



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

Brian K. Meredith^{1,2}, Donagh P. Berry³, Francis Kearney⁴, Emma K. Finlay⁵, Alan G. Fahey², Daniel G. Bradley^{5*} and David J. Lynn^{1*}

¹ Animal and Bioscience Research Department, Teagasc, Animal and Grassland Research and Innovation Centre, Grange, Dunsany, Co. Meath, Ireland

² School of Agriculture and Food Science, University College Dublin, Dublin 4, Ireland

³ Animal and Bioscience Research Department, Teagasc, Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland

⁴ Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland

⁵ Molecular Population Genetics, Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland

Edited by:

Tad S. Sonstegard, US Department of Agriculture, Agricultural Research Service, USA

Reviewed by:

Marcos V. Silva, Embrapa, Brazil
Dan Nonneman, US Department of Agriculture, Agricultural Research Service, USA

*Correspondence:

Daniel G. Bradley, Molecular Population Genetics, Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland
e-mail: dbradley@tcd.ie;

David J. Lynn, Animal and Bioscience Research Department, Teagasc, Animal and Grassland Research and Innovation Centre, Grange, Dunsany, Co. Meath, Ireland
e-mail: david.lynn@teagasc.ie

Mastitis is an inflammation-driven disease of the bovine mammary gland that occurs in response to physical damage or infection and is one of the most costly production-related diseases in the dairy industry worldwide. We performed a genome-wide association study (GWAS) to identify genetic loci associated with somatic cell score (SCS), an indicator trait of mammary gland inflammation. A total of 702 Holstein-Friesian bulls were genotyped for 777,962 single nucleotide polymorphisms (SNPs) and associated with SCS phenotypes. The SCS phenotypes were expressed as daughter yield deviations (DYD) based on a large number of progeny performance records. A total of 138 SNPs on 15 different chromosomes reached genome-wide significance (corrected p -value ≤ 0.05) for association with SCS (after correction for multiple testing). We defined 28 distinct QTL regions and a number of candidate genes located in these QTL regions were identified. The most significant association (p -value = 1.70×10^{-7}) was observed on chromosome 6. This QTL had no known genes annotated within it, however, the Ensembl Genome Browser predicted the presence of a small non-coding RNA (a Y RNA gene) in this genomic region. This Y RNA gene was 99% identical to human RNY4. Y RNAs are a rare type of non-coding RNA that were originally discovered due to their association with the autoimmune disease, systemic lupus erythematosus. Examining small-RNA sequencing (RNAseq) data being generated by us in multiple different mastitis-pathogen challenged cell-types has revealed that this Y RNA is expressed (but not differentially expressed) in these cells. Other QTL regions identified in this study also encoded strong candidate genes for mastitis susceptibility. A QTL region on chromosome 13, for example, was found to contain a cluster of β -defensin genes, a gene family with known roles in innate immunity. Due to the increased SNP density, this study also refined the boundaries for several known QTL for SCS and mastitis.

Keywords: GWAS, mastitis, SNP, somatic cell score, Holstein-Friesian

INTRODUCTION

Mastitis is an inflammatory disease of the bovine mammary gland that occurs in response to physical damage or infection with pathogenic microorganisms such as *Escherichia coli*, *Streptococcus uberis*, and *Staphylococcus aureus* (Schukken et al., 2011). Upon infection of the mammary gland, macrophages and epithelial cells release cytokines which cause polymorphonuclear neutrophils (PMNs), monocytes, and other leukocytes to migrate from the blood to the site of infection in the mammary tissue (Paape et al., 2003). This influx of leukocytes results in an increased level of what is commonly termed “somatic cells” in the milk. Clinical mastitis (CM) is common in dairy herds and is estimated to occur

in 20–40% of cows in a dairy herd annually (Heringstad et al., 2000). The costs associated with mastitis are high with mastitis-related costs in the EU estimated at €1.55 billion annually (<http://www.sabre-eu.eu/>). The frequency and cost of mastitis, along with increased public interest in animal welfare, and concerns regarding the use of antibiotics in livestock, ensure that mastitis is one of the most important production diseases in the dairy industry and therefore any method to reduce the incidence of mastitis would be welcomed. CM has a heritability of 0.03 which means the response to selection will be proportionally small; however, selection for resistance to mastitis is still possible (Heringstad et al., 2000; Carlen et al., 2004). Obtaining direct

measures of CM for a sufficiently large number of progeny can, however, be difficult and in the recent past, CM incidence has only been recorded routinely in some Nordic countries (Heringstad et al., 2000). In Ireland and in many other countries, national dairy industries routinely record somatic cell count (SCC) as an indicator trait. As mentioned previously, the host response to intra-mammary infection results in an influx of somatic cells into the milk. The SCC phenotype measures the concentration (cells/ml) of these cells in a milk sample and therefore acts as an indicator of intra-mammary infection. The log transformation of SCC, called somatic cell score (SCS), exhibits a strong genetic correlation (~ 0.70) with CM (Rupp and Boichard, 2003; Bloemhof et al., 2009) and with a heritability of ~ 0.15 (Mrode and Swanson, 1996) it is a useful phenotype which has been used in a number of association studies investigating the genetic basis of mastitis resistance (Ashwell et al., 1998; Heyen et al., 1999).

The advent of high-throughput genotyping arrays has enabled researchers to perform association studies on a genome-wide scale. A number of genome-wide association studies (GWAS) have detected QTL for SCS in cattle and these are concentrated on chromosomes 5, 8, 11, 18, and 23 (Khatkar et al., 2004; Smaragdov, 2006). These associations help us to better understand the mechanisms underlying mastitis resistance, to identify likely candidate genes and illuminate the underlying complex nature of the SCS phenotype. An additional role for these QTL regions may be to provide prior information to statistical models used in the selection of animals for breeding. A recent study tested this hypothesis and found that inclusion of QTL as prior information could lead to increased accuracy of selection in some traits (Brøndum et al., 2012).

When designing a GWAS there are two important factors, amongst others (e.g., effect size), which may affect the statistical power of that study to detect associations: (1) the number of markers genotyped and (2) the number of samples which have been genotyped. The GWAS approach relies on the concept that markers are in linkage disequilibrium (LD) with the causal variant and greater LD between the two results in greater power to detect the association (Spencer et al., 2009). In this regard it has been observed that genotyping larger number of SNPs through the use of higher density genotyping arrays can increase the power of the resultant GWAS (Spencer et al., 2009). A variety of genotyping arrays are available in cattle which differ in the number of SNPs which they can genotype ranging from low density ($\sim 7,000$ SNPs) to medium density ($\sim 54,000$ SNPs) and upwards to high density ($\sim 780,000$ SNPs). In general, recent GWAS in Holstein-Friesian dairy cattle have utilized genotyping arrays which assay $\sim 54,000$ SNPs (Cole et al., 2009; Pryce et al., 2010; Meredith et al., 2012). This is reasonable because it was estimated that only $\sim 50,000$ SNPs would be required to achieve sufficient LD ($r^2 \geq 0.2$) for a GWAS in Holstein-Friesian cattle due to the extensive LD in the breed (de Roos et al., 2008). However, as discussed by Khatkar et al. (2008), increased SNP density should lead to higher-powered GWAS due to the stronger average LD between SNPs and coverage achieved across the genome. Therefore, the use of a higher-density genotyping array in the form of the Illumina High-Density Bovine BeadChip (Illumina Inc., San Diego, CA, USA) paves the way for such a GWAS to be performed.

