Livestock Science ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Genome-wide association study for somatic cell score in Argentinean dairy cattle

Juan P Nani [a], Maria A Raschia [b], Mario A Poli [b], Luis F Calvinho [a], Ariel F Amadio [a,c,*]

[a] E.E.A. Rafaela, Instituto Nacional de Tecnología Agropecuaria, Rafaela, Santa Fe, Argentina
[b] Instituto de Genética "Ewald A. Favret", CICVyA, Instituto Nacional de Tecnología Agropecuaria, Castelar, Buenos Aires, Argentina
[c] Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

ARTICLE INFO

Article history:
Received 17 June 2014
Received in revised form
30 January 2015
Accepted 6 February 2015

Keywords:
Mastitis
Resistance
GWAS
Milk

ABSTRACT

This study aimed to understand the genomic architecture of Argentinean dairy herds by measuring linkage disequilibrium (LD) and identifying loci associated with parameters calculated from somatic cell count (SCC). Phenotypic data consisted of 3530 SCC records from 544 Holstein and Holstein x Jersey cows owned by a single dairy company located in the Central dairy area of Argentina. SCC was recorded every 40 days. After quality control, genotypic data consisted in 38,872 single nucleotide polymorphisms (SNP). The squared correlation of the alleles at two loci (r^2) was computed for all SNP pairs on each chromosome. At marker distances less than 10 kb the average r^2 was 0.40. Between 40 and 50 kb the average r^2 was 0.25 and 0.18 for 100 kb apart. Three different variables were calculated from the somatic cell score (SCS): the arithmetic mean (AM), the maximum value (MAX) and the arithmetic mean of the top 3 values (TOP3). Few significant SNP associations were found. As expected, polygenic traits such as SCC are influenced by multiple loci throughout the genome, each with a relatively small effect. AM on one side and TOP3 and MAX on the other, showed different SNP associated showing that they capture different aspects of mastitis response. AM was significantly associated with two SNP: ARS-BFGL-NGS-114608 (BTA1) and Hapmap60306-rs29023088 (BTA5). MAX and TOP3 were significantly associated with four SNP: ARS-BFGL-NGS-107594, ARS-BFGL-NGS-104220 (BTA10), BTA-43543-no-rs (BTA18) and ARS-BFGL-NGS-109705 (BTA26). MAX and TOP3 were equivalent phenotypic variables to be used in a GWAS. These results contribute to gain insight into the genomic regions influencing the SCC in Argentinean herds.

1. Introduction

Mastitis is the most prevalent disease of dairy cattle, causing high economic losses mainly due to decreased milk production (Seegers et al., 2003). Somatic cell count (SCC) is an indirect method of diagnosing subclinical mastitis, positively correlated with mammary gland inflammatory response and closely associated with udder health (Schukken and Wilson, 2003). For monitoring udder health and performing genetic evaluations, SCC is commonly log-transformed to somatic cell score (SCS) to fit a normal distribution.

Several previous studies analyzing variables derived from SCC have been published (Green et al., 2004; de Haas et al., 2008; Urioste et al., 2010, 2012; Koeck et al., 2012). These studies defined alternatives traits to try to cover different

* Corresponding author at: E.E.A. Rafaela, Instituto Nacional de Tecnología Agropecuaria, Rafaela, Santa Fe, Argentina.
E-mail address:
amadio.ariel@inta.gob.ar (A. Amadio).

http://dx.doi.org/10.1016/j.livsci.2015.02.007
1871-1413/

aspects of SCC variation along the lactation. Those traits showed higher heritability than clinical mastitis, indicating good possibilities for selection. Among the traits studied, the most remarkable are, among others, the average SCC or SCS for different intervals of DIM (days in milk), SCS-SD (the standard deviation of SCS), the number of test-day (TD) with values higher than a threshold (e.g. TD > 200 for TD with more than 200,000 cell/ml), patterns of peaks in SCC. Lactation-average SCS is historically used for genetic improvement of mastitis resistance, although it has limitations (Green et al., 2004). Also, the maximum log SCC was used by Green et al. (2004) and Koeck et al. (2012). Green et al. (2004) described that MAX and SCS-SD were the best indicators of clinical mastitis caused by a range of pathogens. Koeck et al. (2012) used MAX for the first time in genetic studies, and showed that MAX, TD > 200, TD > 500 and peaks describe variability in SCC during lactation better than mean SCS.

Dairy breeding programs have initially focused on selecting animals for increased milk production, then for protein yield and more recently for increasing fat and protein content, leading to a decline in health traits. More recently, breeding programs are including health and functional traits to reverse the negative trend in these traits (Miglior et al., 2005). The use of molecular markers has helped in the identification of polymorphisms associated with phenotypic variation (Dekkers, 2004). These molecular markers have been used also to increase the accuracy in the prediction of estimated breeding values (EBV), where genome-wide dense marker maps are useful to predict genomic breeding values (GEBV) called genomic selection (Meuwissen et al., 2001). The sequencing of the bovine genome led to the subsequent discovery of millions of single-nucleotide polymorphisms (SNP) across the bovine genome (Elsik et al., 2009). High-throughput multiplex SNP genotyping allowed alleles identification in large populations at relatively low costs (Matukumalli et al., 2009) and genome-wide association studies (GWAS) can be routinely performed in cattle populations. A GWAS tests each SNP for an association with variation in a trait assuming that the SNP is in linkage disequilibrium (LD) with, or close to, a causative mutation affecting the trait (Hirschhorn and Daly, 2005) allowing for an efficient way to evaluate genetic variants throughout the genome. Measuring the extent of LD or the non-random association of alleles at different loci throughout the genome can reveal historical recombination events which is crucial when mapping for important traits helping to determine the density of SNP needed to detect causative variants as well as for genomic selection (Goddard and Hayes, 2009). The LD is population specific so it needs to be determined for the population under study. In livestock where effective population size are relatively small and populations are under intense selection, the extent of LD tend to be high, spanning large distances throughout the genome (de Roos et al., 2008; Sargolzaei et al., 2008).

