Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel

Macciotta NPP; Gaspa G; Bomba L; Vicario D; Dimauro C; Cellesi M; Ajmone-Marsan P.

Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel

JOURNAL OF DAIRY SCIENCE, 98:8175:8185, 2015,
doi: 10.3168/jds.2015-9500

GWAS FOR LACTATION CURVE TRAITS

Interpretive summary

Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low density (7K) SNP panel.

by Macciotta et al.

A genome wide association study on PCA-derived lactation curve traits was performed on a sample of Italian Simmental cows genotyped with a low density SNP panel. Eighteen significant SNP were detected. Gene discovery highlighted some interesting candidate genes Results suggest interesting perspectives for the use of low density genotyped females for GWAS,.

Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low density (7K) SNP panel

N. P. P. Macciotta[*1], G. Gaspa[*], L. Bomba[‡], D. Vicario[†], C. Dimauro[*], M. Cellesi[*], and P. Ajmone-Marsan[‡]

[*] Dipartimento di Agraria,Università di Sassari, Sassari, Italy,

[†]ANAPRI, Udine, Italy,, Università di Sassari,

[‡]Università Cattolica del Sacro Cuore, Piacenza, Italy

[1]Corresponding author: Nicolò P.P. Macciotta, Dipartimento di Agraria, Sezione Scienze Zootecniche, Università di Sassari, viale Italia 39, 07100 Sassari, Italy. Phone number: 0039 079229299. Fax number: 0039 079229302. e-mail: macciott@uniss.it

ABSTRACT

High-throughput cow genotyping has opened new perspectives for genome wide association studies (GWAS). Directly recorded phenotypes and several records per animal could be used. In this study, a GWAS on lactation curve traits of 337 Italian Simmental cows genotyped with the Illumina low density beadchip (7K) was carried out. Scores of the first two principal components extracted from test day records (seven for each lactation) for milk yield, fat and protein percentages, and somatic cell score (SCS) were used as phenotypes. The first component described the average level (LEVEL) of the lactation curve, whereas the second summarized its shape (SHAPE). Data were analyzed with a mixed linear model that included fixed effects of herd, calving month, calving year, parity, SNP genotype, and random effects of animal and permanent environment. All statistically significant markers (Bonferroni corrected $P<0.05$) were associated to the LEVEL component (two

for milk yield, nine for fat percentage, six for protein percentages and one for SCS, respectively). No markers were found to be associated to the lactation curve shape. Gene discovery was performed using windows of variable size, according to the linkage disequilibrium level of the specific genomic region. Several suggestive candidate genes were indentified, some of which already reported to be associated with dairy traits as *DGAT1*. Others were involved in lipid metabolism, in protein synthesis, in the immune response, in cellular processes, and in early development. The large number of genes flagged in the present study suggests interesting perspectives for the use of low density genotyped females for GWAS, also for novel phenotypes that are not currently considered as breeding goals.

Key Words: Lactation curve, Principal Component Analysis, GWAS, LD panel.

INTRODUCTION

High throughput SNP platforms have been used in several genome-wide association studies (GWAS) in dairy cattle. These have been often carried out within Genomic Selection (GS) programmes, where bulls are preferentially genotyped. However, the need to enlarge the size of the reference populations has led to start genotyping cows. In most cases females are genotyped with low density (LD) SNP panels, that yield less accurate genomic evaluations but at approximately half the price (Wiggans et al., 2013).

The use of cow data in GS programs has raised some issues. For example, phenotypes need to be adjusted in order to be comparable with those of bulls (Wiggans et al., 2011). However, advantages in terms of direct genomic value (DGV) accuracy and in the economic returns of GS schemes that include females have been underlined in studies carried out both on simulated and real data (Lourenco et al., 2014; Thomasen et al., 2014).

So far, polygenic estimated breeding values (EBV), daughter yield deviations or deregressed proofs for traits of interest have been the most frequently used dependent variables in GWAS for

dairy traits. The use of genotyped females allows the direct modelling of recorded phenotypes. Moreover, multiple records are often available for each cow (e.g. multiple measures per lactation, successive lactations). An appealing, straightforward consequence of the use of repeated records is that SNP effects can be fitted along the whole lactation curve using longitudinal models (Strucken et al., 2015).

Candidate gene effects on lactation curves for milk production traits have been investigated using mathematical functions (Strucken et al., 2011; Szyda et al., 2014). In particular, values of function parameters or coefficients of orthogonal polynomials estimated for individual lactation curves were used as dependent variables. A GWAS for different measures of lactation persistency from random regression models was performed on primiparous Holstein and Jersey cows (Pryce et al., 2010). Genomic regions associated to lactation persistency in both breeds were detected on BTA6 and BTA26. A study carried out on German Holstein cows divergently selected for milk yield used Wilmink function parameters as phenotypes (Strucken et al., 2012). SNP associated to lactation curve traits were detected, especially for lactation persistency. Moreover, a dependence of results on parity order was highlighted.

Actually, modelling individual lactation curves is hampered by the large variability between cows (Olori et al., 1999). Individual differences are further enhanced by mathematical artefacts (e.g. negative or very high predicted values of test day yields at the beginning and at the end of the lactation trajectory). These problems often occur when models are fitted to patterns that markedly differ from the shape of the standard lactation curve (Macciotta et al., 2011). The resulting huge variability of parameter estimates, both in magnitude and sign, suggests a great care when these are used in further analyses (Macciotta et al., 2005). As an alternative, model-free algorithms able to derive measures of lactation curve traits without specific assumptions on data structure could be used. An example is principal component analysis (PCA). PCA carried out on test-day records for milk yield treated as different traits yielded two transformed variables related to i) the average production over the entire lactation (LEVEL) and ii) the shape of the lactation curve (SHAPE)

(Macciotta et al., 2006). Both traits had moderate values of heritability (h^2).

The aim of the present work was to perform a GWAS on lactation curve traits using model-free principal component analysis: dependent variables were obtained from the PCA of test-day records for milk production traits. Cows were genotyped with a low density (7K) SNP panel, as common in GS programs. All animals belonged to the Italian Simmental, the third-ranked breed in Italy for milk production (42.133 recorded lactations in 2013 in the Herd book) after Italian Holstein and Italian Brown (ICAR 2014).

MATERIALS AND METHODS

Data consisted of 10,605 test-day records for milk production and milk composition traits (milk yield, fat and protein percentages, and somatic cell score). Records were from 1,515 lactations of 337 Italian Simmental cows. Animals were farmed in 120 herds. Seven test-day (TD) records were considered for each lactation. Extra TD for lactations with more than 7 records were deleted. All cows were genotyped with the 7K Illumina bead-chip. Marker edits were on call rate (>.99) and minor allele frequency (>.01). After edits, 6,891 markers were retained for the analysis.