The objective of this study was to perform a GWAS using a group of 702 Holstein-Friesian bulls which have been genotyped using the Illumina High-density Bovine BeadChip (777,962 SNPs) to detect associations with high-reliability SCS phenotypes, estimated using a large number of progeny performance records. We identified a number of QTL regions and genes within and around these regions which may play a role in the SCS trait, and by extension resistance to mastitis.

MATERIALS AND METHODS

DNA EXTRACTION

Thawed frozen semen was washed twice in phosphate-buffered saline (pH 7.4), and cell pellets were harvested via centrifugation and re-suspended in 450 μ L of pre-warmed extraction buffer (10 mM Tris, pH 8.0; 10 mM EDTA, pH 8.0; 1% SDS; 100 mM NaCl); 15 μ L of 2-mercaptoethanol was added. Samples were incubated at 55°C for 15 min followed by the addition of 10 μ L of proteinase K (20 mg/mL). Lysis occurred following overnight incubation at 60°C. DNA was then extracted using the Maxwell instrument (Promega Corp., Madison, WI) according to the manufacturer's instructions. Genotyping using the Illumina High-density Bovine BeadChip (Illumina Inc., San Diego, CA) and subsequent genotype calling were carried out by Geneseek (Lincoln, NE).

ILLUMINA HIGH-DENSITY BOVINE GENOTYPES

Seven hundred and forty-seven Holstein-Friesian sires were genotyped (777,962 SNPs) on the Illumina High-density Bovine BeadChip. SNP positions were based on the UMD 3.1 bovine genome assembly (Zimin et al., 2009). The genotype data underwent a series of quality control checks and a number of animals and SNPs were removed in the following order; (1) 8 animals and 14,730 SNPs were removed which exhibited over 5% Mendelian inconsistencies (2) 15 animals and 10,412 SNPs had greater than 10% missing SNP calls and were removed, (3) 40 SNPs were removed due to the proportion of heterozygous genotypes being $>90\%$, (4) 6,409 SNPs violating Hardy-Weinberg equilibrium (p -value <0.000001) were removed and (5) 3,153 SNPs located in mitochondrial DNA, on the Y chromosome or which had an unknown genomic position were removed. A further 165,037 SNPs were removed before association analysis due to a minor allele frequency <0.05 ; of these, 97,325 SNPs were monomorphic. An additional 22 animals were removed from the association analysis that did not have a daughter yield deviation (DYD) phenotype for SCS. Following all of these edits, a group of 702 Holstein-Friesian sires genotyped at 578,181 SNPs was available for association testing.

DAUGHTER YIELD DEVIATIONS FOR SOMATIC CELL SCORE

The phenotypes for SCS were expressed as DYD and were based on genetic evaluations carried out in January 2012 by the Irish Cattle Breeding Federation (ICBF). The DYD phenotypes are genetic evaluations undertaken using the performance records from an animal's own progeny. The bulls in this study were widely-used for artificial insemination in Ireland and hence have a large number of progeny performance records. Therefore, animal phenotypes are not just based on a single measure of a trait

but instead on a composite of all their progeny performance data. These DYD phenotypes for average SCS are estimated in Ireland using a repeatability animal model across the first five lactations according to the protocol of Berry et al. (2007). The reliability of a phenotype is a measure of the amount of information used in the genetic evaluation of that animal. The animals in this study were bulls widely-used for artificial insemination in Ireland and so they have a large number of progeny and therefore high phenotypic reliabilities. Reliabilities range from 0-1 with zero being the lowest reliability possible. Animals with high reliability (0.99) have a large number of progeny (~1200 + progeny) and the reliability drops as the number of progeny (and hence performance records) decreases (e.g., a reliability of 0.60 equates to an animal with ~30 progeny). Sires in this study had a mean (S.D.) reliability of 0.88 (0.13) which is roughly equivalent to an animal with 150 progeny. The SCS phenotype was checked for normality using the R statistical software (R Core Team, 2012). A pedigree file, containing at least four generations of the genotyped animal's ancestors, was created for the 702 animals resulting in a pedigree file with 4,739 animals.

STATISTICAL ANALYSIS

Associations (702 animals genotyped at 578,181 SNPs) were quantified using a single SNP regression approach. The dependent variable (i.e., the respective trait expressed as a DYD) was regressed on each SNP individually in a mixed animal model fitted in DMU (Madsen and Jensen, 2010). Each SNP was fitted as a fixed, continuous effect. The individual animal was fitted as a random effect with relationships among animals accounted for using the additive relationship matrix based on pedigree information. Correction for multiple testing was performed using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) as implemented by the "p.adjust" function in version 2.14.1 of the R statistical package. A genome-wide significance threshold was set as a corrected p -value ≤ 0.05 which equated to an uncorrected p -value of $\sim 1.19 \times 10^{-5}$.

DEFINING QTL REGIONS BASED ON LOCAL LD STRUCTURE

A quantitative trait locus (QTL) region surrounding a detected significant association was defined based on the local LD structure for each SNP, individually. Pairwise LD, as measured by r^2 , between all SNPs within 5 Mb upstream and downstream of a significant SNP was calculated using PLINK (Purcell et al., 2007). This allowed us to graphically examine the LD structure surrounding each significant SNP. Within this 10 Mb region (5 Mb upstream and downstream) a cut-off of $r^2 \geq 0.5$ was used and the SNP with this level of LD which were furthest upstream or downstream of the significant SNP were noted, thereby forming a QTL region. Any adjacent QTL regions which overlapped were combined into a single QTL region using the maximum and minimum genomic positions as boundaries; this was carried out in R/Bioconductor (Gentleman et al., 2004) using the reduce function in the "GenomicRanges" package (Aboyoun et al., 2010). In this way, the significant SNPs were used to identify a list of QTL regions. These regions were subsequently investigated to ascertain which, if any, genes were located in each region.

RESULTS

SAMPLE AND SNP SUMMARY STATISTICS

The SCS phenotypes were normally distributed with a mean (S.D.) of 36.57 (104.10) with minimum and maximum values of -236 and 560, respectively. The 702 animals had a mean (S.D.) SCS phenotype reliability of 0.88 (0.13). The mean distance between SNPs across all the chromosomes was 6,174 bp but was lower when only considering the autosomes (4,342 bp).