The objectives of this study were to analyze the extent of LD in commercial Argentinean herds and to detect molecular markers associated with different variables obtained from SCS through a GWAS to include this information in a breeding program.

2. Materials and methods

2.1. Population and phenotypes

Phenotypic data consisted of 3530 SCC records of 379 Holstein and 165 Holstein X Jersey crosses (in total 544 animals). All cows were daughters of 22 bulls and belonged to 14 dairy farms property of a single dairy company, located within 10 km distance from each other, in the province of Santa Fe, Argentina. All animals were kept under similar feeding and sanitary management conditions and were milked mechanically twice a day. SCC was recorded every 40 days by official Dairy Herd Improvement (DHI) service.

For SCC determination, milk samples were preserved with azidiol (0.3%) at 4 °C and analyzed within 24 h. The SCC determinations were performed by a commercial laboratory (Laboratorio Regional de Servicios Analíticos, Esperanza) using an automated counter Somacount 300 (Bentley Instruments, Minesotta, USA). Milk samples were warmed up at 39–40 °C to facilitate homogenization. All cows included in the study had at least 4 records per lactation (average 6). The analyzed data corresponded from 20 to 400 days in milk records. The lactation number varied from 2 to 6 depending on the cow, but only one lactation of each cow was kept for analysis.

SCC was transformed into SCS as described by Shook (1982). Three different variables were calculated from the SCS per lactation: the arithmetic mean (AM), the maximum value (MAX) and the arithmetic mean of the top 3 values (TOP3). MAX was first used in genetic studies by Koeck et al. (2012) and showed a high genetic correlation with TD > 200, TD > 500 and PK (binary variable for presence or absence of peaks of SCC during lactation). These variables showed a stronger genetic correlation with mastitis than averaged SCS in different periods of lactation (Koeck et al., 2012). TOP3 was defined as the arithmetic mean of the top 3 values of SCS for a cow. We used TOP3 to try to differentiate situations in which a cow shows a unique strong peak of SCC with those with more than one peak. In this way, with TOP3 we tried to capture events with sustained or repetitive high values of SCS, rather than only one peak.

2.2. DNA extraction

Bovine blood samples were collected from the caudal veins into tubes containing EDTA as anticoagulant. Genomic DNA was extracted from fresh blood using a commercial kit (AxyPrep Blood Genomic DNA Miniprep Kit, Axygen Biosciences, Union City, CA) according to the manufacturer's directions. DNA quantity and quality were determined by measuring absorbance ($A_{260;}$ $A_{260/280}$ ratio respectively) in a Thermo Scientific NanoDrop™ 1000 Spectrophotometer.

2.3. Genotyping and quality control

Genotyping was carried out using Illumina BovineSNP50v2 BeadChip (Illumina Inc., San Diego, CA, USA) at a commercial laboratory (GeneSeek, Lincoln, NE, USA). Genotyping quality control was performed using PLINK v1.07 (Purcell et al., 2007).

Please cite this article as: Nani, J.P, et al., Genome-wide association study for somatic cell score in Argentinean dairy cattle. Livestock Science (2015), http://dx.doi.org/10.1016/j.livsci.2015.02.007

For the initial 54,609 genotyped SNP, only SNPs that passed the following criteria were included in the analysis: (a) Minor allele frequency (MAF) $\geq 5\%$; (b) percentage of missing genotypes for the SNP across all samples $<5\%$, and (c) no strong deviation from Hardy–Weinberg proportions ($p<0.001$). All SNP with no map position reported were also excluded from further analysis.

Mendelian inconsistences were detected with PLINK and set to missing genotypes. A Mendelian inconsistency refers to the detection of an allele absent in the progeny. All incoherent genotypes (homozygote or heterozygote) were marked as missing. Missing genotypes were imputed with BEAGLE v 3.3.2 program (Browning and Browning, 2009) using 20 iterations of the phasing algorithm. After quality pruning, there were 38,872 SNP and 540 animals left.

2.4. Linkage disequilibrium

LD was quantified by the squared correlation of the alleles at 2 loci (r^2) and was computed using PLINK. r^2 was computed for all SNP pairs on each chromosome using 576 animals (Hill and Robertson, 1968). SNP pairs were grouped into intervals of 10 kb from 0 to 10 Mb. In order to assess the extent of LD at various distances in each chromosome, the arithmetic mean of r^2 for SNP pairs in each interval was calculated. The r^2 was calculated as follows:

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

where $D=f(AB)-f(A)f(B)$, and $f(AB)$, $f(A)$, $f(a)$, $f(B)$ and $f(b)$ were observed frequencies of haplotype AB and of alleles A, a, B, and b, respectively.

2.5. Association tests

Genetic parameters for the three variables were estimated by the Restricted Maximum Likelihood method using WOMBAT (Meyer, 2007). A multi trait animal model was fitted in order to estimate genetic and phenotypic correlations between all traits. Data were analyzed using the following model:

$$y_{ijkl} = \mu + L_i + HYS_j + PH_k + A_l + e_{ijkl} \quad (1)$$

where y_{ijkl} is the response variable corresponding to AM, MAX and TOP3; μ is the overall mean; L_i is lactation number ($i=2$ to 6) of cow l, from HYS_j-group, the fixed effect accounting for the combination of herd (H) in which the record was produced, year (Y) and season (S) of calving ($j=1$ to 115); PH_k is the fixed effect of proportion of Holstein blood in the animal; A_l is the random additive genetic effect of the cow distributed as $\sim \boldsymbol{N}(\mathbf{0}, \mathbf{A}\sigma_a^2)$, which accounted for (co) variances between animals due to genetic relationships by formation of a matrix $\mathbf{A}$ based on pedigree records (consisted of 1252 final individuals) and e_{ijk}, distributed as $\sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{I}\sigma_e^2)$ is the random residual effect.

