ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12886

Identification of candidate genes affecting chronic subclinical mastitis in Norwegian Red cattle: combining genome-wide association study, topologically associated domains and pathway enrichment analysis

E. Kirsanova*, B. Heringstad[†,‡], A. Lewandowska-Sabat* and I. Olsaker*

*Department of Basic Sciences and Aquatic Medicine, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway. [†]Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Oslo, Norway. [‡]Geno Breeding and A.I. Association, Hamar, Norway.

Summary

The aim of this study was to identify genes associated with chronic subclinical mastitis (SCM) in Norwegian Red (NR) cattle. Twelve SCM traits defined based on fixed threshold for test-day somatic cell count (SCC) were, together with lactation-average somatic cell score (LSCS) used for association and pathway enrichment analyses. A GWAS was performed on 3795 genotyped NR bulls with 777K SNP data and phenotypic information from 7 300 847 test-day SCC observations from 3 543 764 cows. At 5% chromosome-wide significance level 36 unique SNP were detected to be associated with one or more of the traits. These SNPs were analysed for linked genes using genomic positions of topologically associated domains (TAD). For the SCM traits with SCC >50 000 and >100 000 cells/ml on two test-days in a row and LSCS, the same top significant genes were identified – checkpoint clamp loader component (*RAD17*) and cyclin B1 (*CCNB1*). The SCM traits with SCC >250 000, 300 000, 350 000 or 400 000 cells/ml on two test-days in a row and D400 (number of days before the first case with SCC >400 000 cells/ml) displayed similar top significant genes: acyl-CoA thioesterase 2 and 4 (*ACOT2*; *ACOT4*). For the traits SCM200_3 (SCC >200 000 cells/ml on three test-days in a row) and SCM150, SCM200 (SCC >150 000; 200 000 cells/ml on two test-days in a row) a group of chemokine (C–X–C motif) ligand genes and the Fos proto-oncogene, AP-1 transcription factor subunit (*FOS*) gene, were identified. Further functional studies of these identified candidate genes are necessary to clarify their actual role in development of chronic SCM in NR cattle.

Keywords genome-wide association study, somatic cell count

Introduction

Chronic subclinical mastitis (SCM) is a complex trait characterised by high somatic cell count (SCC) for a prolonged period of time, often with SCC> 200 000 cells/ml (Harmon, 1994). In Norway, SCC has been recorded since the 1970s in the Norwegian Dairy Herd Recording System (Østerås *et al.* 2007). The data have been used in genetic analyses (e.g. Ødegård *et al.* 2003; Ødegård *et al.* 2004; Svendsen & Heringstad 2006; Haugaard *et al.* 2013) and associations studies (e.g. Sodeland *et al.* 2011; Olsen *et al.* 2016). GWASs (Delvin & Risch 1995) have been widely used for identification of chromosomal regions and for mapping of functional traits such as SCC (*e.g.* Kuhn *et al.* 2003; Sahana *et al.* 2010; Sodeland *et al.* 2011) and clinical mastitis (CM) in dairy cattle (*e.g.* Sahana *et al.* 2013; Wu *et al.* 2015; Olsen *et al.* 2016). In Norwegian Red (NR) cattle, significant QTL that affect somatic cell score have been reported on bovine chromosome (BTA–*Bos Taurus* autosome) 12, 19 and 26 (Sodeland *et al.* 2011), and for NR cattle on BTA5, -6, -13, -16, -19 and -20 (Sahana *et al.* 2014). GWAS combined with pathway-based analysis is often used to understand biological functions involved in complex traits. There are several methods used to select the genes for pathway analysis. Using topologically associated

Address of correspondence

E. Kirsanova, Department of Basic Sciences and Aquatic Medicine, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway.
E-mail: elena.kirsanova@nmbu.no