GENOME-WIDE ASSOCIATIONS

Genome-wide associations were estimated using the 702 animals at 578,181 SNPs and estimates of SNP effect, standard error, t statistic, uncorrected and corrected p -values for all SNPs can be viewed in Supplementary Table 1. Twenty eight distinct QTL regions (Table 1) were significantly associated with SCS, with 138 SNPs exceeding the threshold for genome-wide significance after correction for multiple testing. The significantly associated SNPs were located across 15 different chromosomes (chromosomes 1, 3, 4, 5, 6, 9, 10, 13, 17, 20, 21, 22, 23, 24, 25, and 26) (Figure 1). These QTL regions were cross-referenced to known QTLs for SCS / SCC and CM in the cattle QTL database (<http://www.animalgenome.org/QTLdb/cattle.html>) as of May 2013 and 11 QTLs fell within previously identified QTL regions; this left 17 remaining novel QTL regions. In general, the QTL regions identified using the high-density chip were more strongly associated with SCS and the boundaries of the QTL regions were more precise than those identified in our previous SCS GWAS in a similar population of animals using the lower-density Illumina BovineSNP50 BeadChip (Meredith et al., 2012).

The strongest association signal in this study was detected on chromosome 6 (located at 76.6–77.9 Mb) with 82 SNPs exceeding the genome-wide significance threshold. The most significant SNP had a p -value of 1.70×10^{-7} (Figure 2). No known genes were annotated in this region, however, there was a predicted Y RNA gene (ENSBTAG00000043492), which was located 188,460 bp downstream of the most significant SNP (BovineHD0600021283). Y RNAs are a rare type of small non-coding RNA which were first detected in the investigation of the autoimmune disease systemic lupus erythematosus (Lerner et al., 1981). Y RNAs are evolutionarily conserved (Mosig et al., 2007) and this bovine Y RNA was found to be 99% identical to human Y RNA 4 (RNY4). Little is currently known regarding the function of Y RNAs, except that they are components of Ro ribonucleoproteins (RNPs) (Verhagen and Pruijn, 2011) and may have a role in RNA quality control and in the initiation of DNA replication (Krude et al., 2009).

Examination of small RNAseq data recently published by our group (Lawless et al., 2013) showed that this bovine predicted Y RNA was in fact a real expressed gene and was expressed [albeit at a relatively low level—average 12 reads per million (RPM)] in bovine mammary epithelial (BME) cells infected with a mastitis causing pathogen, *Streptococcus uberis*. This Y RNA did not, however, appear to be differentially expressed in BMEs in comparison to uninfected controls. There is no evidence to link differential expression of this Y RNA gene to mastitis and so we hypothesize that SNPs may occur within the Y RNA which could lead to structural differences in the Y RNA, affecting its function.

Table 1 | QTL regions identified for SCS.

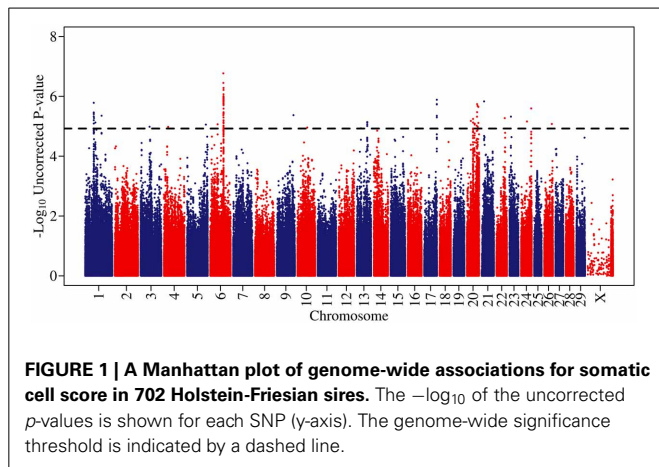
Chr	QTL start (bp)	QTL end (bp)	QTL length (bp)	No. of SNPs	No. of Genes	Gene names*
1	49,885,358	50,452,539	567,182	10	1	ALCAM
1	51,679,722	51,695,056	15,335	1	0	
1	59,651,839	59,799,918	148,080	1	1	ZBTB20
1	96,026,399	96,600,720	574,322	1	3	FNDC3B TMEM212 PRKCSH
3	52,269,691	52,300,879	31,189	1	0	
4	24,799,926	24,800,412	487	1	1	ISPD
5	110,275,795	110,275,795	1	1	1	SOX10
6	43,203,039	43,259,077	56,039	1	0	
6	76,649,555	77,943,085	1,293,531	82	1	Y_RNA
9	97,459,242	97,459,242	1	1	1	WTAP
10	57,715,416	57,715,416	1	1	1	BT.37181
13	24,548,898	24,720,132	171,235	1	2	C10ORF67 OTUD1
13	61,182,238	61,619,701	437,464	7	13	TRIB3NRSN2 ENSBTAG0000 0027386 C20ORF96 DEFB129 DEFB119 DEFB122A DEFB122 BT.54136 BT.100199 REM1 ZCCHC3 DEFB117
17	73,874,968	73,891,077	16,110	6	1	ENSBTAG0000 0023007 EGFLAM
20	36,064,784	36,222,775	157,992	4	1	
20	45,581,169	45,638,063	56,895	4	0	
20	57,650,372	57,746,837	96,466	1	0	
20	59,314,402	59,314,402	1	1	1	DNAH5
20	59,848,168	59,899,899	51,732	2	0	
20	62,174,229	63,901,891	1,727,663	2	13	CTNND2 DAP ANKRD33B BT.22715 MARCH6 ENSBTAG0000 0022809 CMBL BT.68634 BT.65753 BT.76062 U2 SNORD123 U6
20	67,038,772	67,654,036	615,265	2	0	
20	23,201,811	23,616,460	414,650	1	5	BT.66498 IL31RA BT.35994 SLC38A9 SNORA18
21	13,978,417	14,022,698	44,282	1	0	
22	48,483,717	48,603,179	119,463	1	2	BT.103068 TMEM110
23	8,449,971	8,449,971	1	1	0	

(Continued)

Table 1 | Continued

Chr	QTL start (bp)	QTL end (bp)	QTL length (bp)	No. of SNPs	No. of Genes	Gene names*
24	34,723,910	34,723,910	1	1	0	
24	59,670,860	59,759,141	88,282	1	1	MC4R
26	45,116,066	45,149,850	33785	1	0	

The Ensembl gene identifier is provided where a gene name was not available.