To test the association of each SNP with the different variables a univariate model was fitted using Eq. (1) to adjust the phenotypes for environmental and genetic effects. Once the corrected phenotypes were estimated, associations between each individual SNP and the corrected phenotype for AM, MAX and TOP3 were calculated using a general linear model in PLINK by the following equation:

$$y_i^* = \mu + \mathrm{SNP}_i + e_i$$

where y_i^* is the adjusted phenotype (corresponding to AM, MAX and TOP3); μ is the overall mean; SNP_i is the fixed effect of SNP genotype; and e_i is the random residual. To account for the risk of false positives due to the multiple testing problem, a false discovery rate (FDR) was used. Using the package "q value" in the statistical program R (Storey and Tibshirani, 2003) the q-values were calculated. A q-value threshold of 0.20 was set.

3. Results

Aspects of SNP markers across all *Bos taurus* chromosomes are summarized in Table 1. The longest BTA was 1 with 161.1 Mb and the shortest one was BTA25 with 44.1 Mb of total length. A total of 383 SNP markers did not have a defined position in the genome based on UMD3.1 assembly, and were not used for analysis. BTA1 showed the highest number of SNP markers with 2461 SNP and BTA27 had the lowest number of SNP markers, with 699. Each chromosome exhibited different markers density, varying from 17.10 SNP/Mb on BTA25 to 8.68 SNP/Mb on chromosome X, and an average density on all chromosomes of 14.86 SNP/Mb, approximately 1 SNP marker each 67 kb.

Table 1
Summary of analyzed SNP markers at each *B. taurus* chromosome[a].

Chr.	Size (Mb)	No. of SNP	SNP density (SNP/Mb)	Mean distance ± SD[b] (Kb)
Null	–	383	–	–
1	161.1	2,461	15.28	65.4 ± 67.8
2	140.8	1,989	14.13	69.9 ± 77.0
3	127.9	1,845	14.43	68.0 ± 96.6
4	124.5	1,792	14.39	67.0 ± 58.6
5	125.8	1,575	12.52	79.1 ± 113.6
6	122.6	1,831	14.93	65.0 ± 69.2
7	112.1	1,683	15.01	66.8 ± 76.8
8	116.9	1,788	15.30	63.6 ± 53.5
9	108.1	1,498	13.86	70.4 ± 65.9
10	106.4	1,600	15.04	65.1 ± 85.8
11	110.2	1,658	15.05	64.7 ± 59.1
12	85.4	1,243	14.56	73.2 ± 96.9
13	84.4	1,358	16.09	61.6 ± 52.4
14	81.3	1,338	16.46	62.2 ± 56.1
15	84.6	1,294	15.30	64.8 ± 61.1
16	77.9	1,227	15.75	66.3 ± 68.6
17	76.5	1,162	15.19	64.4 ± 67.4
18	66.1	1,020	15.43	64.2 ± 63.2
19	65.3	1,074	16.45	59.2 ± 52.1
20	75.8	1,153	15.21	62.5 ± 57.3
21	69.1	1,034	14.96	68.8 ± 69.4
22	61.8	933	15.10	65.7 ± 55.5
23	53.4	818	15.32	63.8 ± 59.4
24	65.3	948	14.52	66.4 ± 61.0
25	44.1	754	17.10	57.4 ± 45.9
26	51.8	809	15.62	63.0 ± 50.7
27	48.7	699	14.35	64.9 ± 65.3
28	46.1	736	15.97	62.8 ± 53.6
29	60.0	784	13.07	65.2 ± 68.4
X	88.5	768	8.68	193.8 ± 389.0
Overall	**2,642.5**	**39,255**	**14.86**	**62.0 ± 74.1**

[a] Based on UMD3.1.
[b] Between adjacent SNP.

Please cite this article as: Nani, J.P, et al., Genome-wide association study for somatic cell score in Argentinean dairy cattle. Livestock Science (2015), http://dx.doi.org/10.1016/j.livsci.2015.02.007

3.1. Linkage disequilibrium

The average distance (and standard deviation) between all adjacent SNP pairs across all chromosomes was 62.0 (74.1 s.d.) kb. Fifty two percent of the SNP were between 20 and 50 kb apart (20,078 SNP pairs) and only 17% of the SNP were more than 100 kb apart (6657 SNP pairs). The highest distance between 2 adjacent SNP was found in chromosome X, where the SNP pair was 4.7 Mb apart and the shortest was found in BTA13 with only 65 bp of separation.

As expected, the average r^2 decreased rapidly with increasing genetic distance, showing an exponential trend. High LD values were observed at smaller distances between SNPs (Fig. 1). At marker distances less than 10 kb the average r^2 was 0.40. Between 40 and 50 kb the average r^2 was 0.25 and for 100 kb apart the average r^2 was 0.18.

Table 2 shows a summary of the LD among adjacent SNP. The mean r^2 ($\pm$SD) between adjacent SNP ranged from 0.20 ± 0.25 for BTA27 to 0.28 ± 0.29 for BTA10 and the overall average was 0.25 ± 0.28. BTA1 had the highest percentage (6.38%) of adjacent SNP with $r^2 = 1$ and BTA25 had the lowest percentage (2.92%); 4.92% being the genome-wide average. The longest distance with total LD ($r^2 = 1$) between two adjacent SNP was found on BTA10 spanning 5.77 Mb, and the shortest was found on BTA25 with 0.70 Mb.

3.2. Phenotypic data and GWAS

Mean standard deviation, coefficient of variation, maximum and minimum values and heritability for each variable are listed in Table 3. The estimated genetic and phenotypic correlations between all traits are listed in Table 4. A significant correlation was found between TOP3 and MAX.

The GWAS performed with the three different variables and a false discovery rate ≤ 0.20 showed that AM was significantly associated ($p < 0.000015$) with two SNP: ARS-BFGL-NGS-114608 located at position 37,909,243 bp (UMD3.1) on BTA1 and Hapmap60306-rs29023088 located at 60,665,676 bp on BTA5. MAX and TOP3 were significantly associated with four SNP: ARS-BFGL-NGS-107594 (position 40,460,414 bp), ARS-BFGL-NGS-104220 (position 42,555,055 bp) located on BTA10, BTA-43543-no-rs (position 46,124,864 bp) on BTA18, and ARS-BFGL-NGS-109705 (position 8129,819 bp) on BTA26 (Table 5).

In agreement with the phenotypic and genetic correlations found, the association study of MAX and TOP3 had

Table 2
Summary of adjacent SNP markers at each *B. taurus* chromosome[a].

