

RESEARCH

Open Access



# Single nucleotide polymorphisms in candidate genes associated with milk yield in Argentinean Holstein and Holstein x Jersey cows

María Agustina Raschia<sup>1\*</sup>, Juan Pablo Nani<sup>2</sup>, Daniel Omar Maizon<sup>3</sup>, María José Beribe<sup>4</sup>, Ariel Fernando Amadio<sup>2,5</sup> and Mario Andrés Poli<sup>1</sup>

## Abstract

**Background:** Research on loci influencing milk production traits of dairy cattle is one of the main topics of investigation in livestock. Many genomic regions and polymorphisms associated with dairy production have been reported worldwide. In this context, the purpose of this study was to identify candidate loci associated with milk yield in Argentinean dairy cattle. A database of candidate genes and single nucleotide polymorphisms (SNPs) for milk production and composition was developed. Thirty-nine SNPs belonging to 22 candidate genes were genotyped on 1643 animals (Holstein and Holstein x Jersey). The genotypes obtained were subjected to association studies considering the whole population and discriminating the population by Holstein breed percentage. Phenotypic data consisted of milk production values recorded during the first lactation of 1156 Holstein and 462 Holstein x Jersey cows from 18 dairy farms located in the central dairy area of Argentina. From these records, 305-day cumulative milk production values were predicted.

**Results:** Eight SNPs (rs43375517, rs29004488, rs132812135, rs137651874, rs109191047, rs135164815, rs43706485, and rs41255693), located on six *Bos taurus* autosomes (BTA4, BTA6, BTA19, BTA20, BTA22, and BTA26), showed suggestive associations with 305-day cumulative milk production (under Benjamini-Hochberg procedure with a false discovery rate of 0.1). Two of those SNPs (rs43375517 and rs135164815) were significantly associated with milk production (Bonferroni adjusted *p*-values < 0.05) when considering the Holstein population.

**Conclusions:** The results obtained are consistent with previously reported associations in other Holstein populations. Furthermore, the SNPs found to influence bovine milk production in this study may be used as possible candidate SNPs for marker-assisted selection programs in Argentinean dairy cattle.

**Keywords:** Milk production, Dairy cattle, Candidate genes, Association study, Single nucleotide polymorphisms

## Background

Within Argentina's economy, the dairy industry is considered one of the most important and dynamic agri-food sector. Its structure is the result of a process of concentration and specialization of several years, with a decrease in the number of dairy farms and an increase in their productive scale [1]. Currently, this country has

approximately 1.8 million dairy cows distributed in 11,750 productive units [2]. The average milk production for the period 2011–2015 was of 11,167.9 million of liters/year [3]. From the total milk production, approximately 7.5% is sold through the informal market and/or consumed by farm households, while 92.5% is processed as fluid milk or as manufactured dairy products [4]. As production capacity far exceeds the volume required to meet domestic demand, between 15 and 25% of the total milk production is sold to external markets [3]. Dairy production system is primarily based on extensive grazing; however, an increased use of supplements is giving

\* Correspondence: [raschia.maria@inta.gov.ar](mailto:raschia.maria@inta.gov.ar)

<sup>1</sup>Instituto Nacional de Tecnología Agropecuaria (INTA), Centro de Investigación en Ciencias Veterinarias y Agronómicas (CICVyA), Instituto de Genética "Ewald A. Favret", Nicolás Repetto y de los Reseros s/n, Hurlingham B1686, Argentina

Full list of author information is available at the end of the article



rise to a mixed system. Regarding current breeding strategies, there is no national scheme with a single objective but each producer follows its own strategy. Furthermore, at present, no breeding strategies using genomic information are being applied at national level.

Most applications of genetic information in selection programs are preceded by analyses aimed at quantitative trait loci (QTL) detection, which should be conducted in the populations that are going to be used for genetic improvement [5, 6]. Only QTLs that are shown to have a significant effect on phenotype are subsequently used for selection [5]. In this context, extensive genetic research on lactation and milk production in cattle has been performed. Consequently, several countries implemented in their breeding programs or in their genetic evaluation systems the information made available on many candidate regions, QTLs, and genome-wide molecular markers, for genetic improvement on milk production traits. France [7], Germany [8] and New Zealand [9] applied marker-assisted selection programs in dairy cattle breeding. Furthermore, genomic selection strategies were implemented in the United States, Canada, Great Britain, Ireland, New Zealand, Australia, France, the Netherlands, Germany, and the Scandinavian countries [10].

The criteria for selecting candidate regions, genes, and markers combine information from different sources that support the candidate status of these regions. Key genes responsible for milk production or composition traits should be identified on the basis of a multidisciplinary approach combining different pieces of evidence such as gene location, gene expression profile, regulation of gene expression, function of coded protein, and previously reported associations between markers on the gene and the phenotypes under study. SNPs within genes directly, indirectly, or potentially related to those traits are convenient markers to use for the identification of causal loci.

Over the last decades, the advances in DNA technology have given rise to many high throughput SNP genotyping platforms, leading to automation and a significant reduction in costs. These advantages, along with the abundance and distribution of SNPs in the genome, facilitate studies on the statistical genetic association of neighboring alleles or linkage disequilibrium, with the objective to identify the location of genes influencing the variation in certain traits. Furthermore, SNPs are also important intrinsic candidates as causal variants of these traits [11–13].

To date, no studies in Argentinean Holstein and Holstein x Jersey cows have reported associations between genetic markers and milk production. The objective of this study was to detect associations between SNP markers in candidate genes suspected to influence milk production and composition traits, and

305-day cumulative milk production, predicted for a representative population of Argentinean dairy cattle. The information obtained aims to improve milk production breeding programs.

## Methods

### Animals, blood samples, and phenotypic records

The animals used belonged to 18 commercial dairy farms owned by a single dairy company located in the central dairy area of Argentina. Blood samples were taken from 1618 cows comprising 1156 Holstein and 462 Holstein x Jersey crosses, and semen samples were obtained from 25 bulls (20 Holstein and 5 Jersey).