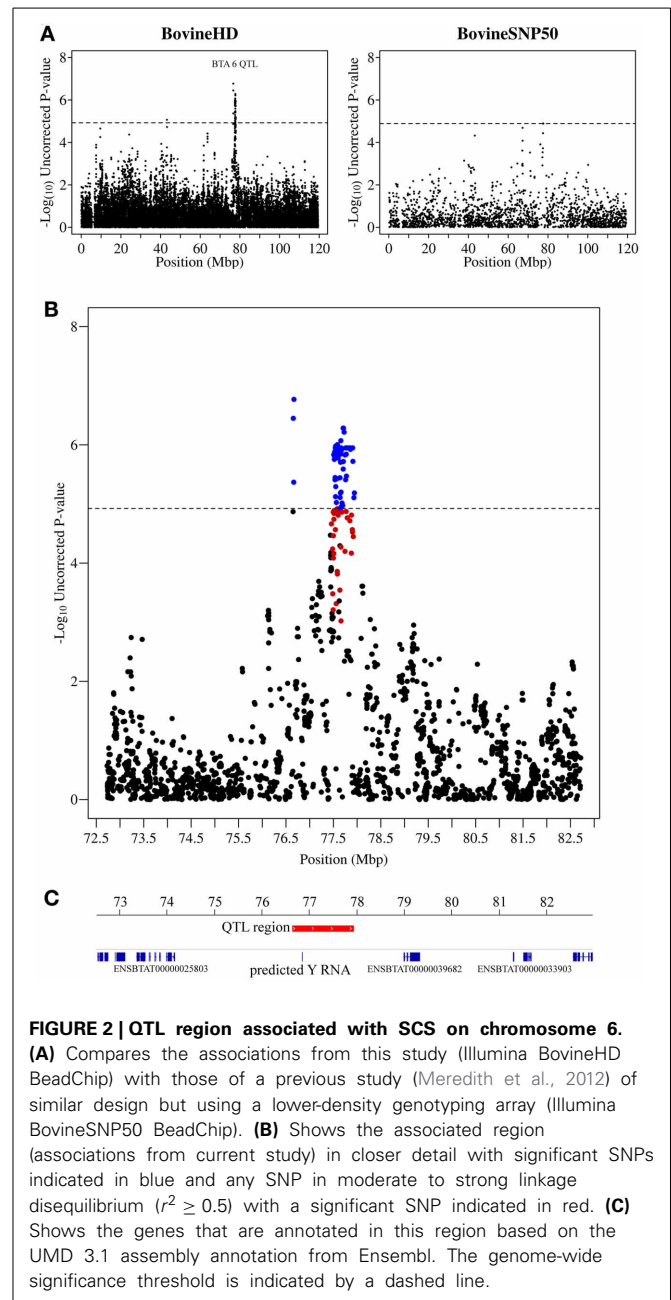


Unfortunately, no SNP genotyped by the Illumina bovine high-density BeadChip was located within this gene.

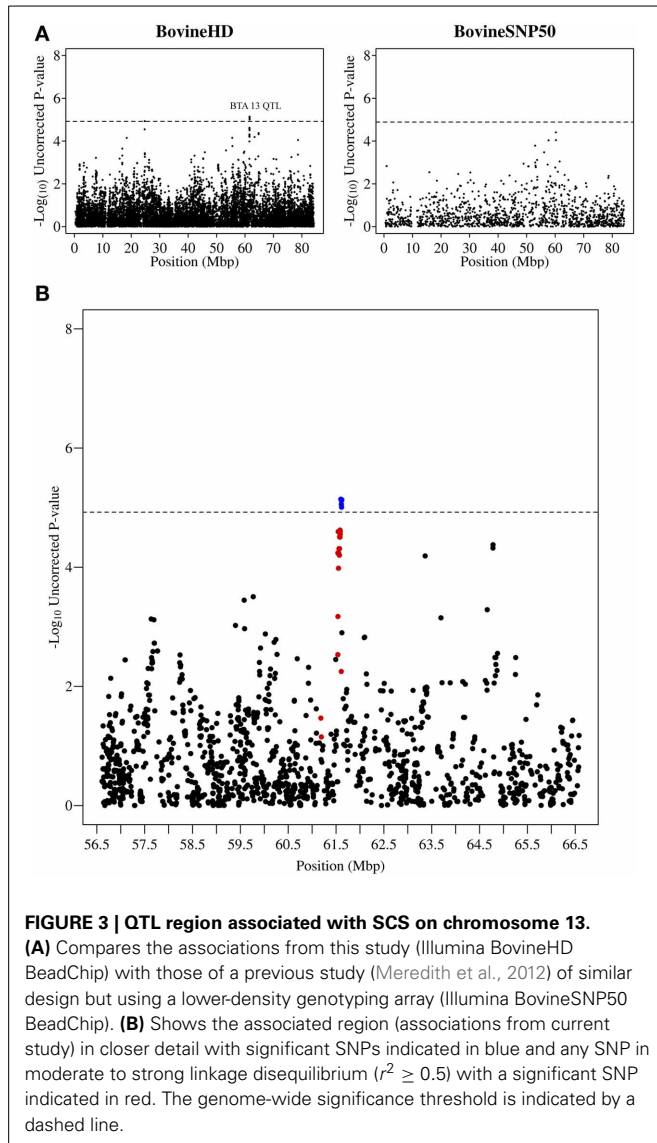
Other QTL regions identified in this study also contained strong candidate genes for mastitis susceptibility. A significant association was detected on chromosome 13 at 61.2–61.6 Mb with the most significant SNP having a p -value of 7.23×10^{-6} (Figure 3). The association signal was observed to be a more precise and significant signal than had been identified in our previous 50K GWAS study (Meredith et al., 2012). This QTL region, which extends for 437 kb, overlaps a known cluster of β -defensin genes (Cormican et al., 2008) which encompasses five β -defensin genes (*DEFB129*, *DEFB119*, *DEFB122A*, *DEFB122*, *DEFB117*) and two genes for β -defensin precursors (*BT.54136*, *BT.100199*). β -defensins are a family of host defence peptides involved in the innate immune response and which are found across numerous species including cattle, humans and chickens (Ganz, 2003).

A peak of six significantly associated SNPs was also detected near the end of chromosome 17 at 73.87–73.89 Mb with the most significant of these SNPs having an uncorrected p -value of 1.3×10^{-6} (Figure 4). A strong candidate gene, Mitogen-activated protein kinase 1 (*MAPK1*), which has been shown to be significantly down-regulated in bovine primary BME cells in response to challenge with *Escherichia coli* (Gunther et al., 2011), was found to be located 120 kb downstream of the detected QTL region.