Principal component analysis was performed on the seven test day records, separately for each lactation. Scores of the first two Principal Components (PC1 and PC2) for each lactation were then calculated by multiplying the row vector of the standardized original variables by the column vector of the corresponding eigenvector coefficients. Scores were used as dependent variables in the association study carried out with the following mixed linear model:

$$y_{ijklmnop} = \mu + H_i + M_j + Y_k + Par_l + SNP_m + a_n + p_o + e_{ijklmnop}$$

where y = PC score; μ = overall mean; H = fixed effect of the i-th herd, M = fixed effect of the j-th calving month (12 months); Y= fixed effect of the k-th calving year (from 2002 to 2013); Par = fixed effect of l-th parity (from 1 to 6, >6); SNP = fixed covariable of the m-th SNP marker genotype (coded as 0,1,2 according to the copies of the second allele); a = random additive effect of

122 the n-th animal; p =random permanent environmental effect of the o-th lactation; e =random
123 residual. The animal and permanent environment random effects were assumed to be normally
124 distributed as $\mathbf{N}(0,\mathbf{A}\sigma^2_A)$ and $\mathbf{N}(0,\mathbf{I}\sigma^2_p)$, respectively, where **A** is the pedigree relationship matrix
125 and σ^2_A and σ^2_p are additive genetic and permanent environmental variances, respectively. The
126 mixed model was solved using a REML algorithm implemented in the ASREML software (Gilmour
127 et al., 2000). Values published by Macciotta et al. (2006) were used as variance priors (PC1=
128 0.4427, 0.4654, 1.4005; PC2= 0.05264, 0.02954, 0.66050, for σ^2_A, σ^2_P,and σ^2_e, respectively).
129 Bonferroni corrected significance levels for the SNP effects were calculated to account for
130 multiple testing: uncorrected *P* values were multiplied by the number of tests performed (i.e.,
131 6,891). SNP were considered significantly associated to the considered trait when the corrected *P*
132 value was lower than 0.05.
133 Gene discovery was performed in the genomic regions located around the significant SNP.
134 The width of these intervals was based on the linkage disequilibrium (LD) of the region. The LD
135 was calculated on a sample of 479 Italian Simmental bulls genotyped with the 54K Illumina bead-
136 chip in a previous study (Pintus et al., 2012). For each significant SNP detected in the present study,
137 the value of the r^2 statistic with all other SNP located in the same chromosome was calculated using
138 the Haploview software (Barret et al., 2005). Then the distance between the significant SNP and the
139 farthest SNP having an r^2 >0.15 was calculated. A window was then defined by considering such
140 distance both upstream and downstream the position of the significant SNP. Annotated genes
141 located in the windows were derived from UCSC Genome Browser Gateway
142 (http://genome.ucsc.edu./). SNP and gene positions were obtained from the UMD3.1 Bovine
143 genome assembly (Zimin et al., 2009).
144
145 **RESULTS**
146 ***Principal component analysis***

147 The first two principal components extracted from the correlation matrix of the test day
148 records accounted for most of the original variance. The sum of the first two eigenvalues was 83%
149 for milk yield, 50% for fat percentage, 57% for protein percentage, and 42% for SCS, respectively
150 (Table 1). The structure of eigenvectors showed an association between PC1 and all the test-day
151 records. PC2 was positively associated with the first and negatively with the second part of
152 lactation, respectively. Such a structure could be observed for all the four considered traits. The first
153 eigenvalue was always markedly larger than the second in agreement with previous studies
154 (Macciotta et al., 2006). Differences in magnitude can be observed between traits (Table 1) .
155 Figures 1-4 report average lactation patterns for animals grouped according to PC1 and PC2
156 scores for the four considered traits. In particular, figures 1 (a and b) highlight the role of the two
157 PC as phenotypic indices of level of production (LEVEL) and lactation curve shape (SHAPE) for
158 milk yield, respectively. The first two PC have a similar meaning also for fat percentage (Figures 2a
159 and 2b), protein percentage (Figures 3a and 3b) and somatic cell score (Figures 4a and 4b).
160
161 ***Association study***
162 Eighteen SNP significantly associated (Bonferroni corrected $P<0.05$) to the PC scores for
163 the considered traits (Table 2) were detected. Most of them were for fat and protein percentage (9
164 and 6, respectively). All significant SNP were associated to PC1, i.e., the variable that expressed the
165 level at which the lactation curve is located. No SNP were found to be significantly associated with
166 lactation curve shape (i.e., PC2).
167
168 ***Milk yield***
169 The top significant SNP for milk yield LEVEL (corrected $P=0.003$) was located on BTA6 at
170 approximately 89Mb. A strong association between this region and clinical mastitis, milk yield, and
171 protein percentage was reported in dairy cattle (Sahana et al., 2014). Some interesting genes are
172 located in the interval of approximately 0.48 Mb calculated around the significant SNP (Table 2).

One is the *GC* (Groups specific Component), a gene that encodes for a vitamin D binding protein. It is involved in several physiological functions as the modulation of inflammatory and immune response, binding of fatty acids and bone development (Speeckaert et al., 2014). It has been suggested as a putative candidate for milk yield (Raven et al., 2014) and clinical mastitis (Sahana et al., 2014). The casein cluster is also located in this genomic region, but it did not fall in the considered interval. The second SNP in the rank mapped on BTA21, close to the *NRTK3* (neurotrophic tyrosine kinase, receptor, type 3) locus (Table 3). This gene encodes for a membrane protein receptor and it is involved in the determinism of a type of breast carcinoma in humans, (Tognon et al., 2002). Other interesting genes located in this region are two mitochondrial ribosomal proteins (*MRLP1* and *MRPS11*) that have been reported as candidates for mitochondrial disorders in humans (Vasta et al., 2009).

Fat percentage

The most significant SNP ($P<0.001$) associated with the fat content LEVEL was located on BTA23 (Figure 5). A suggestive candidate gene that maps in the window around this marker is the desmoplakin locus. This gene encodes for a protein involved in the structure of desmosomes, intercellular junctions that provide tissue integrity at epithelial level (Garrod and Chidgey, 2008). The second significant marker was located on BTA7 at approximately 71.5 Mb. The clathrin interactor 1 (*CLINT1*) locus that encodes for a protein which is involved in the vesicle trafficking (Dodd et al., 2009) maps in this region. This gene has been found to be associated with skin colour in chicken (Sun et al., 2013).

Two significant SNP for LEVEL of fat content were highlighted on BTA14. They both map in a very dense region where the *DGAT1* is located. This gene has a well known major effect on milk fat in cattle (Grisart et al., 2002). The first marker (ARS-BFGL-NGS-34135) was reported to be statistically associated to milk yield, fat and protein percentages in a large multibreed study (Raven et al., 2014). The Zinc Finger Protein 34 (*ZNF 34*) and the Glutamic Pyruvate Transaminase

(*GPT*) are also located in the interval around this SNP. Associations between these genes and fat yield and percentage have been detected in Chinese Holstein (Jiang et al., 2014). The second SNP (ARS-BFGL-NGS-4939) was highlighted in US (Cole et al. 2011) and German (Wang et al., 2014) Holsteins, and in a multibreed population (Raven et al., 2014). The window considered for this marker was fairly large, due to the high linkage disequilibrium of this genomic region (Table 2). It includes the cytochrome P450, subfamily XI B, polypeptide 1 (*CYP11B1*). This gene has been suggested as a second relevant QTL on BTA14 affecting fat yield and content in cattle (Jiang et al., 2014; Kaupe et al., 2007; Mai et al., 2010).

A significant SNP for fat percentage mapped on BTA3 at around 99.6 Mb. The glutathione reductase gene is located in this region. Associations between genes involved in the metabolism of glutathione and milk yield, fat and protein percentage have been found in cattle (Raven et al., 2014). Moreover, the reductase gene family has been reported under balancing selection in a recent comparison between *Bos taurus* and *Bos indicus* genomes (Porto-Neto et al., 2013). The marker found on BTA 17 (Table 2) defined a 0.53 Mb window where maps the Claudin 5 (*CLDN5*). This gene econdes for a membrane protein that is a component of the thigh junctions. The significant SNP on BTA2 was found in a region that harbours several interesting genes. One is the Long-chain Acyl coenzyme A dehydrogenase (*ACAD*) gene, which encodes for a key enzyme of the fatty acid metabolism in the liver (Lia et al., 2013). Another gene located in this genomic region is the myosin light chain 1 (*MYL1*), whose expression has been related to the physiological status (lactation vs puberty) in cattle (Ron et al., 2007).