The phenotypic records consisted of data from the official dairy control obtained during the first lactation of the cows under study. All the animals were kept in similar feeding and sanitary conditions. The cows were machine milked twice a day and milk records were taken every ~40 days. Every cow included in this study had at least four dairy records during their first lactation. Procedures for blood samples collection and for milk production evaluation were approved by the Institutional Committee for Care and Use of Experimental Animals (CICUAE) of the National Institute of Agricultural Technology (INTA) under protocol number 35/2010 and were carried out in strict accordance of the guidelines specified in the institutional manual.

### DNA extraction

Fresh blood was obtained from the jugular vein of cows using EDTA as anticoagulant. Genomic DNA was extracted from blood samples using a commercial kit (AxyPrep Blood Genomic DNA Miniprep Kit, Axygen Biosciences, Union City, CA, USA), following the manufacturer's protocol. Genomic DNA from sires was obtained from semen straws using a standard phenol-chloroform extraction. DNA quantity and quality were assessed by measuring DNA absorbance at 260 nm and evaluating 260/280 and 260/230 absorbance ratios, respectively, using a NanoDrop™ 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

### Gene and marker selection

The candidate genes were selected by combining different sources of information available on databases and public literature. The chosen genes were those that met at least one of the following criteria, proposed by Ogorevc et al. [14]: 1) the gene encodes a protein directly associated with milk components or milk metabolism that exists in different genetic variants; 2) markers on the gene have proven association with milk traits; 3) the gene expression profile is related to the phenotype under study; 4) the gene is located within a cattle QTL for the studied phenotype; 5) miRNAs directed to the mRNA of the gene are

expressed in bovine mammary gland and regulate its expression post-transcriptionally.

Three strategies were employed for marker selection: 1) markers from candidate genes previously found to be associated with milk production or composition traits; 2) markers from genes not directly related to milk production, but significantly associated with the phenotype under study; 3) newly identified SNPs by sequencing candidate genes on the evaluated sire population.

Primers were designed to amplify regions in 14 candidate genes: *LEP*, *ABCG2*, *OPN*, *PPARGC1A*, *CSN1S1*, *CSN2*, *CSN3*, *LGB*, *DGATI*, *GH*, *GHR*, *PRLR*, *LTF*, and *PRL*, using BatchPrimer3 [15]. The amplified region was sequenced both to check the existence of known polymorphisms and to discover new ones. Additional file 1 indicates primer sequences, GenBank accession numbers, length of the amplicons, and optimal annealing temperatures used.

The PCR reactions were performed using the enzyme Paq5000<sup>®</sup> DNA Polymerase and its reaction buffer (Agilent Technologies, Stratagene, Santa Clara, CA, USA) following the manufacturer's instructions. Conditions for amplification were 94 °C for 3 min, followed by 35 cycles of 20 s at 94 °C, 20 s at an optimal annealing temperature, and 1 min at 72 °C. Reactions were ended with a 72 °C, 5 min final extension stage. Amplified fragments from sire DNA were visualized in agarose gels stained with ethidium bromide and then sequenced using BigDye<sup>®</sup> chemistry (Perkin-Elmer, Wellesley, MA, USA) according to the manufacturer's protocol, on an ABI3130xl sequencer (Applied Biosystems, Foster City, CA, USA). Sequencing reactions were conducted with the same primers as those used for PCR reactions, with a 1:10 dilution of the PCR product as template. Sequence traces assembly was performed with Gap4 [16]. The positions of the SNPs in the chromosomes were determined by mapping the SNPs including their 200 bp context against assembly UMD3.1 [17] using the Exonerate software [18].

### Genotyping and quality control

Forty-six markers from 23 candidate genes were taken into consideration for the SNPlex genotyping panel design. Considered candidate genes comprised *ARL4A* (ADP ribosylation factor like GTPase 4A), *ETVI* (ETS variant 1), *SNX13* (sorting nexin 13), *LEP* (leptin), *LALBA* (lactalbumin alpha), *OLRI* (oxidized low density lipoprotein receptor 1), *ABCG2* (ATP binding cassette subfamily G member 2), *OPN* (osteopontin), *PPARGC1A* (peroxisome proliferator-activated receptor gamma coactivator 1-alpha), *CSN1S1* (casein alpha s1), *CSN2* (casein beta), *CSN3* (casein kappa), *LGB* (beta-lactoglobulin), *DGATI* (diacylglycerol O-acyltransferase 1), *STAT5A* (signal transducer and activator of transcription 5A), *GH* (growth hormone), *FASN* (fatty acid synthase), *GHR* (growth hormone

receptor), *PRLR* (prolactin receptor), *UTMP* (uterine milk protein precursor), *LTF* (lactotransferrin), *PRL* (prolactin), and *SCD1* (stearoyl-CoA desaturase). Selection criteria required nucleotide substitutions in those genes to involve only two possible bases and to map to only one position in the genome. Forty-six SNPs were proposed for custom assay design. However, seven of the selected markers were not included in the panel because they failed to meet the manufacturer's specifications. As a result, the designed panel consisted of 39 markers from 22 candidate genes to be genotyped with the SNPlex Genotyping System platform (Applied Biosystems, Foster City, CA, USA; [19]) on the commercial population of Holstein and Holstein x Jersey cattle under study. Genotyping reactions and analyses were done following provider protocols using an ABI3130xl sequencer and GeneMapper v4.0 software (Applied Biosystems).

Call rate, as well as genotypic and allelic frequencies, were calculated for all genotyped SNPs. Genotyping quality assurance was performed using PLINK v1.07 [20]. Only the SNPs that satisfied the following criteria were retained: (a) minor allele frequency (MAF) > 5%, and (b) percentage of missing genotypes across all samples < 10%. After quality pruning, 22 SNPs were included in the analysis. Concerning animals, those with more than 10% missing genotypes across all SNPs were removed. Mendelian errors, accounting for all assigned genotypes that generate genotypic inconsistencies in sire-mother-daughter or sire-daughter familiar nucleus, were detected with PLINK and set to missing genotypes. As expected, loci in this population did not match Hardy-Weinberg postulates. Furthermore, SNPs associated with milk production might have been under selection pressure in the dairy population used and therefore they might not have been in Hardy-Weinberg equilibrium (HWE). For the reasons previously stated, no consideration was given to deviation from HWE in the quality control test.