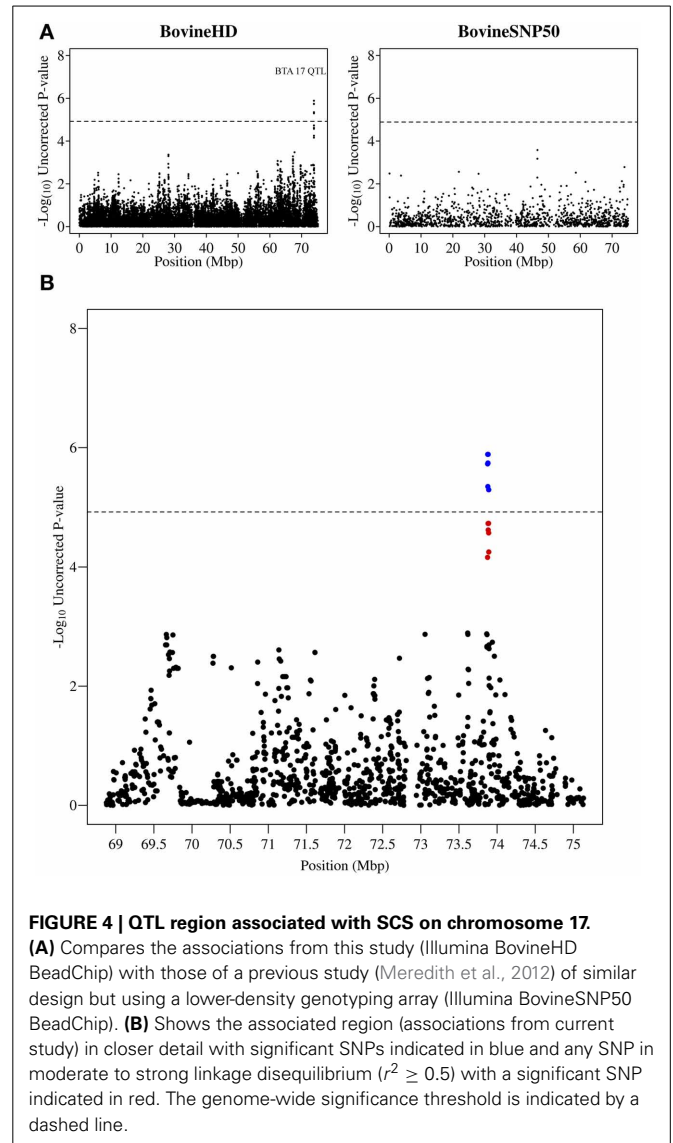
A significant association signal was also detected on chromosome 1 with a QTL region extending from 49.9–50.5 Mb; the most significant SNP (BovineHD0100014076) in this region had an uncorrected p -value of 1.64×10^{-6} . Only one gene, activated leukocyte cell adhesion molecule (*ALCAM*), overlapped with this QTL region.



On chromosome 20, a number of significantly associated QTL regions were detected. One of these extended from 23.2–23.6 Mb (the most significant SNP in this region was BovineHD2000007011—uncorrected p -value = 6.64×10^{-6}).



One gene of interest in this region, which contained a significantly-associated SNP, was interleukin-31 receptor subunit alpha precursor (*IL31R α*) gene. Another QTL region of interest on chromosome 20 was identified from 62.2–63.9 Mb with the most associated SNP, BovineHD2000017761, having an uncorrected p -value of 2.08×10^{-6} . This region was gene-rich with a total of 13 bovine genes located within its boundaries. In addition, another QTL region on chromosome 20 located 36.0–36.2 Mb was detected which had a number of nearby adjacent complement genes (*C9*, *C6*, *BT.69954*, and *CIQTNF3*). These complement genes are part of the complement system, a collection of serum proteins and cell surface receptors, which form an integral part of the innate immune system that are activated via host pattern-recognition receptors which recognize pathogen-associated molecular patterns (PAMPs) on pathogenic microorganisms (Medzhitov and Janeway, 2002; Medzhitov, 2007). The complement system is involved in innate immune responses such as inflammation, opsonization, chemotaxis of phagocytes to the



site of infection, and cell lysis, and also is believed to play a role in adaptive immunity via the regulation of B and T cell response (Carroll, 2004; Medzhitov, 2007). Therefore, these complement genes are good candidate genes to be involved in resistance to mastitis.

Finally, a QTL region detected on chromosome 26 extending from 45.12–45.15 Mb fell within a larger QTL region previously associated with SCS (Zhang et al., 1998) and CM (Lund et al., 2008). However, our detected QTL is much narrower than those previously identified.

IMPUTATION

Imputation of lower-density (BovineSNP50) genotypes from 525 animals was also carried out using high-density genotypes from the 724 animals used in our study as a reference using the BEAGLE software (Browning and Browning, 2009) (Materials and Methods in Supplementary Document 1). The average imputation accuracy for the animals was 0.98 (ranging

from 0.79–0.99). SNPs had an average imputation accuracy of 0.98 (ranging from 0.47–1.00). We repeated our GWAS using a combined dataset of both the high-density genotyped and imputed datasets (1,227 animals). The associations that were detected in the original GWAS failed to reach genome-wide significance in the combined dataset. As discussed below this may be due to a loss of power when including the imputed data.

DISCUSSION

This study identified a large number of SNPs significantly associated with SCS, a trait that is highly correlated with bovine mastitis, one of the most costly production diseases in dairy cattle. To the best of our knowledge this study is the first to use the Illumina BovineHD BeadChip to undertake a GWAS for SCS. The substantially increased SNP density of the BovineHD platform (1 SNP/~4,500 bp vs. 1 SNP/~62,000 bp for the BovineSNP50) appears to have provided increased resolution and power to detect QTL associated with SCS in this study (see **Figures 2–4**). Furthermore, this study enabled more precise mapping of both previously identified and novel QTL due to the increased SNP density. Twenty eight distinct QTL regions were defined and a number of promising candidate genes located in these QTL regions were identified.

The strongest association with SCS in our study was identified on chromosome 6. The only gene that appears to be located within this QTL encodes a rare type small non-coding RNA called a Y RNA. We have found that this Y RNA gene is 99% identical to human Y RNA 4 (RNY4) and was expressed but not differentially expressed in BME cells infected with a mastitis causing pathogens (Lawless et al., 2013). Further work is needed to characterize this gene and determine whether it plays a role in susceptibility to mastitis. Several previous studies have also implicated regions of chromosome 6 as being associated with SCS or mastitis, although none have identified the same QTL as uncovered in this study. Bennewitz et al., for example used microsatellite markers to identify a large QTL (~108 Mb) on BTA6 (Bennewitz et al., 2004). The size of the QTL, however, made it impossible to pinpoint particular candidate genes that might be associated with the trait. An association with CM near the casein gene complex at ~90 Mb on chromosome 6 has also been identified (Nilsen et al., 2009; Sodeland et al., 2011). This region was not significantly associated with SCS in our study, however. This may be explained by differences in the phenotypes used in the association study. Recently, Sahana et al. (2013) found a region on chromosome 6 extending from 83–93 Mb associated with CM which is closer to our detected QTL region; however, the association detected in our study is still ~6 Mb away from this region and so can be considered a distinct QTL region. A QTL region associated with SCS located at 78.7 Mb on BTA6 was also detected in a previous study of Irish Holstein-Friesian cattle (Meredith et al., 2012).