The second significant SNP found on BTA23 was located in a region that harbours some genes of potential interest. One is the Apolipoprotein B mRNA-editing enzyme catalytic subunit 2 (*APOBEC2*). It is expressed in the muscle and affects muscle fibre ratio and body mass in mice (Sato et al., 2010). Another is the MyoD family inhibitor (*MDFI*) that was proposed as a suggestive candidate gene for a QTL that affects fatness traits in pigs (Huang et al., 2011). Finally, also the interval around the significant marker found on BTA19 included different genes of potential

interest. The *ATP2A3* that is involved in calcium mobilization in the cell, could be mentioned. This gene has been recently reported as a selection signature in Ethiopian cattle populations (Edea et al., 2014). Another is the olfactory receptor, family 1, subfamily E, member 2-like (*LOC618124*). Genes of the olfactory receptor family have been found in selection signatures in cattle (Qanbari et al., 2010).

Protein percentage

The most significant SNP for protein percentage was located on BTA16 (Figure 6). This marker showed the largest LD (Table 2). Among the genes that map in the surrounding interval, the *EFCAB2* is of potential interest. It is involved in the micro architecture of the bone in humans (Mohan et al., 2013). Another interesting gene is the *SYMD3*, that contributes to neuromuscular processes and that has been found in a region of deleted CNV in Korean cattle (Shin et al, 2014). The second marker in order of importance for protein percentage was located on BTA12 at approximately 16 Mb. In this region maps the spermatid associated (*SPERT*), which has been reported in a region of selection signature in German Holsteins (Qanbari et al., 2010). Another interesting gene located close to the significant marker is the lymphocyte cytosolic protein 1 (*LCP1*), a highly conserved protein of the cattle genome (Lemay et al, 2009). The expression of this gene in the mammary tissue has been found to be associated with infections by *Staphilococcus aureus* (Huang et al., 2014).

The third marker was detected on BTA6. It is placed close to a gene cluster that included the leucine-rich repeat LGI family, member 2 (*LGI2*), and the DEAH (Asp-Glu-Ala-His) box polypeptide 15 (*DHX15*), involved in the immune response mechanism. Associations between these two genes and fat and protein percentages in German Holstein and Fleckvieh have been reported, respectively (Weikard et al., 2011). Another interesting gene located in this region is the superoxide dismutase 3 (*SOD3*). It has been found to be associated with intake in beef cattle (Al-Husseini et al., 2013). The fourth marker affecting the PC1 for protein percentage was located on BTA7, in a

region characterized by a relevant LD (Table 2). Several interesting genes could be found in the calculated window. Of particular interest is the Eukaryotic Translation Elongation Factor 2 (*EEF2*), involved in milk protein synthesis in the mammary gland as a mediator of the effect of growth hormone (Hayashi et al., 2009). Moreover, it has been found to be down regulated in cattle experimentally infected by Bovine tubercolosis (Meade et al., 2007). Another gene of interest in this region is the integrin beta 1 binding protein 3 (*ITGB1BP3*), that was found under positive selection in a comparison between Jersey, Guensey, and Zebuine breeds (Porto-Neto et al, 2013). Intervals surrounding the last two markers for protein percentage, located on BTA13 and BTA5 respectively, did not include any annotated gene for the bovine genome.

Somatic cell score

The only significant marker for somatic cell score was located at the beginning of BTA22. In this region, the vesicular, over expressed in cancer, prosurvival protein 1 (VOPP1) maps. This gene encodes for a protein that is involved in cellular apoptosis in vertebrates (Pei and Grishin, 2012).

DISCUSSION

Principal component analysis

PCA performed on test day records for milk production traits was able to synthesise the main aspects of the lactation pattern, i.e., the general level of production and its shape. The use of PCA sometimes results in a first component correlated with almost all original variables that absorbs most of the total variance. In these cases, the eigenvectors of the other extracted PC often show some variations (values that increase or decrease, or that change sign) in comparison with PC1 (Stearns et al., 2005). These eigenvector structures allow to infer specific aspects of relationships between groups of variables that cannot be seen in the first PC (Jombart et al., 2009). In the case of the lactation curve, the predominance of the first eigenvalue over the second may be also an expression of the larger heritability of milk yield compared to lactation persistency (Cole

277 and Null, 2009; Cole and VanRaden, 2006). The interpretation of eigenvalues as expression of the
278 genetic contribution of principal components to phenotypes was also proposed by Kirkpatrick and
279 Meyer (2004).
280 Of interest is also the difference that can be observed between the eigenvalues of the first
281 principal component in the four traits (Table 1). The largest value was for milk yield, a trait
282 characterised by a polygenic background. The smallest eigenvalue was for fat percentage, a trait
283 that is genetically determined by few genes with large effects (Hayes et al., 2010). From these
284 figures, a relationship between magnitude of the eigenvalue of PC1 and the genetic determinism of
285 the trait could be inferred.
286 PCA could be therefore seen not only as a dimension-reduction technique, but also as an
287 approach for investigating the genetic determinism of traits. Studies carried out both on dairy and
288 meat traits underlined a higher efficiency of PCA in comparison with univariate analyses for the
289 detection of SNP associations and of QTL with pleiotropic effects (Bolormaa et al., 2010; Stearns et
290 al. 2005).
291
292 ***Association analysis***
293 A relevant number of significant associations were detected in spite of the low density
294 marker map used. Such a result may be somewhat unexpected because the extent of linkage
295 disequilibrium between marker and QTL is one of the main factors affecting the power of GWAS
296 (Powell et al., 2011). On the other hand, a large number of significant markers was reported in a
297 previous research carried out on females (Cole et al., 2011). Furthermore, some of the most
298 significant SNP markers found in this study confirmed previous reports on dairy cattle (Cole et al.
299 2011; Raven et al., 2014; Wang et al., 2012).
300 It is worth remembering that GWAS on males are usually carried out on progeny tested or
301 on AI bulls, i.e., on the top animals of the breed. A reduction of the genetic variability in
302 comparison with females could be therefore expected. On the other hand, it has to be remembered

303 that the Italian Simmental is a dual-purpose breed. Bulls are selected by combining performance
304 test for beef traits and progeny test for dairy traits respectively. Moreover there is limitation in the
305 number of semen doses that can be used per bull (7,000). These two aspects may have contributed
306 to reduce the selection pressure and to maintain genetic variation.
307 The kind of analysed traits could be a further cause of results here obtained. Strucken et al.
308 (2011) found that the use of lactation curve parameters instead of yield data as dependent variables
309 provided greater power in detecting associations. Authors explained their results with an increase of
310 the genetic variance of the considered trait. In the present study, the Bonferroni correction of the
311 SNP significance level was implemented to prevent the occurrence of false positives. Moreover,
312 population stratification was accounted for by including the polygenic effect in the model.
313 All significant associations were found only for the principal component that described the
314 level at which the lactation curve is located. No markers were detected for the shape of the lactation
315 curve. This result is in contrast with previous studies that detected SNP associated to lactation
316 persistency. Some points could be discussed in order to find possible reasons for these
317 discrepancies. First of all the kind of dependent variable used. There is a lack of consensus on a
318 suitable measure of lactation persistency and this trait has been defined in many ways. The measure
319 of lactation curve shape used in the present paper is completely uncorrelated from milk yield (i.e.
320 $r_{PC1,PC2}=0$), and it explains a reduced quota of variation compared to the LEVEL component. This
321 may be also an explanation of the absence of common significant markers across traits. On the other
322 hand, in the paper by Strucken et al. (2011) persistency is defined by using the parameter *c* of the
323 Wilmink function, which is usually highly correlated with the parameter *a*, i.e., the one related with
324 yield (Macciotta et al., 2005). A second point is represented by the effect of parity. It is widely
325 acknowledged that first calving cows have a markedly flatter curve in comparison with older
326 parities. In the paper of Pryce et al. (2010), only primiparous cows were considered. Moreover,
327 Strucken et al. (2011) found significant associations only when first lactations were analysed
328 separately from those of older animals. In the present paper, cows of different parity were included