Chr.	SNP[b]	Max distance[c]	Average $r^2 \pm$ SD	$r^2 = 1$ ratio[d] (%)
1	156	1.75	0.25 ± 0.29	6.38
2	105	1.53	0.25 ± 0.29	5.29
3	94	2.32	0.25 ± 0.28	5.10
4	107	2.59	0.25 ± 0.29	5.98
5	82	3.69	0.26 ± 0.29	5.22
6	98	1.53	0.26 ± 0.28	5.37
7	82	1.30	0.27 ± 0.29	4.88
8	101	1.27	0.26 ± 0.29	5.66
9	64	1.83	0.24 ± 0.28	4.29
10	89	5.77	0.28 ± 0.29	5.58
11	84	1.31	0.25 ± 0.29	5.09
12	37	1.51	0.22 ± 0.26	2.98
13	52	1.60	0.25 ± 0.28	3.83
14	64	1.12	0.28 ± 0.30	4.80
15	49	1.31	0.21 ± 0.26	3.80
16	65	1.60	0.25 ± 0.29	5.32
17	50	1.24	0.24 ± 0.28	4.31
18	44	1.10	0.23 ± 0.27	4.32
19	52	1.90	0.22 ± 0.27	4.86
20	52	1.14	0.25 ± 0.28	4.53
21	53	1.82	0.26 ± 0.29	5.17
22	42	1.62	0.24 ± 0.28	4.54
23	46	2.04	0.21 ± 0.25	5.63
24	55	1.52	0.25 ± 0.29	5.81
25	22	0.70	0.22 ± 0.26	2.92
26	49	1.08	0.25 ± 0.29	6.08
27	21	0.93	0.20 ± 0.25	3.01
28	22	2.09	0.22 ± 0.26	3.03
29	27	1.00	0.22 ± 0.26	3.45
X	38	1.00	0.23 ± 0.28	5.54
Overall	1902	1.71 ± 0.96	0.25 ± 0.28	4.92

[a] Based on UMD3.1.
[b] Number of adjacent SNP with $r^2 = 1$.
[c] Longest distance in Mb between adjacent SNP with $r^2 = 1$.
[d] $N_{(r^2=1)}$/SNP.

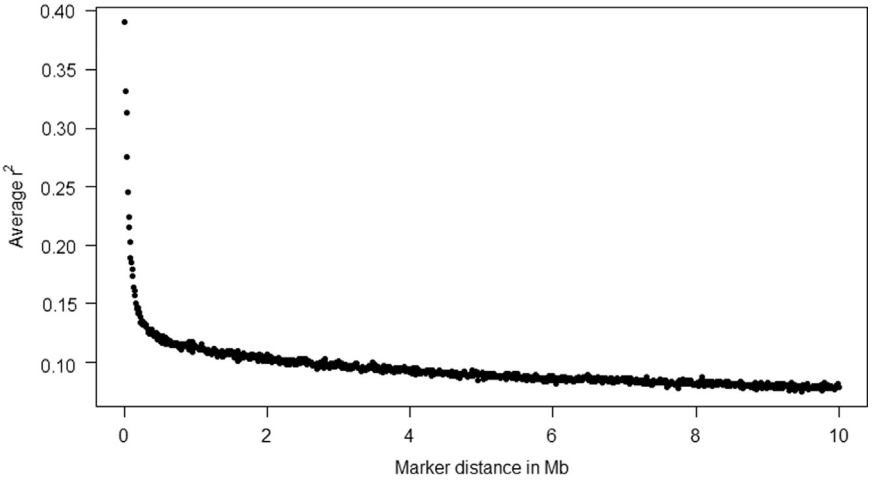

Fig. 1. Average r^2 over genetic distance for all *B. taurus* chromosomes every 10 kb to 10 Mb.

Please cite this article as: Nani, J.P, et al., Genome-wide association study for somatic cell score in Argentinean dairy cattle. Livestock Science (2015), http://dx.doi.org/10.1016/j.livsci.2015.02.007

Table 3
Mean standard deviation, coefficient of variation, maximum, minimum and heritability values for each variable.

Variable[a]	Mean	SD	CV	Max value	Min value	h^2 (SD)
TOP3	5.10	1.14	0.22	9.02	1.33	0.03 (0.01)
MAX	5.83	1.32	0.23	9.00	2.01	0.02 (0.01)
AM	4.05	1.04	0.26	7.00	1.00	0.04 (0.02)

[a] Variables calculated from the SCS per lactation: the arithmetic mean (AM), the maximum value (MAX) and the arithmetic mean of the top 3 values (TOP3).

Table 4
Genetic (above diagonal) and phenotypic (below) correlations (SE) between all traits.

	TOP	MAX	AM
TOP	–	**0.89(0.06)**	0.49(0.02)
MAX	**0.83(0.03)**	–	0.37(0.01)
AM	0.69(0.04)	0.53(0.02)	–

Table 5
SNP markers significantly associated with the three variables[a] calculated from SCS.

Variable[a]	SNP	BTA	Location[b] (bp)	*p* Value
AM	ARS-BFGL-NGS-114608	1	37,909,243	5.16E−06
AM	Hapmap60306-rs29023088	5	60,665,676	8.20E−06
TOP3	ARS-BFGL-NGS-107594	10	40,460,414	1.43E−06
TOP3	ARS-BFGL-NGS-104220	10	42,555,055	7.07E−06
TOP3	BTA-43543-no-rs	18	46,124,864	9.04E−06
TOP3	ARS-BFGL-NGS-109705	26	8129,819	1.02E−05
MAX	ARS-BFGL-NGS-107594	10	40,460,414	2.00E−06
MAX	ARS-BFGL-NGS-104220	10	42,555,055	4.00E−06
MAX	ARS-BFGL-NGS-109705	26	8129,819	8.00E−06
MAX	BTA-43543-no-rs	18	46,124,864	1.10E−05

[a] Variables calculated from the SCS per lactation: the arithmetic mean (AM), the maximum value (MAX) and the arithmetic mean of the top 3 values (TOP3).
[b] Based on UMD3.1

the same four SNP significantly associated (ARS-BFGL-NGS-107594, ARS-BFGL-NGS-104220, ARS-BFGL-NGS-109705 and BTA-43543-no-rs) (Fig. 2).