Our study also detected a novel QTL region located at 61.1–61.6 Mb on chromosome 13 associated with SCS. Previous studies have identified genetic markers associated with SCS nearby on chromosome 13 at 57.5–58.5 Mb (Lund et al., 2008; Sahana et al., 2013) but none of them overlap our associations. The QTL region in this study overlaps a well-known cluster of β -defensin genes and these genes, known for their host defence

function, would appear to be strong candidate genes for having a role in mastitis. β -defensin genes have also been shown to be expressed in bovine udder tissue in response to infection (Goldammer et al., 2004; Swanson et al., 2004) and expression levels of particular β -defensins have been shown to be positively correlated with SCS (Swanson et al., 2004). In addition, β -defensins are also believed to influence chemotaxis of white blood cells to the site of infection (Kaiser and Diamond, 2000; Ganz, 2003) which provides another mechanism by which these genes may play an important functional role in mastitis resistance in the dairy cow.

The strong, novel associations detected on chromosome 17 provided a number of other candidate genes which were in close proximity to the QTL region. The *MAPK1* gene located within 125 kb of this QTL is a member of the MAPK family which is a key pathway involved in the recruitment of leukocytes to the site of infection (Herlaar and Brown, 1999; Kaminska, 2005). However, a number of other nearby genes also have merit including (1) immunoglobulin lambda-like polypeptide 1 precursor (*IgLL-1*) which has been observed to have increased expression in bovine udder tissue in response to *Streptococcus uberis* challenge (Swanson et al., 2009), (2) Synaptosomal-associated protein 29 (*SNAP29*) whose increased expression has been linked to enhanced internalization and killing of *E. coli* by mast cells (Wesolowski et al., 2012), (3) Stromal cell-derived factor 2-like protein 1 (*SDF2L1*) which is involved in the processing of a wide range of defensins and (4) mir301b, a microRNA which has been shown to be significantly down-regulated in response to lipopolysaccharide (LPS) (Zheng et al., 2012).

A number of separate strongly associated QTL regions were identified on chromosome 20 in this study. Several studies have reported associations with SCS/CM with genetic markers on chromosome 20 (Sodeland et al., 2011; Meredith et al., 2012; Sahana et al., 2013) with many of them overlapping or nearby the regions detected in this study. Nevertheless, a number of possible candidate genes have been identified including the *IL31R α* gene which acts as part of a receptor complex for the *IL-31* cytokine in the activation of signaling pathways including the *JAK-STAT* and *MAPK* pathways which can initiate a wide range of immunological processes (Zhang et al., 2008; Cornelissen et al., 2012). The detection of a QTL region in close proximity to a number of genes of the complement system also looks promising given the importance of the complement system in both host innate and adaptive immunity.

The associated region on chromosome 1 from 49.9–50.5 Mb contained a single gene called *ALCAM*. This gene, also known as *CD166 antigen*, is a transmembrane receptor which has been examined extensively in human cancer research and has been implicated in leukocyte adhesion/migration and T cell activation (Hassan et al., 2004; Zimmermann et al., 2006; Cayrol et al., 2007). In addition, in a study of milk somatic cells which had been challenged with *Staphylococcus epidermidis* or *Staphylococcus aureus*, *ALCAM* was shown to be over-expressed in the milk somatic cells of a mastitis-resistant line of sheep in comparison with a mastitis-susceptible line (Bonnetfont et al., 2011). Therefore, *ALCAM* seems a plausible candidate gene to be involved in mastitis resistance.

Statistical power is a key issue in any GWAS and we endeavored to improve our study power via imputation of ungenotyped markers in additional animals. However, the associations were not significant on a genome-wide scale when using the combined dataset with the imputed genotypes. There are several possible reasons which may underlie this result. The imputation accuracy observed across all animals was high (0.98), however, a number of animals had an imputation accuracy score <0.90 with one animal having an accuracy score of 0.79. This variability in imputation accuracy may have occurred due to several reasons. Firstly, it is possible the size of our reference panel ($n = 724$) is inadequate to accurately impute ungenotyped SNPs for animals genotyped on the Bovine SNP50 chip leading to a loss of certainty of imputed genotypes. The accuracy of imputation can vary widely for a particular SNP depending on the surrounding LD structure (de Bakker et al., 2008) and this may have contributed to non-uniform imputation accuracy. Huang et al. (2009) looked at the relationship between the accuracy of imputation and the power of a subsequent association study using imputed markers. They found that even small decreases in the imputation accuracy can lead to large reductions in the power of GWAS and that a substantial increase in sample size would be required to maintain the same level of power. Given the variability in imputation accuracy that we observed in our study, this reduction in power due to imperfect imputation may explain the lack of associations with genome-wide significance observed when including the imputed animal in the association dataset.

CONCLUSIONS

The use of the Illumina High-density BeadChip afforded our study a SNP density and resolution far above that previously reported for a GWAS for SCS in dairy cattle. We have identified 138 SNPs significantly associated with SCS which make up 28 distinct QTL regions across 15 chromosomes. Several of these QTL regions are novel associations. Additionally, the increased SNP density in our study allowed us to more precisely map QTL intervals with some previously known QTL intervals being narrowed. Many of these strongly associated QTL regions encompass promising candidate genes such as the β -defensins on chromosome 13 and a Y RNA gene on chromosome 6. These QTL regions and their candidate genes can form the basis of more detailed research into the genetic mechanisms that underlie resistance to mastitis in the dairy cow.

AUTHOR CONTRIBUTIONS

Experimental design was carried out by Brian K. Meredith, David J. Lynn, Daniel G. Bradley and Donagh P. Berry. DNA extraction was carried out by Emma K. Finlay. Phenotypes were provided by Donagh P. Berry and Francis Kearney. Analysis was undertaken by Brian K. Meredith and David J. Lynn. Manuscript preparation was carried out by Brian K. Meredith, David J. Lynn, Daniel G. Bradley, Donagh P. Berry and Alan G. Fahey.

ACKNOWLEDGMENTS

This material is based upon works supported by the Science Foundation Ireland under Grant No. [09/IN.1/B2642]. We also acknowledge funding provided for this project through the

Teagasc Walsh Fellowship scheme and Teagasc RMIS 6018 as well as The Irish Department of Agriculture, Food and Marine Research Stimulus Fund (11/S/112). Further acknowledgments must go to the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support and to Per Madsen for his generous help with the DMU software.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2013.00229/abstract>