329 in the data set and primiparous represented about 25% of the records. Finally, it should be
330 remembered that variation of persistency explained by significant SNP in previous researches was
331 rather low, 1 to 2% in the study of Pryce et al. (2010).
332
333 ***Gene discovery***
334 The intervals surrounding the significant SNP allowed for the detection of several suggestive
335 candidate genes. Most of them were found for milk fat percentage. Some of these genes were
336 already known to affect dairy traits, as the *DGAT1*. This is a rather common result for GWAS but it
337 was somewhat unexpected for Italian Simmental. Pintus et al. (2012) hypothesised that the low
338 DGV accuracy for fat percentage obtained in Italian Simmental bulls was due to the absence of
339 allele segregation at the *DGAT1* locus in this breed. The frequency of the favourable *DGAT1* allele
340 was >0.99 in a previous work carried out on 95 cows (Scotti et al., 2010). Results of the present
341 study have been obtained on larger sample. Moreover, the two significant markers found in the
342 region of BTA14 that harbours the *DGAT1* locus have been reported also for other breeds (Cole et
343 al., 2011; Raven et al., 2011; Wang et al., 2012). A recent study carried out on Italian Simmental
344 bulls reported a SNP in the promoter region of the *DGAT1*, but no associations with dairy traits
345 were found (Chessa et al. 2015). In any case it is worth remembering that several genes map in this
346 genomic region. Thus SNP significance could be also due to the effect of other genes.
347 Most of identified putative candidate genes have a biological connection with the lactation.
348 Examples are those involved in the lipid metabolism (*ACAD*), in protein synthesis (*EEF2*), in the
349 integrity of the epithelium (*CLDN5* and the desmoplakin), and in the immune response (*DHX15*,
350 *GC*, and *LCP1*). Of interest is also the gene found on BTA6 that encodes antioxidant enzyme
351 *SOD3*, whose expression has been found to be downregulated in the mammary tissue of cows fed a
352 diet rich in polyunsaturated acid (Cortes et al., 2012). This gene is involved in the determinism of
353 feed intake in cattle (Al-Husseini et al., 2013). Other genes, that map in genomic regions where
354 selection signatures have been detected, are of more general functions. Examples are olfactory

355 receptors, transmembrane proteins involved cellular processes, and mediators of early development
356 as *LOC618124*,*VOPP1*, and *ITGB1BP3*, respectively.
357
358 **CONCLUSIONS**
359 The GWAS carried out in Italian Simmental cows highlighted some markers statistically
360 associated with lactation curve traits. In particular, significant associations were found for the
361 principal component that describes the production level at which the lactation curve for different
362 dairy traits is located. No significant SNP were found for lactation curve shape. This result
363 disagrees with previous studies on dairy cattle that report significant associations between genomic
364 regions and lactation persistency Reasons can be found in the kind of measure used, in the dual
365 purpose aptitude of the Italian Simmental breed, and in the composition of the sample of animals
366 considered (especially concerning parity). On the other hand, some putative candidate genes
367 detected in the present study were found to be associated to production traits in previous researches.
368 In spite of the low density marker map used, a relatively large number of significant markers was
369 detected. These results suggest the use of low density genotyped females for GWAS also for novel
370 phenotypes (e.g. milk fatty acid spectrum, milk coagulation properties) that are not currently
371 measured in breeding programs.
372
373 **ACKNOWLEDGEMENTS**
374 This research has been funded by the Italian Ministry of Agriculture (Grant INNOVAGEN, Project
375 Leader prof. Alessio Valentini).
376
377
378
379 **REFERENCES**
380
381
382 Al-Husseini, W. , C. Gondro, K. Quin, R. M. Herd, J. P. Gibson and Y. Chen. 2013. Expression of
383 candidate genes for residual feed intake in Angus cattle. Anim. Genet. 45: 12–19.
384

Barrett, J.C, B. Fry, J. Maller, and M.J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinf. 21:263-265.

Bolormaa, S., J.E. Pryce , B. J. Hayes , and M. E. Goddard. 2010. Multivariate analysis of a genome-wide association study in dairy cattle. J. Dairy Sci. 93 :3818–3833.

Chessa, S., E.L. Nicolazzi, L. Nicoloso, R. Negrini, R. Marino, D. Vicario, P. Ajmone Marsan, A. Valentini, B. Stefanon. 2015. Analysis of candidate SNPs affecting milk and functional traits in the dual-purpose Italian Simmental cattle. Livest. Sci.173:1.8.

Cole, J. B., and P. M. VanRaden. 2006. Genetic Evaluation and Best Prediction of Lactation Persistency. J. Dairy Sci. 89:2722–2728.

Cole, J.B., and D.J. Null. 2009. Genetic evaluation of lactation persistency for five breeds of dairy cattle. J. Dairy Sci. 92:2248–2258.

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

Cortes, C., M.F. Palin, N. Gagnon, C. Benchaar, P. Lacasse, and H. V. Petit. 2012. Mammary gene expression and activity of antioxidant enzymes and concentration of the mammalian lignin enterolactone in milk and plasma of dairy cows fed flax lignans and infused with flax oil in the abomasums. British J. Nutr. 108: 1390–1398.

Dodd, M. E., J. Hatzold, J. R. Mathias, K. B. Walters, D. A. Bennin, J. Rhodes, J. P. Kanki, A. T. Look, M. Hammerschmidt, and A. Huttenlocher. 2009. The ENTH domain protein Clint1 is required for epidermal homeostasis in zebrafish. Develop. 136: 2591-2600.

Edea, Z., H. Dadi, S.-W. Kim, J.-H. Park, G.-H. Shin, T. Dessie and K.-S. Kim. 2014. Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in Ethiopia. J. Anim. Breed. Genet. 131: 358–366

Garrod, D., and M. Chidgey. 2008. Desmosome structure, composition and function. Biochimica et Biophys. Acta 1778: 572–587.

Gilmour, A. R., B. R. Cullis, S. J. Welhan, and R. Thompson. 2000. ASREML manual. New South Wales Dept. Agric., Orange, Australia.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. (2002) Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12: 222-231.

Hayashi, A.A., K. Nones, N. C. Roy, W. C. McNabb, D. S. Mackenzie, D. Pacheco, and S. McCoard. 2009. Initiation and elongation steps of mRNA translation are involved in the increase in milk protein yield caused by growth hormone administration during lactation. J. Dairy Sci. 92:1889–1899.

Hayes, B. J., J. Pryce, A.J. Chamberlain, P.J. Bowman, and M.E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. Plos Genet. 6:e1001139.

Huang, W.H., Z.X. Ma, Z.Y. Xu, Y.Z. Xiong and B. Zuo. 2011. Detection of novel SNPs and mapping of the fatness QTL on pig chromosome 7q1.1-1.4 region. Genet. Mol. Res. 10: 3090-3097.