### Prediction of 305-day cumulative milk production

Lactation curves were adjusted on the basis of days in milk production, through Legendre orthogonal polynomials of fourth degree [21]. All curves were treated as random regression models, which made it possible to adjust a lactation curve for each individual -random regression-, expressed as the deviation from an average population curve -steady regression- [22]. The 305-day cumulative milk production was predicted for 1618 cows with at least four production records during their first lactation (MJ Beribe, personal communication).

### Statistical analysis and association test

Associations between each individual SNP and the 305-day cumulative milk production of 1346 cows (those

that passed the genotype quality control criteria) were calculated using a general linear model in PLINK by the following equation:

$$Y_{ijklmn} = \mu + b_i + yb_j + herd_k + SYFL_l + sire_m + SNP_n + e_{ijklmn}$$

where  $y_{ijklmn}$  is the 305-day cumulative milk production previously predicted,  $\mu$  is the overall population mean,  $b_i$  is the breed fixed effect,  $yb_j$  is the year of birth fixed effect,  $herd_k$  is the farm fixed effect,  $SYFL_l$  is the combined fixed effect of season and year of first lactation,  $sire_m$  is the sire effect,  $SNP_n$  is the effect of the SNP genotype, and  $e_{ijklmn}$  is the random residual. The breed effect comprised five categories: 100, 75, 50, 25 and 12.5% of Holstein genetic background. Cows with production records were born between the years 2000 and 2006, belonged to 18 different herds, and were sired by 20 Holstein and 4 Jersey bulls. The fixed effect accounting for the combination of season and year of first lactation considered four seasons (fall, winter, spring, and summer) and five years (2004 to 2008). Association tests were conducted on the whole population (1346 cows) as well as on pure Holstein (978 animals) and on Holstein x Jersey crosses (368 cows) separately. The contribution of each covariate included in the model was checked by PLINK software for each association study performed, and only those found to have significant effects were considered.

To account for the risk of false positives due to the multiple testing problem,  $p$ -values were adjusted by Bonferroni correction and Benjamini-Hochberg procedure using PLINK. Bonferroni adjusted  $p$ -values  $< 0.05$  were accepted to represent a proof of significant association, while Benjamini-Hochberg adjusted  $p$ -values  $< 0.1$  were considered as indicators of suggestive associations.

Student's  $t$ -tests were conducted to compare the 305-day cumulative milk production of cows with different genotypes, on each SNP detected as significantly associated with the trait under study.

## Results

### Gene and marker selection and SNP panel design

Twenty-two candidate genes from bovine loci involved in milk production and composition traits were retrieved from literature [11–13, 23–41] and different open access databases (PubMed, Cattle QTL Database, and Ensembl). The collected data included loci on 11 chromosomes (BTAs 4, 5, 6, 11, 14, 19, 20, 21, 22, 23 and 26), with the highest number of candidate genes on BTA6. The final list of 39 SNPs included in the SNPlex panel is indicated in Table 1.

DNA sequencing of the sires allowed us to detect previously reported SNPs in bovine *LEP*, *ABCG2*, *OPN*,

*PPARGC1A*, *CSN1S1*, *CSN2*, *CSN3*, *LGB*, *DGAT1*, *GH*, *GHR*, *PRLR*, *LTF*, and *PRL* genes. However, we did not identify any new SNP in these regions.

### Genotyping

Five out of the 39 markers included in the panel were discarded (OPN3907, rs41255680, rs43691049, rs109487069, and rs208645216) because of genotyping errors/failure in the genotyping reactions. For the remaining 34 SNPs, the overall success rate, calculated as the ratio between genotype calls and genotyped loci, was 0.94. However, 12 markers were removed because they failed to pass call rate (rs109326954, rs109234250, rs41255685, rs41919985, rs109625649, CSN1\_AB, and rs109019599) or MAF (rs43702337, rs43706475, rs433385179, rs43703017, and rs43703013) thresholds. The SNPs rs43706475 and rs433385179 were monomorphic in the study population. Call rate, MAF, and HWE  $p$ -value for these 34 SNPs are indicated in Table 1. Additional file 2 shows the genotypic and allelic frequencies obtained for the 22 remaining SNPs, considered for subsequent analyses. Frequencies were discriminated for Holstein and Holstein x Jersey cows, without differentiating the degree in the cross.

An estimated 1.24% of the cows' assigned genotypes showed Mendelian inconsistencies with their parent genotypes and consequently were considered as missing genotypes in the subsequent analysis. After filtering on call rate, 1346 out of 1618 cows were conserved for the association test.

### Prediction of 305-day milk production

When the whole population was evaluated, the phenotypic mean and the standard deviation for first lactation 305-day cumulative milk production predictions were 5967.38 and 880.52 kg, respectively. Furthermore, considering Holstein and Holstein x Jersey cows separately, phenotypic means and standard deviations were 6191.27  $\pm$  806.76 and 5427.73  $\pm$  815.05 kg, respectively.

### Statistical analysis and association test

When performing the association study on the whole population, every effect considered in the linear mixed model used was significant. Even though no significant associations after Bonferroni correction were detected in this test, 5 out of the 22 SNPs assessed showed suggestive associations with the 305-day cumulative milk production (Table 2). Those associations were detected on bovine autosomes 4, 6, 19, and 26. The greatest number of associations was found on BTA4, with two associations (rs29004488 and rs43375517), while the other three chromosomes (BTA6, 19, and 26) revealed a single SNP associated with the trait (rs132812135, rs137651874, and rs41255693, respectively). These five SNPs correspond to