Five of the six SNP (all except ARS-BFGL-NGS-109705) found to be significantly associated with the different variables, presented homozygous genotypes as the most common, with genotype frequency from 0.65 to 0.87) and MAF from 0.07 to 0.2. In the same way, these five SNP had the favorable alleles as most prevalent, ranging from 0.80 to 0.93, showing a negative value in the SNP effect (Table 6). The untransformed SCC values for the favorable genotypes corresponded up to 25,000 cell/mL lower compared with those of the oposite homozygous genotypes (case of AA genotypes for ARS-BFGL-NGS-104220 and BTA-43543-no-rs).

4. Discussion

This study identified six SNP associated with traits derived from SCC, which is highly correlated with mastitis. In the past years several authors have studied different variables derived from SCC to improve the possibilities of genetic progress in selection against mastitis (Green et al., 2004; de Haas et al., 2008; Urioste et al., 2010, 2012; Koeck et al., 2012). In this work, we used three traits derived from the SCS: AM, TOP3 and MAX. From these, AM, has been used for many years to capture the variation over the lactation period of cows. MAX was also used in several studies and was correlated with clinical cases of mastitis (Green et al., 2004; Koeck et al., 2012). More recently, the log-transformed standard deviation from SCC (SCS-SD) was suggested to capture peaks in the SCC, and was related to clinical mastitis (Green et al., 2004; Urioste et al., 2010, 2012). Furthermore, it was recently used in a GWAS (Wijga et al., 2012). To our knowledge, TOP3 (the arithmetic mean of the top 3 values from the log-transformed SCC) was never used before this work.

In the analyzed dataset, MAX and TOP3 had a high genetic correlation (0.89), and seemed to capture the same variation in the SCS. They were also significantly associated with the same four SNP. As expected, all traits showed a low heritability (0.02–0.04) as showed by different studies (de Haas et al., 2008; Urioste et al., 2010; Koeck et al., 2012). The variables used reflect two different genetic aspect of SCS variations, AM reflects more the length and number of periods with elevated SCC (since a higher lactation SCS mean is probably due to recurrent infections or a constant higher count of cells), whilst TOP3 and MAX reflects only the presence and height of peaks in the SCC, only one in the case of MAX and possibly several in the case of TOP3. This difference was also reflected in the results of the association study, giving evidence that the traits used reveal different aspects of the response to infection or at least that different loci involved could be distinguished.

Linkage disequilibrium decays rapidly over short distances, and remains slightly over zero for longer distances. According to Meuwissen et al. (2001) a minimum level of LD for genomic selection should be $r^2=0.2$ in order to achieve an accuracy of 0.85 when computing genomic breeding values. In this study, this value was needed in order for SNP markers to pick up effects of causal variants between two markers. In our population this level of LD ($r^2=0.2$) was observed in SNP pairs separated at approximately up to 82,000 bp. According to the level of LD found in this population, a minimum of approximately 36,500 markers would be required for an association study (assuming that a causative variation or QTL is located in the middle of two SNP) as well as to perform genomic selection. Taking into account that up to 82,000 bp apart the LD is still at a value of 0.2, we considered genes within this region as possible candidates.

The window surrounding ARS-BFGL-NGS-114608 (BTA1) according to the Ensembl database (http://www.ensembl.org/Bos_taurus/) contains 4 genes: PROS1 (protein S alpha), ARL13B (ADP-ribosylation factor-like 13B), NSUN3 (NOP2/Sun domain family, member 3) and STX19 (syntaxin 19). The PROS1 gene encodes a protein called protein S (vitronectin), has an important role in blood clotting control

Please cite this article as: Nani, J.P, et al., Genome-wide association study for somatic cell score in Argentinean dairy cattle. Livestock Science (2015), http://dx.doi.org/10.1016/j.livsci.2015.02.007