Huang, J., G. Luo, Z. Zhang, X. Wang, Z. Ju, C. Qi, Y. Zhang, C. Wang, R. Li, J. Li, W. Yin, Y. Xu, S.J. Moisá, J.J. Loor and J. Zhong1. 2014. iTRAQ-proteomics and bioinformatics analyses of mammary tissue from cows with clinical mastitis due to natural infection with Staphylococci aureus. BMC Genom. 15:839.

International Committee for Animal Recording.2014.Yearly enquiry on the situation of cow milk recording in ICAR member countries. Results for the years 2012-2013.. ICAR. Rome. 2014.

Jiang, L., X. Liu, J. Yang, H. Wang, J. Jiang, L. Liu, S. He, X. Ding, J. Liu, and Q. Zhang. 2014. Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. BMC Genom. 15:1105.

Jombart, T., D. Pontier, and A. B. Dufour. 2009. Genetic markers in the playground of multivariate analysis. Hered. 102:330–341

Kaupe, B., H. Brandt, E. M. Prinzenberg, and G. Erhardt. 2007. Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. J. Anim. Sci. 85:11–21.

Kirkpatrick, M., and K. Meyer. 2004. Direct estimation of genetic principal components: simplified analysis of complex phenotypes. Genet. 168:2295-2004.

Lemay, D.G., D.J. Lynn, W.F. Martin, M.C. Neville, T. M. Casey, G. Rincon, E.V. Kriventseva, W. C. Barris, A.S. Hinrichs, A.J. Molenaar, K.S. Pollard, N.J. Maqbool, K. Singh, R. Murney, E.M. Zdobnov, R.L. Tellam, J.F. Medrano, J. B. German, and M. Rijnkels. 2009. The bovine lactation genome: insights into the evolution of mammalian milk. Genome Biology 2009, 10:R43

Lia, P., Y. Liu, Y. Zhang, M. Long,Y. Guo, Z. Wang, X. Li, C. Zhang, X. Li, J. He, G. Liu. 2013. Effect of Non-Esterified Fatty Acids on Fatty Acid Metabolism-Related Genes in Calf Hepatocytes Cultured in Vitro. Cell. Physiol. Biochem. 32:1509-1516

Lourenco, D.A.L., I. Misztal, S. Tsuruta, I. Aguilar, E. Ezra, M. Ron, A. Shirak, and J. I. Weller.2014. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. J. Dairy Sci. 97 :1742–1752.

Mai, M. D., G. Sahana, F. B. Christiansen, and B. Guldbrandtsen. 2010. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. J. Anim. Sci. 2010. 88:3522–3528.

Macciotta, N. P. P. , D. Vicario, and A. Cappio-Borlino. 2005. Detection of Different Shapes of Lactation Curve for Milk Yield in Dairy Cattle by Empirical Mathematical Models. J. Dairy Sci. 88:1178–1191

Macciotta, N.P.P., D.Vicario, and A. Cappio-Borlino. 2006. Use of multivariate analysis to extract latent variables related to level of production and lactation persistency in dairy cattle. J. Dairy Sci. 89: 3188-3194.

Macciotta, N.P.P. , C.Dimauro, S.P.G. Rassu, R. Steri, G. Pulina. 2011. The mathematical description of lactation curves in dairy cattle. It. J. of Animal Sci. 11:213-223.

Meade K.G., E. Gormley, M.B. Doyle, T. Fitzsimons, C. O'Farrelly, E. Costello, J. Keane, Y. Zhao, and D.E. MacHugh. 2007. Innate gene repression associated with Mycobacterium bovis infection in cattle: toward a gene signature of disease. BMC Genom. 8:40

Mohan, S., Y. Hu, B. Edderkaoui. 2013. Identification of Gender-Specific Candidate Genes that Influence Bone Microarchitecture in Chromosome 1. Calc. Tiss. Int. 92: 362-371

Olori, V.E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. Livest. Prod. Sci. 58: 55-63.

Pei, J., and N. V. Grishin. 2012. Unexpected diversity in Shisa-like proteins suggests the importance of their roles as transmembrane adaptors. Cell. Signall. 24:758–769

Pintus, M. A., G. Gaspa, E. L. Nicolazzi, D. Vicario, A. Rossoni, P. Ajmone-Marsan, A. Nardone, C. Dimauro, and N.P.P. Macciotta. 2012. Prediction of genomic breeding values for dairy traits in Italian Brown and Simmental bulls using a principal component approach. J. Dairy Sci. 95:3390–3400

Porto-Neto, L.R., T.S. Sonstegard, G.E. Liu, D.M. Bickhart, M.V.B. Da Silva, M.A. Machado, Y.T. Utsunomiya, J.F. Garcia, C. Gondro, and C.P. Van Tassell. 2013. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. BMC Genom. 14:876

Powell, J. E., A. Kranis, J. Floyd, J. C. M. Dekkers, S. Knott, and C. S. Haley. 2011. Optimal use of regression models in genome-wide association studies. Anim. Genet. 43: 133–143.

Pryce, J.E., M. Haile-Mariam, K. Verbyla, P.J. Bowman, M.E. Goddard, and B.J. Hayes. 2010. Genetic markers for lactation persistency in primiparous Australian dairy cows J. Dairy Sci.93: 2202-2214

Qanbari, S., E.C.G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi and H. Simianer. 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. Anim. Genet. 41: 377–389

Raven, L.A. , B.G. Cocks, and B.J Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genom. 15:62.

Ron, M., G. Israeli, E. Seroussi, J. I Weller, J.P. Gregg, M. Shani, and J. F. Medrano. 2007. Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. BMC Genom. 8:183

Sahana, G., B. Guldbrandtsen, B. Thomsen, L-E. Holm, F. Panitz, R.F. Brøndum, C. Bendixen, M.S. Lund. 2014. Genome-wide association study using high-density single nucleotide

polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. J. Dairy Sci. 97: 7258–7275

Sato, Y., H.C. Probst, R. Tatsumi, Y. Ikeuchi, M.S. Neuberger, and C. Rada. 2010. Deficiency in APOBEC2 Leads to a Shift in Muscle Fiber Type, Diminished Body Mass, and Myopathy. J. Biol. Chem. 285:7111–7118.

Scotti, E. , L. Fontanesi, F. Schiavini, V. La Mattina, A. Bagnato, V. Russo. 2010. Short communication: DGAT1 p.K232A polymorphism in dairy and dual purpose Italian cattle breeds. It. J. Anim. Sci. 9:e16

Shin, D.H., H.J. Lee, S. Cho, H.J. Kim, J.Y. Hwang, C.K. Lee, J.Y. Jeong, D. Yoon, and H. Kim. 2014. Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. BMC Genom. 15:240

Speeckaert, M.M., R. Speeckaert, N. van Geel, J.R. Delanghe. 2014. Chapter One – Vitamin D Binding Protein: A Multifunctional Protein of Clinical Importance. Adv. Clin. Chem. 63: 1-57.

Stearns, T.M., J.E. Beever, B.R. Southey, M. Ellis, F.M. McKeith, and S.L. Rodrigues-Szas. 2005. Evaluation of approaches to detect quantitave loci for growth, carcass and meat quality on swine chromosomes 2, 6, 13, and 18. II. Multivariate and principal component analyses. J. Anim. Sci. 83:2471-2481.