**Table 1** SNPs included in the SNPlex genotyping panel

Candidate gene	Marker	Location (BTA:bp)	Call rate	MAF	HWE <i>p</i> -value
<i>ARL4A</i>	<b>rs43375517</b>	4:20296999	98.1	0.39	0.14
<i>ETV1</i>	<b>rs42213673</b>	4:22111819	98.8	0.35	0.28
<i>SNX13</i>	<b>rs41595314</b>	4:26425057	98.6	0.07	8.10 <sup>-3</sup>
<i>LEP</i>	<b>rs29004488</b>	4:93262056	97.4	0.33	0.02
<i>OLR1</i>	rs109019599	5:100254823	89.3	0.25	2.10 <sup>-3</sup>
<i>ABCG2</i>	rs43702337	6:38027010	99.7	0.00	1.00
<i>OPN</i>	<b>rs132812135</b>	6:38120968	94.5	0.35	0.55
	<b>rs110930453</b>	6:38122665	93.1	0.42	0.87
	OPN3907 <sup>a</sup>	6:38128806			
<i>PPARGC1A</i>	<b>rs109579682</b>	6:44875251	98.4	0.23	0.03
	<b>rs133669403</b>	6:44875315	97.5	0.08	0.09
	<b>rs17870811</b>	6:44875421	99.5	0.06	0.46
<i>CSN1S1</i>	CSN1_AB	6:87146017	87.9	0.00	1.00
	rs433385179	6:87148464	99.9	0.00	1.00
	<b>rs43703010</b>	6:87157262	96.8	0.15	0.64
<i>CSN2</i>	rs43703013	6:87181453	96.8	0.04	0.02
	<b>rs43703011</b>	6:87181619	97.6	0.27	0.01
<i>CSN3</i>	rs43706475	6:87390479	99.3	0.00	1.00
	<b>rs43703015</b>	6:87390576	99.4	0.19	0.33
	rs43703017	6:87390632	98.2	0.04	0.33
<i>LGB</i>	rs41255680	11:103301242			
	rs41255685	11:103301690	75.3	0.08	< 1.10 <sup>-6</sup>
	rs43691049	11:103303343			
	rs109625649	11:103304757	83.5	0.10	0.56
<i>DGAT1</i>	rs109234250	14:1802265	62.8	0.10	0.61
	rs109326954	14:1802266	62.8	0.10	0.61
<i>STAT5A</i>	rs109487069	19:43047829			
<i>GH</i>	<b>rs109191047</b>	19:48768766	97.5	0.19	0.19
	<b>rs137651874</b>	19:48769040	98.1	0.48	0.10
<i>FASN</i>	rs208645216	19:51400139			
	rs41919985	19:51402032	78.7	0.15	2.10 <sup>-3</sup>
<i>GHR</i>	<b>rs385640152</b>	20:31909478	98.8	0.09	0.84
<i>PRLR</i>	<b>rs135164815</b>	20:39115344	99.2	0.15	0.91
<i>UTMP</i>	<b>rs132991801</b>	21:59667572	96.0	0.42	0.03
<i>LTF</i>	<b>rs41256920</b>	22:53521978	99.2	0.23	0.39
	<b>rs43706485</b>	22:53522038	95.5	0.19	0.93
<i>PRL</i>	<b>rs211032652</b>	23:35106206	99.5	0.08	0.57
	<b>rs110494133</b>	23:35110974	97.4	0.21	3.10 <sup>-6</sup>
<i>SCD1</i>	<b>rs41255693</b>	26:21144708	98.9	0.19	0.17

<sup>a</sup>OPN3907 is a T deletion, not a SNP

Call rate, MAF and HWE *p*-value for the SNPs that did not present failure in the genotyping reactions are indicated. Markers bolded in black passed the quality control checks. Location of the SNPs is based on assembly UMD3.1

**Table 2** SNP markers associated with milk production

SNP	BONF/BH adjusted <i>p</i> -values			Gene	Variation
	WP	HP	HxJ P		
<b>rs43375517</b>	0.06/0.06	0.01*/0.01	1/0.86	ARL4A	C/G, 5'UTR
<b>rs41255693</b>	0.33/0.09	0.43/0.09	1/0.87	SCD1	C/T, missense
<b>rs29004488</b>	0.40/0.09	1/0.50	0.27/0.14	LEP	T/C, missense
<b>rs137651874</b>	0.46/0.09	0.52/0.09	1/0.77	GH	G/A, intronic
<b>rs132812135</b>	0.47/0.09	1/0.20	1/0.66	OPN	A/C, 3'UTR
<b>rs135164815</b>	1/0.21	0.03*/0.02	0.98/0.20	PRLR	G/A, missense
rs43703011	1/0.21	1/0.26	0.94/0.20	CSN2	G/T, missense
<b>rs43706485</b>	1/0.22	0.07/0.02	1/0.30	LTF	C/G, 5'UTR
<b>rs109191047</b>	1/0.22	0.56/0.09	1/0.67	GH	A/C, synonymous
rs110930453	1/0.35	1/0.54	1/0.77	OPN	C/T, intronic
rs385640152	1/0.35	1/0.35	1/0.86	GHR	A/T, missense
rs110494133	1/0.35	1/0.60	1/0.53	PRL	C/T, intronic
rs132991801	1/0.36	1/0.60	1/0.77	UTMP	C/T, synonymous
rs211032652	1/0.36	1/0.26	1/0.86	PRL	G/A, synonymous
rs41595314	1/0.48	1/0.50	1/0.86	SNX13	A/T, intronic
rs42213673	1/0.53	1/0.60	1/0.86	ETV1	G/A, intronic
rs17870811	1/0.53	1/0.17	0.24/0.14	PPARGC1A	C/T, intronic
rs109579682	1/0.53	1/0.95	0.78/0.20	PPARGC1A	A/G, intronic
rs133669403	1/0.53	1/0.60	1/0.77	PPARGC1A	G/A, missense
rs43703015	1/0.53	1/0.34	1/0.77	CSN3	C/T, missense
rs41256920	1/0.86	1/0.60	1/0.67	LTF	C/A, 5'UTR
rs43703010	1/0.90	1/0.60	1/0.77	CSN1S1	A/G, missense