Accepted for publication 10 November 2019

domains (TADs; Dixon *et al.* 2012), representing functional regulatory regions of the genome can improve the search for causative variants compared with using arbitrarily chosen distances from significant SNPs (Wang *et al.* 2018). TADs are highly conserved domains with self-interacting chromatin, and constitute one type of functional annotation of the regulatory structure of the genome (Krefting *et al.* 2018). TAD regions are known to be stable during replication (Pope *et al.* 2014) and they are conserved between different cell types and between species (Dixon *et al.* 2012; Rao *et al.* 2014; Rudan *et al.* 2015). The stability and impact on gene regulation makes TADs an important tool for identification of underlying gene regulation and mechanisms of a trait or disease. In the bovine genome, TADs were first described and analysed by Wang *et al.* (2018). TAD coordinates from available humans, mice, dogs and macaques from several cell types were converted to the bovine genome with 79.63–98.88% mapped. TADs at least 200 kb wide were included to avoid small homological genomic fragments incapable of forming TADs in the bovine genome. Coordinates for TAD regions based on bovine sequence data are not yet available (September 2019). However, the analysis by Wang *et al* (2018) showed that the liftover TAD regions provide useful information about causative regulatory variants in the bovine genome and constitute a sensible search space.

The aim of this study was to identify genes with potential effect on chronic SCM in NR cattle using GWAS on 777K SNP data and the alternative SCM traits described by Kirsanova *et al.* (2019) and subsequently perform enrichment analysis using ingenuity pathway analysis (IPA) with genes identified within TAD regions.

Materials and methods

Phenotypic data

Phenotypic test-day SCC records from 1979 to 2016 for NR cows were obtained from the national Norwegian Dairy Herd Recording System and edited using SAS version 9.4 (SAS Inst. Inc., Cary, NC, USA; www.sas.com). Lactations 1–3 with two or more test-days in a row within a 2-month period, from 3 543 764 cows with 7 300 847 observations, were analysed. The DMU trace (Madsen 2012) was used to build the corresponding pedigree file of 4 126 678 animals.

Traits and statistical analyses

Kirsanova *et al.* (2019) defined 12 alternative SCM traits, and estimated genetic parameters for these together with lactation-average somatic cell score (LSCS). Briefly, there were eight binary SCM traits defined based on whether or not two test-day SCCs in a row were above the given SCC threshold, with threshold from 50 000 to 400 000 cells/ml (SCM50, -100, -150, -200, -250, -300, -350 and -400); two similar traits were based on three test-day SCC records in a row above 200 000 and 400 000 cells/ml respectively (SCM200_3 and SCM400_3); and finally, two traits were defined as the number of days before the first case with SCM50 or SCM400 (D50 and D400). All of the defined traits were within the days in-milk period between 21 and 305 days after calving. Linear animal repeatability models with estimated variance components from Kirsanova *et al.* (2019) were used for analyses in the DMU5 procedure in the DMU (Madsen & Jensen 2013). Solutions for fixed and random effects were used in calculations of daughter-yield-deviation (DYD; VanRaden & Wiggans 1991) for NR sires. DYDs are daughter phenotypes corrected for fixed and non-genetic random effects. The following model was used to calculate yield deviation (YD) for each cow for each trait:

$$\begin{aligned} \mathrm{YD} = {} & \mathrm{trait} - \mathrm{year\_month} - \mathrm{age} - \mathrm{days\_open} \\ & - \mathrm{herd\_year} - pe - 0.5\,\mathrm{EBV}_{\mathrm{mor}} \end{aligned}$$

where year_month is calving year/month; age is age at calving in month by lactation number; and days_open is days open grouped (each 10 days for each lactation number). Non-genetic random effects (herd_year, herd year of calving; *pe*, permanent environment effect) and 0.5 of dam's estimated breeding value ($\mathrm{EBV}_{\mathrm{mor}}$) were used to correct for genetic effect of dam in the model. All the fixed and random effects have been described by Kirsanova *et al.* (2019). The DYDs for sires were calculated as the arithmetic mean of the daughter's YD.