Strucken, E.M., D.J. de Koning, S.A. Rahmatalla, and G.A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442–449
Strucken, E.M., R.H. Bortfeldt, D.J. de Koning, and G.A. Brockmann. 2012. Genome-wide associations for investigating time-dependent genetic effects for milk productionn traits in dairy cattle. Anim. Genet. 43:375–382.

Strucken, EM, Y.C.S.M. Laurenson, and G.A. Brockmann. 2015. Go with the flow—biology and genetics of the lactation cycle. Front. Genet. 6:118. doi: 10.3389/fgene.2015.00118

Sun, Y., G. Zhao, R. Liu, M. Zheng, Y. Hu, D Wu, L. Zhang, P. Li and J. Wen. 2013. The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study. BMC Genom.14:458

Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on complex traits as variables over time. Anim. Genet. 45:322–328.

Tognon, C., S.R. Knezevich, D. Huntsman, C.D. Roskelley, N. Melnyk, J.A. Mathers, L. Becker, F. Carneiro, N. MacPherson, D. Horsman, C. Poremba, P. H.B Sorensen. 2002. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. Cancer Cell 2:367–76.

Thomasen, J.R., A.C. Sørensen, M. S. Lund, and B. Guldbrandtsen. 2014. Adding cows to the reference population makesa small dairy population competitive. J. Dairy Sci. 97 :5822–5832.

Vasta, V., S.B.Ng, E.H. Turner, J. Shendure, and S. Houn Hahn. 2009. Next generation sequence analysis for mitochondrial disorders. Genom. Med. 1:100.

589 Zimin, A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.
590 P. Van Tassell, T.S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S.L.
591 Salzberg. 2009. A whole-genome assembly of the domestic cow, Bos Taurus. Genom. Biol. 10:R42
592
593 Wang, X., C. Wurmser, H. Pausch, S. Jung, F. Reinhardt, J. Tetens, G. Thaller, R. Fries.2104.
594 Identification and Dissection of Four Major QTL Affecting Milk Fat Content in the German
595 Holstein-Friesian Population. PlosOne 7:e40711
596
597 Weikard, R., P. Widmann, J. Buitkamp, R. Emmerling, and C. Kuehn. 2011. Revisiting the
598 quantitative trait loci for milk production traits on BTA6. Anim. Genet. 43: 318–323
599
600 Wiggans, G.R., T.A. Cooper, P.M. VanRaden, and J.B. Cole. 2011. Technical note: Adjustment of
601 traditional cow evaluations to improve accuracy of genomic predictions. J. Dairy Sci. 94 :6188–
602 6193
603
604 Wiggans, G.R., T.A. Cooper, C.P. Van Tassell, T.S. Sonstegard, and E.B. Simpson. 2013. Technical
605 note: Characteristics and use of the Illumina BovineLD and GeneSeek Genomic Profiler low-
606 density bead chips for genomic evaluation. J. Dairy Sci. 96 :1258–1263.
607
608
609
610
611

612 Table 1. Eigenvectors and eigenvalues of the first two principal components (PC1 and PC2)
613 extracted for the seven test day records for milk yield, fat and protein percentage, somatic cell
614 score.

	Milk Yield		Fat		Protein		SCS	
Test day	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Test 1	0.32	0.60	0.19	0.74	0.14	0.75	0.26	0.78
Test 2	0.38	0.38	0.32	0.49	0.34	0.42	0.32	0.48
Test 3	0.40	0.16	0.38	0.11	0.35	0.19	0.38	-0.17
Test 4	0.40	0.02	0.40	-0.07	0.47	-0.01	0.43	-0.22
Test 5	0.40	-0.20	0.44	-0.19	0.43	-0.14	0.44	-0.21
Test 6	0.38	-0.38	0.44	-0.30	0.44	-0.28	0.39	-0.11
Test 7	0.34	-0.54	0.41	0.24	0.35	-0.37	0.38	-0.17
Eigenvalue	0.71	0.12	0.35	0.15	0.42	0.15	0.39	0.13

615
616
617

Table 2. Markers significantly associated (Bonferroni adjusted level of significance<0.05) with scores of principal components representing lactation curve traits

Marker	BTA	Position (bp)	Trait[1]	P	Interval
BTB-01654826	6	88,891,318	PC1 MY	0.003	±242,276
Hapmap38505-BTA-51760	21	19,193,100	PC1 MY	0.015	±690,747
ARS-BFGL-NGS-11659	23	47,677,534	PC1 FP	0.0003	±5,663
ARS-BFGL-NGS-75852	7	71,545,383	PC1 FP	0.001	± 214,702
ARS-BFGL-NGS-34135	14	1,675,278	PC1 FP	0.003	± 783,698
ARS-BFGL-NGS-18926	3	99,592,696	PC1 FP	0.009	± 2,160,054
ARS-BFGL-NGS-24012	17	74,948,921	PC1 FP	0.016	± 267,855
ARS-BFGL-NGS-4939	14	1,801,116	PC1 FP	0.029	± 2,574,789
Hapmap34329-BES11_Contig247_1378	2	98,446,391	PC1 FP	0.031	± 132,872
Hapmap41022-BTA-55560	23	15,147,471	PC1 FP	0.044	± 1,729,322
Hapmap58587-ss4652997	19	24,972,085	PC1 FP	0.05	± 256,339
BTB-01225907	16	32,653,232	PC1 PP	0.0205	± 10,322,590
ARS-BFGL-NGS-55674	12	16,069,827	PC1 PP	0.0215	± 373,760
ARS-BFGL-NGS-97136	6	45,909,053	PC1 PP	0.0240	± 868,964
UA-IFASA-8256	7	21,704,630	PC1 PP	0.0341	± 1,619,254
BTA-102818-no-rs	13	83,632,355	PC1 PP	0.0423	± 244,086
Hapmap50090-BTA-75536	5	11,284,482	PC1 PP	0.0492	± 28,293
ARS-BFGL-NGS-84222	22	531,301	PC1 SCS	0.0360	± 43,004

[1]Trait= PC1MY (first principal component extracted from milk yield test data)
PC1FP (first principal component extracted from fat percentage test data)
PC1PP (first principal component extracted from protein percentage test data)
PC1SCS (first principal component extracted from somatic cell score test data)

628
629
630

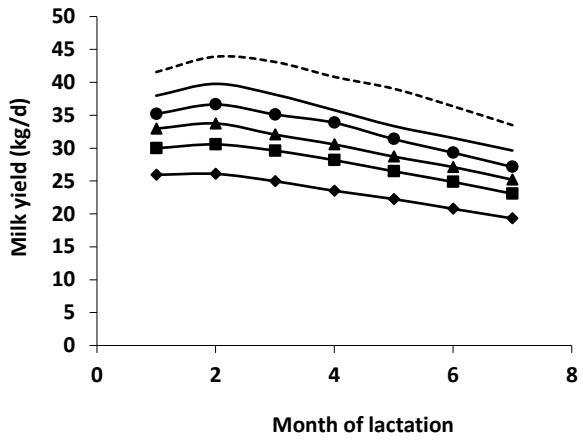

Figure 1a.

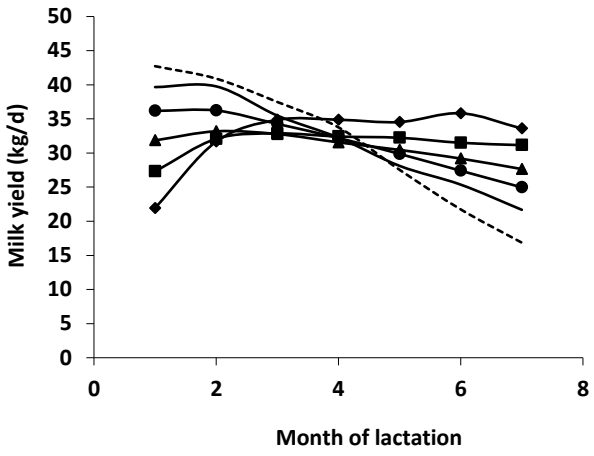

Figure 1b.