\**p* < 0.05

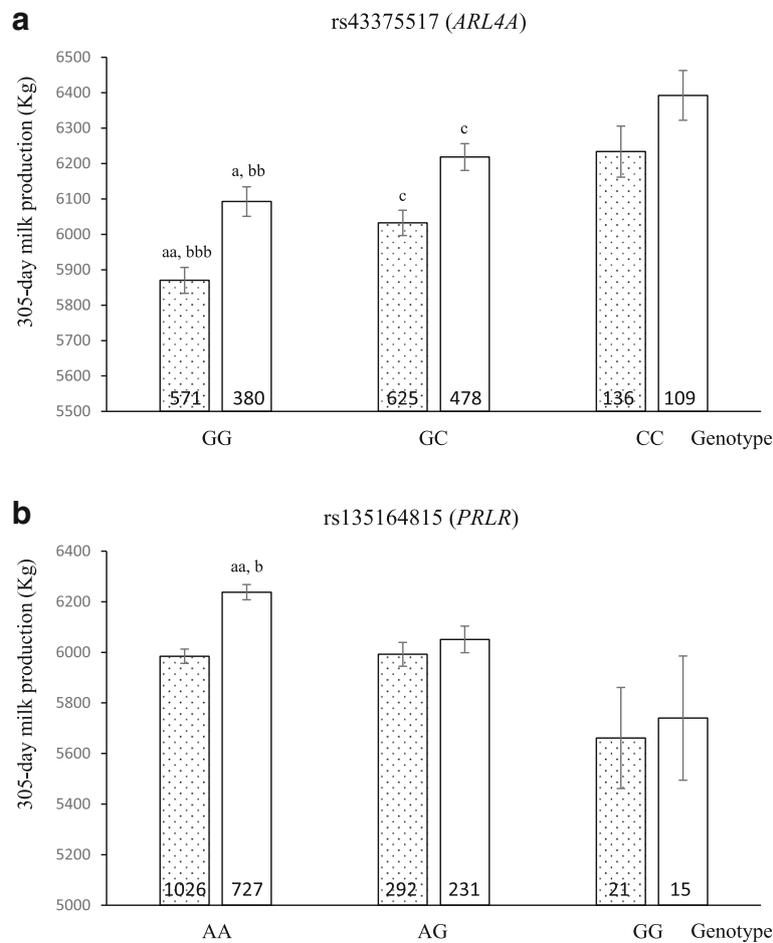
This Table indicates markers adjusted *p*-values based on Bonferroni correction (BONF) and Benjamini-Hochberg procedure (BH) for the whole (WP), Holstein (HP) and Holstein x Jersey (HxJ P) population analyses, corresponding gene names, and nucleotide changes implicated in the polymorphisms. SNPs in bold represent suggestive or significant associations with 305-day milk production

nucleotide variations occurring in *ARL4A*, *LEP*, *OPN*, *GH*, and *SCD1* genes.

Regarding the Holstein subpopulation, the SNPs rs43375517, located in *ARL4A* (BTA4), and rs135164815, located in *PRLR* (BTA20), were significantly associated with 305-day milk production after Bonferroni correction. Furthermore, other four SNPs showed suggestive associations. Two of them, rs137651874 and rs41255693, were also detected when analyzing the total population, while rs109191047 and rs43706485 were only detected when considering the Holstein population for analysis. SNP rs109191047 is a synonymous substitution in *GH* gene and rs43706485 is a nucleotide variation in the 5' untranslated region of *LTF* gene. This association test was performed using a linear model considering the fixed effects of breed, year of birth, sire, and the combined effect of season and year of first lactation. The fixed herd effect did not show to be significant in this analysis. When performing the association study on the Holstein x Jersey crosses subpopulation, the combined

effect of season and year of first lactation was not found to be significant, so only the effects of the covariates breed, year of birth, sire and herd were considered in the model used. In this test, none SNP showed a Benjamini-Hochberg adjusted *p*-value smaller than the false discovery rate used.

To illustrate the effect of each significantly associated SNP genotype on milk production, we compared 305-day cumulative milk production values of the cows with each of the three possible genotypes. As shown in Fig. 1, the cows with genotype GG in SNP rs43375517 (*ARL4A*) presented lower cumulative milk production than those with the other genotypes (*p* < 0.05 and *p* < 0.005 when comparing homozygous GG vs. heterozygous, and vs. homozygous CC, respectively). Regarding rs135164815 (*PRLR*), Holstein cows with genotype AA presented higher cumulative milk production than those with the other genotypes (*p* < 0.005 and *p* < 0.05 when comparing homozygous AA vs. heterozygous, and vs. homozygous GG, respectively). We further determined



**Fig. 1** Effect of different SNP genotypes on milk production. Genotypes for SNPs from *ARL4A* (a) and *PRLR* (b) genes are shown in the x-axis and 305-day cumulative milk production (Kg) in the y-axis. Dot bars and white bars denote the whole population and the Holstein subpopulation, respectively. The number of animals presenting each genotype is indicated inside each bar. Homozygote GG for rs43375517 or AA for rs135164815 vs heterozygote, <sup>aa</sup> $p < 0.005$ ; <sup>a</sup> $p < 0.05$ . Homozygote 1 vs homozygote 2, <sup>bbb</sup> $p < 0.0005$ ; <sup>bb</sup> $p < 0.005$ ; <sup>b</sup> $p < 0.05$ . Heterozygote vs homozygote CC for rs43375517 or GG for rs135164815, <sup>c</sup> $p < 0.05$

the population frequencies for the favorable alleles of these two SNPs found to be significantly associated with milk production. They were 0.34 for allele C of SNP rs43375517, and 0.88 for allele A of SNP rs135164815.

### Discussion

The economic relevance of lactation and milk production has given rise to numerous research studies on polymorphisms in candidate genes associated with these traits. In dairy cattle, much research has focused on the association of certain genetic variants of milk proteins with milk yield and composition traits [42]. Major milk protein genes have been identified in different genetic variants encoding similar proteins that are slightly different in chemical structure [38]. Based on this information, we selected 23 candidate genes identified in numerous independent studies that used the same or

different approaches. Furthermore, we selected 46 SNPs among these candidate genes for the SNPlex genotyping panel design.

The quality control test performed considerably reduced the number of candidate SNPs, mainly because of a low call rate and failure in the genotyping reactions. However, the proportion of discarded SNPs was similar to that of previously reported studies which used the same genotyping system [43–47], even though some of them [43–45] set a less stringent minimum call rate than the one used in this study. Herein, SNPs rs109234250 and rs109326954, located in the *DGAT* gene, showed the highest percentage of missing genotypes (37.2%) and therefore we did not take them into account for further analysis, although *DGAT* is a major QTL for milk production traits. Likewise, we could not evaluate the association of *LGB* variants with milk yield as we

removed the four SNPs within this candidate gene locus due to the genotyping errors in the reactions (rs41255680 and rs43691049) or low call rate (rs41255685 and rs109625649) that we found.