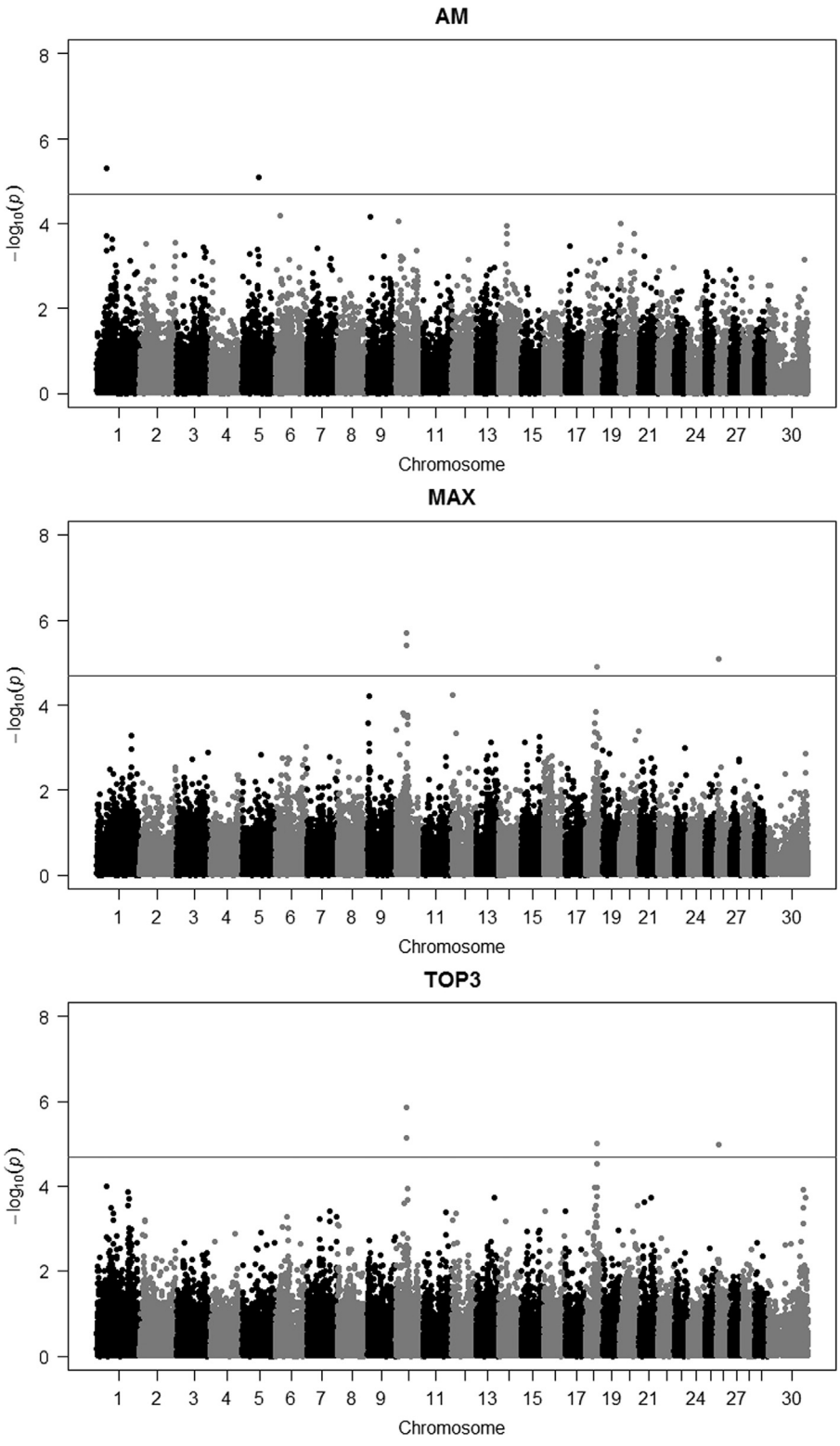

Fig. 2. Manhattan plot for the −log10 of the *p* values from all SNP for each variable. Chromosomes arranged from left to right from BTA1 to 30 (sex chromosome X). Above the horizontal line are the SNP that passed the FDR threshold of 0.20.

Please cite this article as: Nani, J.P, et al., Genome-wide association study for somatic cell score in Argentinean dairy cattle. Livestock Science (2015), http://dx.doi.org/10.1016/j.livsci.2015.02.007

Table 6
Effect sizes (SD) per genotype class and MAF for the SNP significantly associated with the different variables calculated from SCS.

SNP	BTA	MA (F)	Variable[a]	*B*/*B*	*B*/*A*	*A*/*A*
ARS-BFGL-NGS-114608	1	*B*(0.20)	AM	0.65 (0.42)	0	−0.11 (0.08)
Genotype frequency				0.05	0.31	0.65
Hapmap60306-rs29023088	5	*A*(0.14)	AM	0.22 (0.20)	0	−0.27 (0.21)
Genotype frequency				0.01	0.26	0.73
ARS-BFGL-NGS-107594	10	*B*(0.07)	TOP3	0.79 (0.61)	0	−0.53 (0.16)
			MAX	0.72 (0.69)	0	−0.67 (0.44)
Genotype frequency				0	0.13	0.87
ARS-BFGL-NGS-104220	10	*B*(0.07)	TOP3	2.04 (0.68)	0	−0.40 (0.41)
			MAX	1.51 (1.05)	0	−0.53 (0.36)
Genotype frequency				0	0.13	0.87
BTA-43543-no-rs	18	*B*(0.07)	TOP3	0.40 (0.15)	0	−0.05 (0.15)
			MAX	1.52 (0.42)	0	−0.52 (0.27)
Genotype frequency				0	0.13	0.87
ARS-BFGL-NGS-109705	26	*A*(0.45)	TOP3	−0.04 (0.25)	0	0.56 (0.23)
			MAX	−0.17 (0.05)	0	0.53 (0.49)
Genotype frequency				0.29	0.52	0.18

[a] Variables calculated from the SCS per lactation: the arithmetic mean (AM), the maximum value (MAX) and the arithmetic mean of the top 3 values (TOP3).

(Castoldi and Hackeng, 2008). Filippsen et al. (1990) investigated the binding of bovine complement S protein to *Streptococcus dysgalactiae* isolates from cattle with mastitis, and the S protein's role in bacteria adherence to bovine epithelial cells. They found that bovine S protein could be an important mediator of adherence of *S. dysgalactiae* to bovine epithelial cells. In addition, it has been demonstrated that vitronectin augments ingestion of opsonized bacteria by monocytes through stimulation of cell receptors, thus supporting the host immune system (Preissner, 1989).

The window around the marker Hapmap60306-rs29023088 on BTA5 contained 5 genes: SNRPF (small nuclear ribonucleoprotein F), CCDC38 (coiled-coil domain containing 38), AMDHD1 (amidohydrolase domain containing 1), LTA4H (Leukotriene A-4 hydrolase) and HAL (histidine ammonia-lyase). Within this region Ashwell et al. (2004) positioned a QTL affecting SCS. The LTA4H codes for the Leukotriene A-4 hydrolase, an epoxy hydrolase that catalyzes the final step in the biosynthesis of the proinflammatory mediator leukotriene B4 (Haeggström et al., 1990). Leukotriene B4 is a potent chemotactic agent, known to induce recruitment and accumulation of neutrophils in the bovine mammary gland (Boutet et al., 2003).

Within the region surrounding the marker ARS-BFGL-NGS-107594 (BTA10) Boichard et al. (2003) located a QTL affecting SCS in French dairy cattle breeds (including Holstein). No genes were located within the 82,000 bp window around ARS-BFGL-NGS-104220; however, Kühn et al. (2003) positioned a QTL affecting SCC at about 1 cM away.