631

632

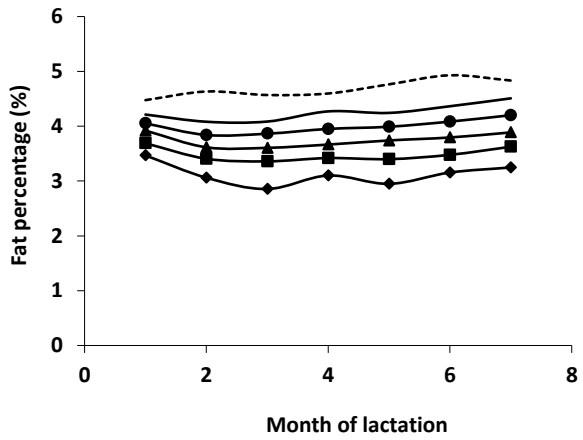

Figure 2a.

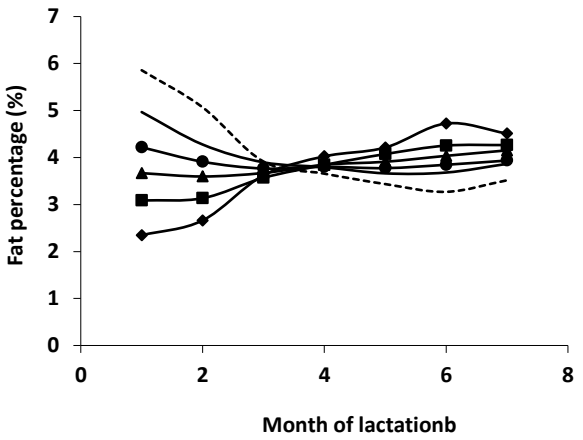

Figure 2b.