REFERENCES

- Aboyoun, P., Pages, H., and Lawrence, M. (2010). *GenomicRanges: Representation and Manipulation of Genomic Intervals*. R package version 1.10.7.
- Ashwell, M. S., Da, Y., Van Tassell, C. P., Vanraden, P. M., Miller, R. H., and Rexroad, C. E. Jr. (1998). Detection of putative loci affecting milk production and composition, health, and type traits in a United States Holstein population. *J. Dairy Sci.* 81, 3309–3314. doi: 10.3168/jds.S0022-0302(98)75896-5
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.2307/2346101
- Bennewitz, J., Reinsch, N., Guiard, V., Fritz, S., Thomsen, H., Looft, C., et al. (2004). Multiple quantitative trait loci mapping with cofactors and application of alternative variants of the false discovery rate in an enlarged granddaughter design. *Genetics* 168, 1019–1027. doi: 10.1534/genetics.104.030296
- Berry, D. P., Shalloo, L., Cromie, A. R., Veerkamp, R. F., Dillon, P., Amer, P. R., et al. (2007). “The economic breeding index: a generation on” in *Technical Report to the Irish Cattle Breeding Federation*, (Bandon; County Cork: ICBF), 29–34. Available online at: www.icbf.com
- Bloemhof, S., de Jong, G., and de Haas, Y. (2009). Genetic parameters for clinical mastitis in the first three lactations of Dutch Holstein cattle. *Vet. Microbiol.* 134, 165–171. doi: 10.1016/j.vetmic.2008.09.024
- Bonnefont, C., Toufeer, M., Caubet, C., Foulon, E., Tasca, C., Aurel, M.-R., et al. (2011). Transcriptomic analysis of milk somatic cells in mastitis resistant and susceptible sheep upon challenge with *Staphylococcus epidermidis* and *Staphylococcus aureus*. *BMC Genomics* 12:208. doi: 10.1186/1471-2164-12-208
- Brøndum, R., Su, G., Lund, M., Bowman, P., Goddard, M., and Hayes, B. (2012). Genome position specific priors for genomic prediction. *BMC Genomics* 13:543. doi: 10.1186/1471-2164-13-543
- Browning, B. L., and 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. doi: 10.1016/j.ajhg.2009.01.005
- Carlen, E., Strandberg, E., and Roth, A. (2004). Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish Holstein cows. *J. Dairy Sci.* 87, 3062–3070. doi: 10.3168/jds.S0022-0302(04)73439-6
- Carroll, M. C. (2004). The complement system in regulation of adaptive immunity. *Nat. Immunol.* 5, 981–986. doi: 10.1038/ni1113
- Cayrol, R., Wosik, K., Berard, J. L., Dodelet-Devillers, A., Ifergan, I., Kebir, H., et al. (2007). Activated leukocyte cell adhesion molecule promotes leukocyte trafficking into the central nervous system. *Nat. Immunol.* 9, 137–145. doi: 10.1038/ni1551
- Cole, J. B., VanRaden, P. M., O’Connell, J. R., Tassell, C. P. V., Sonstegard, T. S., Schnabel, R. D., et al. (2009). Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.* 92, 2931–2946. doi: 10.3168/jds.2008-1762
- Cormican, P., Meade, K. G., Cahalane, S., Narciandi, F., Chapwanya, A., Lloyd, A. T., et al. (2008). Evolution, expression and effectiveness in a cluster of novel bovine β -defensins. *Immunogenetics* 60, 147–156. doi: 10.1007/s00251-007-0269-8
- Cornelissen, C., Lüscher-Firzlaß, J., Baron, J. M., and Lüscher, B. (2012). Signaling by IL-31 and functional consequences. *Eur. J. Cell Biol.* 91, 552–566. doi: 10.1016/j.ejcb.2011.07.006
- de Bakker, P. I. W., Ferreira, M. A. R., Jia, X., Neale, B. M., Raychaudhuri, S., and Voight, B. F. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* 17, R122–R128. doi: 10.1093/hmg/ddn288

