart, C. R. Long, J. H. Pryor, T. J. King, S. G. Lillico, A. J. Mileham, D. G. McLaren and C. B. A. Whitelaw. 2015. Genome edited sheep and cattle. Transgenic Res. 24:147-153.

Schrodinger, E. 1946. What is life?: The physical aspect of the living cell.

Stamatoyannopoulos, J. A. 2012. What does our genome encode? Genome Res. 22:1602-1611.

World Health Organization and Food and Agriculture Organization of the United Nations. General standard for cheese. CODEX STAN. 283-1978:.

CURRICULUM VITAE

Christos Dadousis (Χρήστος Δαδούσης) was born on 24th of May 1983 in Alexandroupolis, Hellas. He was raised in Didimoticho. In 2001 he graduated from high school in Didimoticho and enrolled at Aristotle University of Thessaloniki, School of Agriculture, Department of Animal Production for his BSc studies (5 years period in Hellas). In 2006, he obtained his BSc degree in agriculture, specialization in animal science. Between 2008-2010 he followed the MSc program entitled "Sustainable production systems and environment in agriculture" at Democritus University of Thrace (DUTH), Department of Rural Development. Under the supervision of Prof. Zafeirios Ampas, he worked in the pioneering field of genomic selection. The MSc thesis investigated the application of different genomic selection breeding schemes in Chios dairy sheep. During this period, he had the chance to collaborate and work in the lab of Prof. Peristera Paschou at the Department of Molecular Biology and Genetics (DUTH). Between 2010-2012 he followed the European Master in Animal Breeding and Genetics (EMABG; under Erasmus Mundus scholarship), spending the first year at the Norwegian University of Life Sciences (Norway) and the second at Wageningen University (The Netherlands). His thesis was performed under the supervision of Dr. Mario Calus (Wageningen UR Livestock Research) and Dr. Bjorg Heringstad (Norwegian University of Life Sciences), studying the potential use of principal component analysis for predicting genomic breeding values. He continued with a research period (1,5 years) at Technical University of Munich, Department of Plant Breeding. His research was focused in genomic predictions in maize and sunflower, collaborating with Prof. Chris-Carolin Schoen and Prof. Daniel Gianola (UW-Madison). In January 2014, he enrolled in the PhD School in Animal & Food Science at the Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE) University of Padova (scholar of the CARIPARO - Cassa di Risparmio

di Padova e Rovigo - Foundation, Padua, Italy). Having a pioneering and unique, worldwide, dataset, he investigated i) the genomic background and ii) the identification of latent phenotypes of individual cow cheese-related traits. Both approaches can enhance the dairy cattle breeding. The PhD project was supervised by Prof. Alessio Cecchinato and co-supervised by Prof. Giovanni Bittante.

LIST OF PUBLICATIONS

Manuscripts under review

2016 **Dadousis, C.,** C. Cipolat Gotet G. Bittante and A. Cecchinato. **Inferring individual cow effects, dairy system and genetics on latent variables underlying milk yield and quality, protein composition and cheese-making traits in dairy cattle.** J Dairy Sci. (*in review*)

2016 **Dadousis, C**., S. Pegolo G. Bittante and A. Cecchinato**. Genome-wide Association and Pathway-based Analysis using Latent Variables Related to Milk Yield and Quality, Protein Composition and Cheese-Making Traits in Dairy Cattle.** PLoS ONE *(submitted)*

Publications in international journals

2016 **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. **Genome-wide association study for cheese yield and curd nutrient recovery in dairy cows**. J Dairy Sci. (*accepted*). https://doi.org/10.3168/jds.2016-11586

2016 **Dadousis, C.,** S. Pegolo, G. J. M. Rosa, D. Gianola, G. Bittante and A. Cecchinato. **Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield and curd nutrient recovery in dairy cattle**. J Dairy Sci. (*accepted*) https://doi.org/10.3168/jds.2016-115867