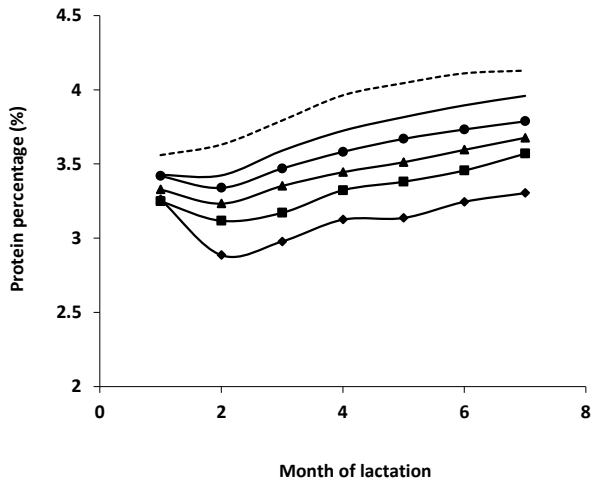

Figure 3a

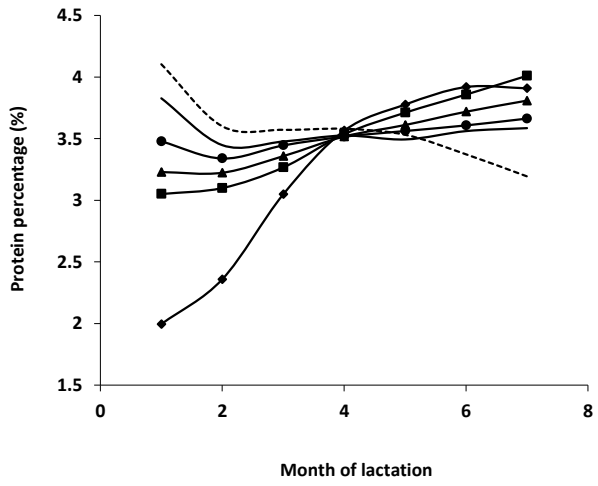

Figure 3b

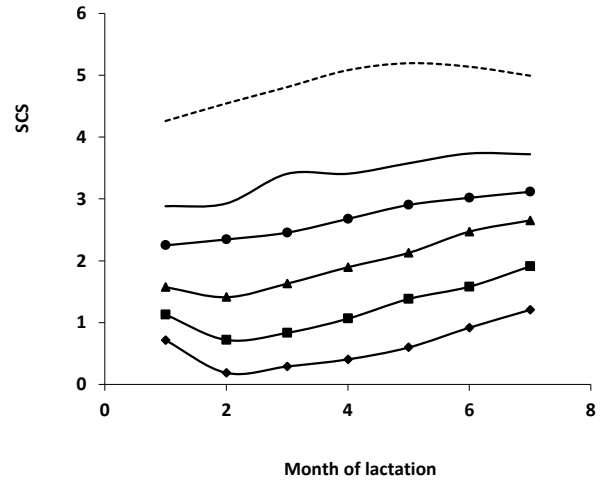

Figure 4a

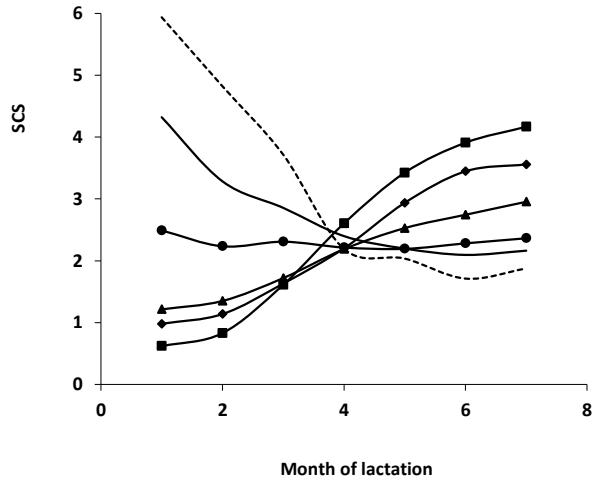

Figure 4b

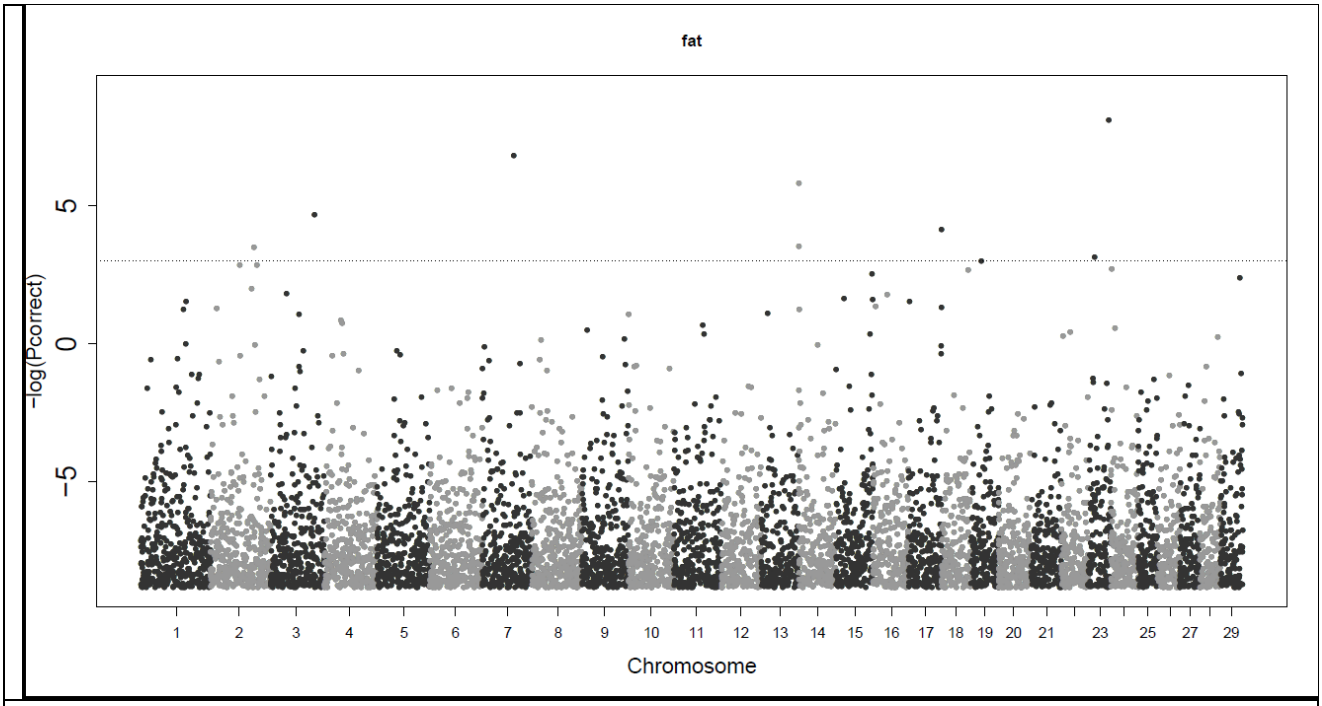
fat
−log(Pcorrect)
5
0
−5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
23
25
27
29
Chromosome

Figure 5

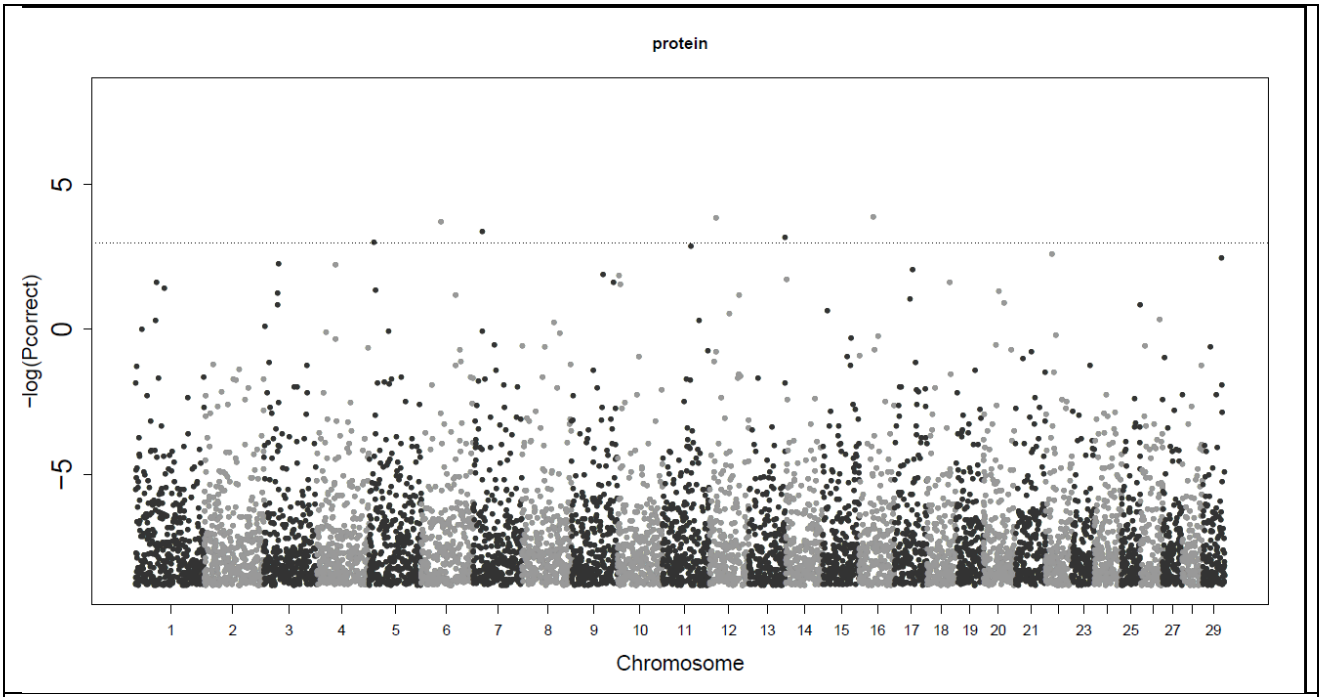
protein
−log(Pcorrect)
5
0
−5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
23
25
27
29
Chromosome

Figure 6

Captions of figures

Figure 1a. Average lactation curves for milk yield of groups of cows of different PCl score classes (◆=<-2 ■=−2 to −1; ▲= −1 to 0; ●= 0 to 1; continuous line = 1 to 2; dotted line=>2). Points are plotted for the average day in milk on each test day.

Figure 1b. Average lactation curves for milk yield of groups of cows of different PC2 score classes (◆=<-2 ■=−2 to −1; ▲= −1 to 0; ●= 0 to 1; continuous line = 1 to 2; dotted line=>2). Points are plotted for the average day in milk on each test day.

Figure 2a. Average lactation curves for fat percentage of groups of cows of different PCl score classes (◆=<-2 ■=−2 to −1; ▲= −1 to 0; ●= 0 to 1; continuous line = 1 to 2; dotted line=>2). Points are plotted for the average day in milk on each test day.

Figure 2b. Average lactation curves for fat percentage of groups of cows of different PC2 score classes (◆=<-2 ■=−2 to −1; ▲= −1 to 0; ●= 0 to 1; continuous line = 1 to 2; dotted line=>2). Points are plotted for the average day in milk on each test day.

Figure 3a. Average lactation curves for protein percentage of groups of cows of different PCl score classes (◆=<-2 ■ =−2 to −1; ▲= −1 to 0; ●= 0 to 1; continuous line = 1 to 2; dotted line=>2). Points are plotted for the average day in milk on each test day.

Figure 3b. Average lactation curves for protein percentage of groups of cows of different PC2 score classes (◆=<-2 ■ =−2 to −1; ▲= −1 to 0; ●= 0 to 1; continuous line = 1 to 2; dotted line=>2). Points are plotted for the average day in milk on each test day.

Figure 4a. Average lactation curves for somatic cell score f of groups of cows of different PCl score classes (◆=<-2 ■ =−2 to −1; ▲= −1 to 0; ●= 0 to 1; continuous line = 1 to 2; dotted line=>2). Points are plotted for the average day in milk on each test day.

Figure 4b. Average lactation curves for somatic cell score of groups of cows of different PC2 score classes (◆=<-2 ■ =−2 to −1; ▲= −1 to 0; ●= 0 to 1; continuous line = 1 to 2; dotted line=>2). Points are plotted for the average day in milk on each test day.

Figure 5. Genome-wide association study of the scores of the first principal component (LEVEL) for fat percentage. The dashed line corresponds to a Bonferroni corrected significance level of 0.05

Figure 5. Genome-wide association study of the scores of the first principal component (LEVEL) for protein percentage. The dashed line corresponds to a Bonferroni corrected significance level of 0.05