For marker BTA-43543-no-rs (BTA18), the window contained 10 genes: FXYD1/3/5/7 (domain-containing ion transport regulator), LIG4 (ATP-dependent ligase IV), FAM187B (family with sequence similarity 187, member B), MAG (myelin associated glycoprotein), LSR (lipolysis stimulated lipoprotein receptor), USF2 (upstream transcription factor 2) and HAMP (hepcidin antimicrobial peptide). There are many previous studies positioning QTL affecting SCS spanning this region: Baes et al. (2009) and Brand et al. (2009) in German Holstein cattle; Schrooten et al. (2000) in Dutch Holstein cattle. A recent GWAS study (Abdel-Shafy et al., 2014) narrowed the region to 0.05 Mb, and positioned the QTL 3 Mb downstream the SNP BTA-43543-no-rs. Research studies by Lund et al. (2007) also positioned a QTL in this region, affecting SCS in Finnish Ayrshire, Swedish Red and White, and Danish Red dairy cattle breeds. Among the 10 genes found within this region only the HAMP gene appears to have a function that could be related to mastitis. The HAMP gene product is a Cysteine-rich protein abundant in different animal tissues with antimicrobial properties that is involved in host defense against bacteria (Park et al., 2000).

Within the 82 kb window around ARS-BFGL-NGS-109705 marker (BTA26) only one gene is located: PRKG1 (protein kinase, cGMP-dependent, type I). Spanning this region there is a large QTL positioned by Ashwell et al. (2004) affecting SCS in American Holstein cattle. The PRKG1 protein is a serine/threonine protein kinase that acts as key mediator of the nitric oxide (NO)/cGMP signaling pathway (Butt et al., 1993). This protein is also involved in the Ca^{2+} signaling, which has a direct role in apoptosis induction. Proteins that are phosphorylated by PRKG1 regulate platelet activation and adhesion as well as smooth muscle contraction (Lohmann et al., 1997).

Several studies conducting GWAS with traits related to mastitis were published recently. Meredith et al. (2013) and Abdel-Shafy et al. (2014) used DYD (Daughter Yield Deviations) for SCS as the phenotype to perform GWAS. Sodeland et al. (2011) also used DYD, but in this case for clinical mastitis and SCS. Only two regions on BTA13 and BTA20 can be considered similar between the report of Meredith et al. (2013) and Sodeland et al. (2011), since they are only 1 Mb apart. However, most of the reported associations, including those we report in the present work, correspond to different genome locations. The closest SNP between our results compared with previous ones, is BTA-43543-no-rs located approximately 3 Mb away from a SNP detected by Abdel-Shafy et al. (2014) on BTA18. This is close enough to overlap with previous detected QTLs, but the region of 0.05 Mb claimed by Abdel-Shafy et al. (2014) for their haplotype-based study do not overlap the SNP detected in this study.

Wijga et al. (2012) conducted a GWAS using lactation average SCS and SCS-SD based on first lactation test-day records. Unlike our results, Wijga et al. (2012) reported a

Please cite this article as: Nani, J.P, et al., Genome-wide association study for somatic cell score in Argentinean dairy cattle. Livestock Science (2015), http://dx.doi.org/10.1016/j.livsci.2015.02.007

high correlation between the studied traits SCS-SD and lactation average SCS, and found the same SNP associated with both variables. In this study, different traits showed different SNP association, and this can help to a better understand the genetic basis of mastitis resistance.

5. Conclusions

This is the first study to collect phenotypic and genotypic information of dairy herds in Argentina. The results of the LD analysis and GWAS contributed to a better understanding of the genomic architecture of Argentinean dairy herds and also gain insight into the genomic regions influencing the SCS. Comparing the three variables calculated from the SCS, MAX and TOP3 behaved similarly, having high genetic and phenotypic correlations that indicating that these variables were equivalent in this GWAS. Moreover, AM on one side and TOP3 and MAX on the other, had different SNPs associated, which showed that they capture different aspects of mastitis response. As expected, no very large QTL effects were detected, showing that the QTLs affecting these variables are dispersed throughout the genome, each with a relatively small effect. While more studies are necessary, this research contributes to the knowledge of the mechanisms of resistance to mastitis to be implemented in breeding programs.

6. Conflict of interest

There are no conflicts of interest.

Acknowledgments

This work was supported by Agencia Nacional de Promoción Científica y Tecnológica (PAE 37143 and PICT 01523), Instituto Nacional de Tecnología Agropecuaria (INTA) PNLEC-071051, INTA AEBIO-245711 and INTA PNBIO-1131033.

References

Abdel-Shafy, H., Bortfeldt, R.H., Tetens, J., Brockmann, G.A., 2014. Single nucleotide polymorphism and haplotype effects associated with somatic cell score in German Holstein cattle. Genet. Sel. Evol. 46, 35.

Ashwell, M.S., Heyen, D.W., Sonstegard, T.S., Van Tassell, C.P., Da, Y., VanRaden, P.M., Ron, M., Weller, J.I., Lewin, H.A., 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. J. Dairy Sci. 87, 468–475.

Baes, C., Brand, B., Mayer, M., Kühn, C., Liu, Z., Reinhardt, F., Reinsch, N., 2009. Refined positioning of a quantitative trait locus affecting somatic cell score on chromosome 18 in the German Holstein using linkage disequilibrium. J. Dairy Sci. 8, 4046–4054.

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M.Y., Levéziel, H., 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genet. Sel. Evol. 35, 77–101.

Boutet, P., Bureau, F., Degand, G., Lekeux, P., 2003. Imbalance between lipoxin A4 and leukotriene B4 in chronic mastitis-affected cows. J. Dairy Sci. 86, 3430–3439.

Brand, B., Baes, C., Mayer, M., Reinsch, N., Kühn, C., 2009. Identification of a two-marker-haplotype on Bos taurus autosome 18 associated with somatic cell score in German Holstein cattle. BMC Genet. 10, 50.

Browning, B.L., Browning, S.R., 2009. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84, 210–223.

Butt, E., Geiger, J., Jarchau, T., Lohmann, S.M., Walter, U., 1993. The cGMP-dependent protein kinase—gene, protein, and function. Neurochem. Res. 18, 27–42.

Castoldi, E., Hackeng, T.M., 2008. Regulation of coagulation by protein S. Curr. Opin. Hematol. 15, 529–536.

de Haas, Y., Ouweltjes, W., ten Napel, J., Windig, J.J., de Jong, G., 2008. Alternative somatic cell count traits as mastitis indicators for genetic selection. J. Dairy Sci. 91 (6), 2501–2511.

de Roos, A.P.W., Hayes, B.J., Spelman, R.J., Goddard, M.E., 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. Genetics 179, 1503–1512.