- de Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–1512. doi: 10.1534/genetics.107.084301
- Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nat. Rev. Immunol.* 3, 710–720. doi: 10.1038/nri1180
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80. doi: 10.1186/gb-2004-5-10-r80
- Goldammer, T., Zerbe, H., Molenaar, A., Schuberth, H. J., Brunner, R. M., Kata, S. R., et al. (2004). Mastitis increases mammary mRNA abundance of β -defensin 5, Toll-Like-Receptor 2 (TLR2), and TLR4 but not TLR9 in cattle. *Clin. Diagn. Lab. Immunol.* 11, 174–185. doi: 10.1128/CDLI.11.1.174-185.2004
- Gunther, J., Esch, K., Poschadel, N., Petzl, W., Zerbe, H., Mitterhuemer, S., et al. (2011). Comparative kinetics of *Escherichia coli*- and *Staphylococcus aureus*-specific activation of key immune pathways in mammary epithelial cells demonstrates that *S. aureus* elicits a delayed response dominated by Interleukin-6 (IL-6) but not by IL-1A or Tumor Necrosis Factor Alpha. *Infect. Immunol.* 79, 695–707. doi: 10.1128/IAI.01071-10
- Hassan, N. J., Barclay, A. N., and Brown, M. H. (2004). Frontline: optimal T cell activation requires the engagement of CD6 and CD166. *Eur. J. Immunol.* 34, 930–940. doi: 10.1002/eji.200424856
- Heringstad, B., Klemetsdal, G., and Ruane, J. (2000). Selection for mastitis resistance in dairy cattle: with focus on the situation in the Nordic countries. *Livest. Prod. Sci.* 64, 95–106. doi: 10.1016/S0301-6226(99)00128-1
- Herlaar, E., and Brown, Z. (1999). p38 MAPK signalling cascades in inflammatory disease. *Mol. Med. Today* 5, 439–447. doi: 10.1016/S1357-4310(99)01544-0
- Heyen, D. W., Weller, J. I., Ron, M., Band, M., Beever, J. E., Feldmesser, E., et al. (1999). A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiol. Genomics* 1, 165–175.
- Huang, L., Wang, C., and Rosenberg, N. A. (2009). The relationship between imputation error and statistical power in genetic association studies in diverse populations. *Am. J. Hum. Genet.* 85, 692–698. doi: 10.1016/j.ajhg.2009.09.017
- Kaiser, V., and Diamond, G. (2000). Expression of mammalian defensin genes. *J. Leukoc. Biol.* 68, 779–784.
- Kaminska, B. (2005). MAPK signalling pathways as molecular targets for anti-inflammatory therapy - from molecular mechanisms to therapeutic benefits. *Biochim. Biophys. Acta* 1754, 253–262. doi: 10.1016/j.bbapap.2005.08.017
- Khatkar, M., Nicholas, F., Collins, A., Zenger, K., Cavanagh, J., Barris, W., et al. (2008). Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics* 9:187. doi: 10.1186/1471-2164-9-187
- Khatkar, M., Thomson, P., Tammen, I., and Raadsma, H. (2004). Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 36, 163–190. doi: 10.1186/1297-9686-36-2-163
- Krude, T., Christov, C. P., Hyrien, O., and Marheineke, K. (2009). Y RNA functions at the initiation step of mammalian chromosomal DNA replication. *J. Cell Sci.* 122, 2836–2845. doi: 10.1242/jcs.047563
- Lawless, N., Foroushani, A. B. K., McCabe, M. S., O'Farrelly, C., and Lynn, D. J. (2013). Next generation sequencing reveals the expression of a unique miRNA profile in response to a gram-positive bacterial infection. *PLoS ONE* 8:e57543. doi: 10.1371/journal.pone.0057543
- Lerner, M. R., Boyle, J. A., Hardin, J. A., and Steitz, J. A. (1981). Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science* 211, 400–402. doi: 10.1126/science.6164096
- Lund, M. S., Guldbandsen, B., Buitenhuis, A. J., Thomsen, B., and Bendixen, C. (2008). Detection of quantitative trait loci in Danish Holstein cattle affecting clinical mastitis, somatic cell score, udder conformation traits, and assessment of associated effects on milk yield. *J. Dairy Sci.* 91, 4028–4036. doi: 10.3168/jds.2007-0290
- Madsen, P., and Jensen, J. (2010). *DMU: a User's Guide. A Package for Analysing Multivariate Mixed Models*. Version 6, release 5.0. Available Online at: http://dmu.agrsci.dk/dmuv6_guide.5.0.pdf.
- Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. *Nature* 449, 819–826. doi: 10.1038/nature06246
- Medzhitov, R., and Janeway, C. A. (2002). Decoding the patterns of self and nonself by the innate immune system. *Science* 296, 298–300. doi: 10.1126/science.1068883
- Meredith, B., Kearney, F., Finlay, E., Bradley, D., Fahey, A., Berry, D., et al. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genet.* 13:21. doi: 10.1186/1471-2156-13-21
- Mosig, A., Guofeng, M., Stadler, B.R.R., and Stadler, P. (2007). Evolution of the vertebrate Y RNA cluster. *Theory. Biosci.* 126, 9–14. doi: 10.1007/s12064-007-0003-y
- Mrode, R. A., and Swanson, G. J. T. (1996). Genetic and statistical properties of somatic cell count and its suitability as an indirect means of reducing the incidence of mastitis in dairy cattle. *Anim. Breed. Abstr.* 64, 847–857.
- Nilsen, H., Olsen, H. G., Hayes, B., Nome, T., Sehested, E., Svendsen, M., et al. (2009). Characterization of a QTL region affecting clinical mastitis and protein yield on BTA6. *Anim. Genet.* 40, 701–712. doi: 10.1111/j.1365-2052.2009.01908.x
- Paape, M. J., Bannerman, D. D., Zhao, X., and Lee, J. (2003). The bovine neutrophil: structure and function in blood and milk. *Vet. Res.* 34, 597–627. doi: 10.1051/vetres:2003024
- Pryce, J. E., Bolormaa, S., Chamberlain, A. J., Bowman, P. J., Savin, K., Goddard, M. E., et al. (2010). A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J. Dairy Sci.* 93, 3331–3345. doi: 10.3168/jds.2009-2893
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- R Core Team. (2012). *R: a Language and Environment for Statistical Computing*. Available online at: <http://www.R-project.org/>.
- Rupp, R., and Boichard, D. (2003). Genetics of resistance to mastitis in dairy cattle. *Vet. Res.* 34, 671–688. doi: 10.1051/vetres:2003020
- Sahana, G., Guldbandsen, B., Thomsen, B., and Lund, M. S. (2013). Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. *Anim. Genet.* 44, 620–626. doi: 10.1111/age.12053
- Schukken, Y. H., Günther, J., Fitzpatrick, J., Fontaine, M. C., Goetze, L., Holst, O., et al. (2011). Host-response patterns of intramammary infections in dairy cows. *Vet. Immunol. Immunopathol.* 144, 270–289. doi: 10.1016/j.vetimm.2011.08.022
- Smaragdov, M. (2006). Genetic mapping of loci responsible for milk production traits in dairy cattle. *Russ. J. Genet.* 42, 1–15. doi: 10.1134/S1022795406010017
- Sodeland, M., Grove, H., Kent, M., Taylor, S., Svendsen, M., Hayes, B., et al. (2011). Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle. *BMC Genet.* 12:70. doi: 10.1186/1471-2156-12-70
- Spencer, C. C. A., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* 5:e1000477. doi: 10.1371/journal.pgen.1000477
- Swanson, K., Gorodetsky, S., Good, L., Davis, S., Musgrave, D., Stelwagen, K., et al. (2004). Expression of a β -Defensin mRNA, lingual antimicrobial peptide, in bovine mammary epithelial tissue is induced by mastitis. *Infect. Immunol.* 72, 7311–7314. doi: 10.1128/IAI.72.12.7311-7314.2004
- Swanson, K. M., Stelwagen, K., Dobson, J., Henderson, H. V., Davis, S. R., Farr, V. C., et al. (2009). Transcriptome profiling of *Streptococcus uberis*-induced mastitis reveals fundamental differences between immune gene expression in the mammary gland and in a primary cell culture model. *J. Dairy Sci.* 92, 117–129. doi: 10.3168/jds.2008-1382
- Verhagen, A. P. M., and Pruijn, G. J. M. (2011). Are the Ro RNP-associated Y RNAs concealing microRNAs? Y RNA-derived miRNAs may be involved in autoimmunity. *Bioessays* 33, 674–682. doi: 10.1002/bies.201100048
- Wesolowski, J., Caldwell, V., and Paumet, F. (2012). A novel function for SNAP29 (Synaptosomal-Associated Protein of 29 kDa) in mast cell phagocytosis. *PLoS ONE* 7:e49886. doi: 10.1371/journal.pone.0049886
- Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., et al. (1998). Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics* 149, 1959–1973.
- Zhang, Q., Putheti, P., Zhou, Q., Liu, Q., and Gao, W. (2008). Structures and biological functions of IL-31 and IL-31 receptors. *Cytokine Growth Factor Rev.* 19, 347–356. doi: 10.1016/j.cytogfr.2008.08.003
- Zheng, Y., Xiong, S., Jiang, P., Liu, R., Liu, X., Qian, J., et al. (2012). Glucocorticoids inhibit lipopolysaccharide-mediated inflammatory response by downregulating

- microRNA-155: a novel anti-inflammation mechanism. *Free Radic.Biol. Med.* 52, 1307–1317. doi: 10.1016/j.freeradbiomed.2012.01.031
- Zimin, A., Delcher, A., Florea, L., Kelley, D., Schatz, M., Puiu, D., et al. (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10, R42. doi: 10.1186/gb-2009-10-4-r42
- Zimmerman, A. W., Joosten, B., Torensma, R., Parnes, J. R., van Leeuwen, F. N., and Figdor, C. G. (2006). Long-term engagement of CD6 and ALCAM is essential for T-cell proliferation induced by dendritic cells. *Blood* 107, 3212–3220. doi: 10.1182/blood-2005-09-3881

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 August 2013; paper pending published: 20 September 2013; accepted: 17 October 2013; published online: 06 November 2013.

Citation: Meredith BK, Berry DP, Kearney F, Finlay EK, Fahey AG, Bradley DG and Lynn DJ (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. doi: 10.3389/fgene.2013.00229

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2013 Meredith, Berry, Kearney, Finlay, Fahey, Bradley and Lynn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.