2016 **Dadousis, C.,** S. Biffani, C. Cipolat Gotet, E. L. Nicolazzi, A. Rossoni, E. Santus, G. Bittante and A. Cecchinato. under review. **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. http://dx.doi.org/10.3168/jds.2015-10078

2014 **Christos Dadousis**, Roel F Veerkamp, Bjørg Heringstad, Marcin Pszczola, Mario PL Calus. **A comparison of principal component regression and genomic REML for genomic prediction across populations**. *Genet Sel Evol* 2014, **46:**60. doi:10.1186/s12711-014-0060-x

2011 Lewis J., Abas Z., **Dadousis C**., Lykidis D., Paschou P., Drineas P., 2011. **Tracing Cattle Breeds With Principal Components Analysis Ancestry Informative SNPs**. PLoS ONE, 6(4): e18007. DOI: 10.1371/journal.pone.0018007

Full papers published in proceedings of international conferences after peer review

2015 **Dadousis C.,** C. Cipolat-Gotet, S. Biffani, E. L. Nicolazzi, G. J. M. Rosa, D. Gianola, G. Bittante and A. Cecchinato. 2015. **Genome-wide association study for cheese yield and curd nutrients recovery in bovine milk.** Page 358 in Book of abstracts of the 66th Annual meeting of the european federation of animal science, Warsaw, Poland. Wageningen Academic Publishers, The Netherlands.

2010 Lewis J., Abas Z., **Dadousis C.**, Lykidis D., Paschou P., Drineas P., 2010. **Tracing The Origin Of Cattle Breeds With PCA-based Ancestry Informative SNPs**. In 9th WCGALP; Leipzig, Germany. 2010.

Conferences

2016 **C. Dadousis**, C. Cipolat-Gotet, S. Biffani, E.L. Nicolazzi, G. Bittante and A. Cecchinato, 2016**. Genome-wide associations for coagulation traits, individual cheese yield and curd nutrient recoveries in**

bovine milk. International Plant & Animal Genome XXIV, San Diego, CA, USA, 9-13 January, 2016.

2016 Sara Pegolo, Alessio Cecchinato, **Christos Dadousis**, Núria Mach, Massimiliano Babbucci, Marianna Pauletto, Luca Bargelloni, Stefano Schiavon, Giovanni Bittante, 2016. **Transcriptomic and phenotypic changes in young bull liver caused by low-impact and nutraceutical diets**. International Plant & Animal Genome XXIV, San Diego, CA, USA, 9-13 January, 2016.

2015 **C. Dadousis**, C. Cipolat-Gotet, S. Biffani, E.L. Nicolazzi, G. Bittante and A. Cecchinato, 2015.**Genome-wide association study to identify chromosomal regions associated with coagulation properties, curd firmness and syneresis modeling, and acidity in milk from Italian Brown Swiss cows.** XXI ASPA Congress, Animal Science and Production Association Milan, Italy, 9-12 June 2015.

2015 A. Cecchinato., C. Cipolat-Gotet, G. Stocco, **C. Dadousis** and G. Bittante, 2015. **Genetic and phenotypic relationships of milk coagulation properties and curd firmness modeling with the cheese yield and curd nutrients recovery in bovine milk.** XXI ASPA Congress, Animal Science and Production Association Milan, Italy, 9-12 June 2015.

2013 **Dadousis C.**, Veerkamp R.F., Heringstad B., Pszczola M., Calus M.P.L. **Principal component regression for genomic predictions in dairy cattle**. Workshop: Biometrische Aspekte der Genomanalyse V, 9.-11. März 2013, Schloss Rauischholzhausen, Germany.

Magazines

Δαδούσης Χ., Άμπας Ζ., 2015. **Γονιδιωματική επιλογή: προβλέψεις παραγωγικότητας των ζώων μέσω DNA**"(http://www.agronews.gr/ekmetaleuseis/ktinotrofikes-farmes/arthro/134384/paragogikotita-kai-dna-stin-ktinotrofia/)