Dekkers, J.C., 2004. Commercial application of marker- and gene assisted selection in livestock: strategies and lessons (E. Suppl.). J. Anim. Sci. 82, E313–E328.

Elsik, C.G., Tellam, R.L., Worley, K.C., The Bovine Genome Sequencing and Analysis Consortium., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324 (5926), 522–528.

Filippsen, L.F., Valentin-Weigand, P., Blobel, H., Preissner, K.T., Chhatwal, G.S., 1990. Role of complement S protein (vitronectin) in adherence of Streptococcus dysgalactiae to bovine epithelial cells. Am. J. Vet. Res. 51, 861–865.

Goddard, M.E., Hayes, B.J., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programs. Nat. Rev. Genet. 10, 381–391.

Green, M.J., Green, L.E., Schukken, Y.H., Bradley, A.J., Peeler, E.J., Barkema, H.W., de Haas, Y., Collis, V.J., Medley, G.F., 2004. Somatic cell count distributions during lactation predict clinical mastitis. J. Dairy Sci. 87 (5), 1256–1264.

Haeggström, J.Z., Wetterholm, A., Vallee, B.L., Samuelsson, B., 1990. Leukotriene A4 hydrolase: an epoxide hydrolase with peptidase activity. Biochem. Biophys. Res. Commun. 173, 431–437.

Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226–231.

Hirschhorn, J.N., Daly, M.J., 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95–108.

Koeck, A., Miglior, F., Kelton, D.F., Schenkel, F.S., 2012. Alternative somatic cell count traits to improve mastitis resistance in Canadian Holsteins. J. Dairy Sci. 95 (1), 432–439.

Kühn, C.H., Bennewitz, J., Reinsch, N., Xu, N., Thomsen, H., Looft, C., Brockmann, G.A., Schwerin, M., Weimann, C., Hiendleder, S., Erhardt, G., Medjugorac, I., Förster, M., Brenig, B., Reinhardt, F., Reents, R., Russ, I., Averdunk, G., Blümel, J., Kalm, E., 2003. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. J. Dairy Sci. 86, 360–368.

Lohmann, S.M., Vaandrager, A.B., Smolenski, A., Walter, U., De Jonge, H.R., 1997. Distinct and specific functions of cGMP-dependent protein kinases. Trends Biochem. Sci. 22, 307–312.

Lund, M.S., Sahana, G., Andersson-Eklund, L., Hastings, N., Fernandez, A., Schulman, N., Thomsen, B., Viitala, S., Williams, J.L., Sabry, A., Viinalass, H., Vilkki, J., 2007. Joint analysis of quantitative trait loci for clinical mastitis and somatic cell score on five chromosomes in three Nordic dairy cattle breeds. J. Dairy Sci. 90, 5282–5290.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4, 4.

Meredith, B.K., Berry, D.P., Kearney, F., Finlay, E.K., Fahey, A.G., Bradley, D. G., Lynn, D.J., 2013. A genome-wide association study for somatic cell score using the Illumina high-density bovine beadchip identifies several novel QTL potentially related to mastitis susceptibility. Front. Genet. 4, 229.

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

Meyer, K., 2007. WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J. Zhejiang Univ. Sci. B 11, 815–821.

Miglior, F., Muir, B.L., Van Doormaal, B.J., 2005. Selection indices in Holstein cattle of various countries. J. Dairy Sci. 88, 1255–1263.

Park, C.H., Valore, E.V., Waring, A.J., Ganz, T., 2000. Hepcidin, a urinary antimicrobial peptide synthesized in the liver. J. Biol. Chem. 276, 7806–7810.

Preissner, K.T., 1989. The role of vitronectin as multifunctional regulator in the hemostatic and immune systems. Blut 59, 419–431.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Sargolzaei, M., Schenkel, F.S., Jansen, G.B., Schaeffer, L.R., 2008. Extent of linkage disequilibrium in Holstein cattle in North America. J. Dairy Sci. 91, 2106–2117.

Schrooten, C., Bovenhuis, H., Coppieters, W., van Arendonk, J.A., 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. J. Dairy Sci. 83, 795–806.

Schukken, Y.H., Wilson, D.J., Welcome, F., Garrison-Tinofsky, L., Gonzales, R.N., 2003. Monitoring udder health and milk quality using somatic cell counts. Vet. Res. 34, 579–596.

Seegers, H., Fourichon, C., Beaudeau, F., 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. Vet. Res. 34, 475–491.

Shook, G. E., 1982, Approaches to summarizing somatic cell count which improve interpretability. In: Proc. Nat'l. Mastitis Council, pp. 150-166.

Sodeland, M., Kent, M.P., Olsen, H.G., Opsal, M.A., Svendsen, M., Sehested, E., Hayes, B.J., Lien, S., 2011. Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. Anim. Genet. 42 (5), 457–465.

Storey, J.D., Tibshirani, R., 2003. Statistical significance for genome-wide experiments. Proc. Natl. Acad. Sci. U.S.A. 100, 9440–9445.

Urioste, J.I., Franzén, J., Strandberg, E., 2010. Phenotypic and genetic characterization of novel somatic cell count traits from weekly or monthly observations. J. Dairy Sci. 93 (12), 5930–5941.

Urioste, J.I., Franzén, J., Windig, J.J., Strandberg, E., 2012. Genetic relationships among mastitis and alternative somatic cell count traits in the first 3 lactations of Swedish Holsteins. J. Dairy Sci. 95 (6), 3428–3434.

Wijga, S., Bastiaansen, J.W., Wall, E., Strandberg, E., de Haas, Y., Giblin, L., Bovenhuis, H., 2012. Genomic associations with somatic cell score in first-lactation Holstein cows. J. Dairy Sci. 95 (2), 899–908.

Please cite this article as: Nani, J.P, et al., Genome-wide association study for somatic cell score in Argentinean dairy cattle. Livestock Science (2015), http://dx.doi.org/10.1016/j.livsci.2015.02.007