Based on the genotypes that passed the quality control criteria, the association test performed on the total population revealed that 22.7% of the SNPs evaluated (5 out of 22 SNPs) showed suggestive associations (adjusted Benjamini-Hochberg  $p$ -values  $< 0.1$ ) with milk production. This high percentage was not surprising because the markers analyzed were located at candidate genes or regions, or they had been previously reported to be associated with bovine milk production or composition traits. To provide additional information and to detect associations which may only occur in either the Holstein or the Holstein x Jersey subpopulations, we performed separate association analyses discriminating animals by breed. Regarding Holstein x Jersey animals, no significant SNPs were identified after multiple testing correction. However, when testing the Holstein subpopulation, two SNPs, rs43375517 and rs135164815, reached significance. The former is located in the 5'untranslated region (5'UTR) of *ARL4A* gene. This SNP was previously associated with predicted transmitting ability for milk in Holstein bulls [32]. The latter SNP constitutes a missense mutation in *PRLR* gene, which was previously reported by Viitala et al. as influencing milk protein and fat yield [33]. We found that the individuals with better performance were those with CC genotype at locus rs43375517 and AA genotype at locus rs135164815. Furthermore, when performing the association test on the Holstein population, other four SNPs showed suggestive associations to milk production according to the Benjamini-Hochberg false discovery rate test. These markers were rs43706485 in *LTF* gene, rs137651874 and rs109191047 both located in *GH* gene, and rs41255693 in *SCD1* gene.

Three of the SNPs identified (rs135164815, rs41255693, and rs29004488) correspond to non-synonymous mutations in *PRLR*, *SCD1*, and *LEP* genes. These missense variations could directly affect the protein biological function. Another suggestive SNP, rs137651874, is an intronic variation of *GH* gene. This polymorphism could trigger an abnormal splicing of mRNA, i.e. making a normal splice site disappear or generating an alternative one, or it could be transcribed to small regulatory RNAs, like micro RNAs, which can regulate gene expression. The SNP rs109191047, a synonymous mutation of *GH* gene, although it does not alter the primary sequence of the protein, it could generate a cryptic or alternative splice site of mRNA, could affect the translation rate based on tRNA availability for the new codon, or could simply be in linkage disequilibrium with another SNP that actually affects the phenotype. The remaining SNPs detected,

rs43375517, rs132812135, and rs43706485 are located in the 5'UTR of *ARL4A*, in the 3'UTR of *OPN* and in the 5'UTR of *LTF*, respectively, and could be affecting post-transcriptional regulation of gene expression.

There is abundant reported evidence that *OPN*, *LEP*, *ARL4A*, *GH*, *PRLR*, and *SCD1* genes are associated with milk production and/or composition traits [26, 30, 32, 33, 41, 48, 49]. Therefore, the results obtained in this study confirmed previously reported associations in other populations. Moreover, these results evidence the so-called "holsteinization" process, which has taken place in Latin America since the 1980s. In Argentina in particular, 65% of dairy cows are inseminated with Holstein semen imported from the United States and Canada. Consequently, it is not surprising that the associations found herein have been previously reported by other authors, based on the common genetic background being due to importations not only of semen but also of embryos and live animals from the northern hemisphere [50].

The approach followed in this study is a plausible way to explore SNPs affecting specific traits, such as milk production. Thereafter, these SNPs already proven to have a significant effect on the desired trait could be suitable as possible candidates for marker-assisted selection (MAS) programs. Moreover, they could even help to develop low density/low cost customized genotyping panels containing a few hundred or a few thousand markers, to be routinely used assisting breeding programs. In the context in which this work was started and developed, it was more cost-effective to genotype a few SNPs on candidate genes than to perform large-scale genotyping. However, with technological advancement in terms of development, availability and cost reduction of high-throughput microarray genotyping, Genomic Selection (GS) has become more affordable. In fact, in several countries, it is customary to use GS based on low or medium density genotyping in dairy cattle breeding. According to genome-wide association studies and genomic selection estimations, dairy traits are affected by 2000–10,000 genes, with the concomitant implication that if many genes affect a trait, individual genes have small effects, thus limiting the efficiency of the MAS approach [51]. Hence, whenever possible, Genomic Selection should be used since it captures a greater amount of the phenotypic variance. However, in a context where economic resources are scarce, MAS would be still applicable.

## Conclusions

This is the first study reporting associations between genetic markers and milk production in Argentinean Holstein and Holstein x Jersey cows. It not only provides useful information to explore the most relevant genes contributing

to the variation in bovine milk production, but also, and most important at country level, it might constitute a first step towards the design of selection programs aimed at increasing profitability of dairy operations by improving milk production performance of Argentinean Holstein and Holstein x Jersey dairy cattle.

## Additional file

**Additional file 1:** Primer sets for PCR used for amplification of 14 candidate genes on sire DNA. (DOC 109 kb)

**Additional file 2:** Genotypic and allelic frequencies obtained for the SNPs that passed the quality control. (DOC 62 kb)

## Abbreviations

BTA: *Bos taurus* autosome; GS: Genomic Selection; HWE: Hardy-Weinberg equilibrium; MAF: Minor allele frequency; MAS: Marker-assisted selection; QTL: Quantitative trait loci; SNP: Single nucleotide polymorphism

## Acknowledgements

We thank Irma Fuxan for careful technical assistance. We also would like to thank Las Taperitas S.A. for giving us access to phenotypic records databases and for enabling us to collect biological samples.

## Funding

This study was supported by Instituto Nacional de Tecnología Agropecuaria (INTA) grants PNBIO1131033 and PNBIO1131043, Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) grant PAE37143, and International Atomic Energy Agency (IAEA) CRPD3.10.28.

## Availability of data and materials

The data that support the findings of this study are available from Las Taperitas S.A. but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Las Taperitas S.A.

## Authors' contributions

MAP and AFA conceived and designed the study. JPN collected blood samples. MAR and JPN performed DNA extraction. AFA selected candidate genes and markers to be considered for the SNPlex genotyping panel design. AFA also sequenced certain candidate genes on the evaluated sire population. MAR performed genotyping quality control. MJB adjusted lactation curves and predicted the 305-day cumulative milk production. MAR carried out the statistical analysis and association test. MAP and MAR analyzed and interpreted the data. MAR wrote the manuscript. MAP, AFA, JPN and DOM made substantial contributions and constructive feedback to the manuscript. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Procedures employed for blood samples collection and for milk production evaluation were approved by the Institutional Committee for Care and Use of Experimental Animals (CICUA) of the National Institute of Agricultural Technology (INTA) under protocol number 35/2010 and followed the guidelines specified in the institutional manual.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Instituto Nacional de Tecnología Agropecuaria (INTA), Centro de Investigación en Ciencias Veterinarias y Agronómicas (CICVyA), Instituto de Genética "Ewald A. Favret", Nicolás Repetto y de los Reseros s/n, Hurlingham B1686, Argentina. <sup>2</sup>Instituto Nacional de Tecnología Agropecuaria (INTA), Estación Experimental Agropecuaria Rafaela, Ruta Nacional 34 Km 227, Rafaela, Argentina. <sup>3</sup>Instituto Nacional de Tecnología Agropecuaria (INTA), Estación Experimental Agropecuaria Anguil, Ruta Nacional 5 Km 580, Anguil, Argentina. <sup>4</sup>Instituto Nacional de Tecnología Agropecuaria (INTA), Estación Experimental Agropecuaria Pergamino, Ruta 32 Km 4.5, Pergamino, Argentina. <sup>5</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad Autónoma de Buenos Aires, Argentina.

Received: 26 July 2018 Accepted: 3 December 2018

Published online: 12 December 2018

## References

1. La lechería argentina: estado actual y su evolución (2008 a 2011). <https://inta.gov.ar/documentos/la-lecheria-argentina-estado-actual-y-su-evolucion-2008-a-2011>. Accessed 10 Oct 2018.
2. SENASA. Tambos. <https://www.argentina.gov.ar/senasa/mercados-y-estadisticas/estadisticas/animal-estadisticas/bovinos/bovinos-y-bubalinos-sector-primario>. Accessed 10 Oct 2018.
3. Producción de leche a nivel nacional. <https://datos.agroindustria.gov.ar/dataset?organization=subse-lecheria&groups=produccion-agroindustrial>. Accessed 10 Oct 2018.
4. Cappellini OR. Dairy development in Argentina. In: Dairy Reports. FAO. 2011.
5. Dekkers JCM, Hospital F. The use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet.* 2002;3:22–32.
6. Dekkers JCM. Application of genomics tools to animal breeding. *Current Genomics.* 2012;13:207–12.
7. Boichard D, Fritz S, Rossignol MN, Boscher MY, Malafosse A, Colleau JJ. Implementation of Marker Assisted Selection in French Dairy Cattle Breeding. In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, August 19-23, vol. 22. France, Session: Montpellier; 2002. p. 22.03.
8. Bennewitz J, Reinsch N, Thomsen H, Szyda J, Reinhardt F, Kühn C, Tuchscherer A, Schwerin M, Weimann C, Erhardt G, Kalm E. Marker Assisted Selection in German Holstein Dairy Cattle Breeding: Outline of the Program and Marker Assisted Breeding Value Estimation. In: 54th Annual Meeting of the European Association for Animal Production, 31 August to 3 2003, Rome, Italy, Session G1.9.
9. Spelman RJ. Utilisation of molecular information in dairy cattle breeding. In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, August 19-23, 2002, Montpellier, France, Session 22, 22.02.
10. Weller JL, Ezra E, Ron N. Invited review: a perspective on the future of genomic selection in dairy cattle. *J Dairy Sci.* 2017;100:1–12.
11. Grisart B, Coppieters W, Farnir F, Karin L, Ford C, et al. 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.* 2002;12:222–31.
12. Weikard R, Kühn C, Goldammer T, Freyer G, Schwerin M. The bovine PPARGC1A gene: molecular characterization and association of an SNP with variation of milk fat synthesis. *Physiol Genomics.* 2005;21:1–13.
13. Roy R, Ordovas L, Zaragoza P, Romero A, Moreno C, Altarriba J, Rodellar C. Association of polymorphisms in the bovine FASN gene with milk-fat content. *Anim Genet.* 2006;37:215–8.
14. Ogorevc J, Kunej T, Razzpet A, Dovc P. Database of cattle candidate genes and genetic markers for milk production and mastitis. *Anim Genet.* 2009;40: 832–51.
15. You FM, Huo N, Gu YQ, Luo M-C, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics.* 2008;9:253.
16. Staden R, Beal KF, Bonfield JK. The Staden package. In: S Misener, SA Krawetz, editors. Computer methods in molecular biology, bioinformatics methods and protocols. The Humana press Inc., Totowa, New Jersey; 1998. Vol 132; p. 115–130.
17. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, et al. A whole-genome assembly of the domestic cow. *Bos taurus* *Genome Biol.* 2009. <https://doi.org/10.1186/gb-2009-10-4-r42>.

18. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005;6:31.
19. Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, et al. The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J Biomol Tech*. 2005;16:396–404.
20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
21. Kirkpatrick M, Lofsvold D, Bulmer M. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*. 1990;124:979–93.
22. Jamrozik J, Schaeffer LR, Dekkers JCM. Genetic evaluation of dairy cattle using test day yields and random regression model. *J Dairy Sci*. 1997;80:1217–26.
23. Cohen-Zinder M, Seroussi E, Larkin DM, Looor JJ, et al. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res*. 2005;15:936–44.
24. Ikonen T, Bovenhuis H, Ojala M, Ruottinen O, Georges M. Associations between casein haplotypes and first lactation milk production traits in Finnish Ayrshire cows. *J Dairy Sci*. 2001;84:507–14.
25. Nilsen H, Olsen H, Hayes B, Sehested E, Svendsen M, Nome T, Meuwissen T, Lien S. Casein haplotypes and their association with milk production traits in Norwegian red cattle. *Genet Sel Evol*. 2009. <https://doi.org/10.1186/1297-9686-41-24>.
26. Khatib H, Zaitoun I, Wiebelhaus-Finger J, Chang YM, Rosa GJM. The association of bovine PPARGC1A and OPN genes with milk composition in two independent Holstein cattle populations. *J Dairy Sci*. 2007;90:2966–70.
27. Leonard S, Khatib H, Schutzkus V, Chang YM, Maltecca C. Effects of the osteopontin gene variants on milk production traits in dairy cattle. *J Dairy Sci*. 2005;88:4083–6.
28. Schnabel RD, Kim J-J, Ashwell MS, Sonstegard TS, Van Tassell CP, Connor EE, Taylor JF. Fine-mapping milk production quantitative trait loci on BTA6: analysis of the bovine osteopontin gene. *Proc Natl Acad Sci U S A*. 2005; 102:6896–901.
29. Blott S, Kim J-J, Moiso S, Schmidt-Küntzel A, Cornet A, et al. Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics*. 2003;163:253–66.
30. Buchanan FC, Van Kessel AG, Waldner C, Christensen DA, Laarveld B, Schmutz SM. Hot topic: an association between a leptin single nucleotide polymorphism and Milk and protein yield. *J Dairy Sci*. 2003;86:3164–6.
31. Komisarek J, Dorynek Z. Effect of ABCG2, PPARGC1A, OLR1 and SCD1 gene polymorphism on estimated breeding values for functional and production traits in polish Holstein-Friesian bulls. *J Appl Genet*. 2009;50:125–32.
32. Rincón G, Islas-Trejo A, Casellas J, Ronin Y, Soller M, Lipkin E, Medrano JF. Fine mapping and association analysis of a quantitative trait locus for milk production traits on *Bos taurus* autosome 4. *J Dairy Sci*. 2009;92:758–64.
33. Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Mäki-Tanila A, Georges M, Vilkki J. The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. *Genetics*. 2006;173:2151–64.
34. Kuss AW, Gogol J, Geldermann H. Associations of a polymorphic AP-2 binding site in the 5'-flanking region of the bovine  $\beta$ -Lactoglobulin gene with milk proteins. *J Dairy Sci*. 2003;86:2213–8.
35. Brym P, Kamiński S, Wójcik E. Nucleotide sequence polymorphism within exon 4 of the bovine prolactin gene and its associations with milk performance traits. *J Appl Genet*. 2005;46:179–85.
36. Khatib H, Schutzkus V, Chang YM, Rosa GJM. Pattern of expression of the uterine Milk protein gene and its association with productive life in dairy cattle. *J Dairy Sci*. 2007;90:2427–33.
37. Liao Y, Du X, Lonnerdal B. miR-214 regulates lactoferrin expression and proapoptotic function in mammary epithelial cells. *J Nutr*. 2010;140:1552–6.
38. Farrell HM, Jimenez-Flores R, Bleck GT, Brown EM, Butler JE, et al. Nomenclature of the proteins of cows' Milk - sixth revision. *J Dairy Sci*. 2004;87:1641–74.
39. Schennink A, Bovenhuis H, Léon-Kloosterziel KM, van Arendonk AM, Visker MHPW. Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. *Anim Genet*. 2009;40:909–16.
40. Olsen HG, Nilsen H, Hayes B, Berg PR, Svendsen M, Lien S, Meuwissen T. Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. *BMC Genet*. 2007;8:32.
41. Mullen MP, Berry DP, Howard DJ, Diskin MG, Lynch CO, Berkowicz EW, Magee DA, MacHugh DE, Waters SM. Associations between novel single nucleotide polymorphisms in the *Bos taurus* growth hormone gene and performance traits in Holstein-Friesian dairy cattle. *J Dairy Sci*. 2010;93:5959–69.
42. Huang W, Peñagaricano F, Ahmad KR, Lucey JA, Weigel KA, Khatib H. Association between milk protein gene variants and protein composition traits in dairy cattle. *J Dairy Sci*. 2012;95:440–9.
43. Chang SC, Chang PY, Butler B, Goldstein BY, Mu L, et al. Single nucleotide polymorphisms of one-carbon metabolism and cancers of the esophagus, stomach, and liver in a Chinese population. *PLoS One*. 2014. <https://doi.org/10.1371/journal.pone.0109235>.
44. Davidson CM, Northrup H, King TM, Fletcher JM, Townsend I, Tyerman GH, Au KS. Genes in glucose metabolism and association with spina bifida. *Reprod Sci*. 2008;15:51–8.
45. Tran PX, Au KS, Morrison AC, Fletcher JM, et al. Association of retinoic acid receptor genes with meningomyelocele. Birth defects research. Part a. *Clinical and molecular teratology*. 2011;91:39–43.
46. Jimenez-Sousa MA, López E, Fernandez-Rodríguez A, Tamayo E, et al. Genetic polymorphisms located in genes related to immune and inflammatory processes are associated with end-stage renal disease: a preliminary study. *BMC Med Genet*. 2012. <https://doi.org/10.1186/1471-2350-13-58>.
47. O'Byrne MR, Au KS, Morrison AC, Lin JJ, Fletcher JM, et al. Association of folate receptor (FOLR1, FOLR2, FOLR3) and reduced folate carrier (SLC19A1) genes with meningomyelocele. Birth defects research. Part A, Clinical and molecular teratology. 2010;88:689–94.
48. Kgwatalala PM, Ibeagha-Awemu EM, Hayes JF, Zhao X. Stearoyl-CoA desaturase 1 3'UTR SNPs and their influence on milk fatty acid composition of Canadian Holstein cows. *J Anim Breed Genet*. 2009;126:394–403.
49. Fontanesi L, Calò DG, Galimberti G, Negrini R, Marino R, et al. A candidate gene association study for nine economically important traits in Italian Holstein cattle. *Anim Genet*. 2014;45:576–80.
50. Casas E, Montaldo H, Nonneman D, Poli M. Potential contribution of genomics and biotechnology in animal production. In: Núñez R, Ramírez R, Fernández S, et al, editors. *La ganadería en América Latina y el Caribe: alternativas para la producción competitiva, sustentable e incluyente de alimentos de origen animal*. 1st ed. Fundación Colegio de Postgraduados, México: Biblioteca Básica de Agricultura; 2015. p. 761–88.
51. Meuwissen T, Hayes B, Goddard M. Genomic selection: a paradigm shift in animal breeding. *Animal Frontiers*. 2016;6:6.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