Genotypes

Sires were genotyped with the Affymetrix 25K SNP array (Affymetrix, Santa Clara, CA, USA), or the Illumina BovineSNP50 BeadChip (54K) (Illumina Inc., San Diego, CA, USA; Matukumalli *et al.* 2009) and combined with the Illumina BovineHD Genotyping BeadChip (777K). Geno SA (www.geno.no), the breeding organisation for NR cattle, performed imputation of missing genotypes. Genotypes from 3795 NR bulls with 613 908 SNP located on 29 chromosomes passed the quality control. A MAF of 1%, a minimum call rate of 95% and an individual call rate of 85% were used. The genome assembly *Bos taurus* UMD 3.1.1 (Zimin *et al.* 2009) was used for location of the SNP positions.

GWAS

GWAS analyses were performed using the R version 3.4.1 package GenABEL version 1.8-0 (http://www.r-project.org/; Aulchenko *et al.* 2007). DYDs were used as phenotypes in the GWAS. GRAMMA-Gamma function, a variance component-based two-step method, was used for unbiased estimation of SNP effects (Svischeva *et al.* 2012). The polygenic function included the kinship relationship matrix calculated and provided by Geno SA, based on 35 652 genotyped SNPs. The *P*-values were corrected by dividing

the observed test statistic by genomic inflation factor estimated in GenABEL. Chromosome-wide significance level was defined according to Sahana *et al.* (2010). The 5% chromosome-wide significance threshold ranged from the point-wise *P*-value of 1.29×10^{-6} on BTA1 to 4.31×10^{-6} on BTA25. Therefore *P*-values of $<5 \times 10^{-6}$ were required for significance. To create Manhattan plots the qqman package version 0.1.4 was used (Turner 2014).

TAD and pathway analysis

Genomic TAD coordinates from mouse cortex (Shen *et al.* 2012) and embryonic stem cells in the mm9 genome (Dixon *et al.* 2012) in addition to TAD coordinates from human embryonic stem cells and foetal lung IMR90 fibroblasts (IMR90) in the hg18 (Dixon *et al.* 2012) were obtained from Bing Ren's Lab, University of California, San Diego (downloaded from: http://chromosome.sdsc.edu/mouse/hi-c/download.html). Using the UCSC Batch Conversion program (liftOver tool; Kuhn *et al.* 2013) provided on the UCSC Genome Browser page (http://genome.ucsc.edu/), all of the TAD coordinates were converted (default settings, as described by Wang *et al.* 2018) to the bovine reference genome *Bos taurus* UMD3.1.1 (UCSC Genome Browser assembly ID: bosTau8). The genome version UMD3.1.1 has more correctly closed gaps and 173 contigs defined as contaminants and suppressed, compared with the version UMD3.1 used in Wang *et al.* 2018. Both human and mouse genomes were converted from hg18 and mm9 to the bosTau8 (NCBI Assembly ID 189361) through hg38 (NCBI Assembly ID 5800238) and mm10 (NCBI Assembly ID 327618) respectively. Bedtools coverage version 2.29.0 (Quinlan & Hall 2010) was used to compute the percentage of converted TADs that covered the reference genome *Bos Taurus* UMD3.1.1 (presented in Supplementary Appendix S1). Additional summary statistics of TAD mapping are described in Table 1. The length of the bovine TADs varied from 275 805 to 5 775 052 bp and from 271 661 to 6 763 506 bp, converted from the human and mouse genome respectively. TAD coordinates with at least 200 kb length (Wang *et al.* 2018) that matched converted TAD from the human and mouse genomes for all types of tissue were defined as stable and used for further analysis. For two SNP no annotated genes were found within the TAD regions containing the SNP. In order to include these in further analyses close TAD regions were used. For trait SCM400_3, SNP BTA2:84213972, BovineHD0200024041, the TAD region was located 100 158 bp from the SNP and for trait SCM50, SNP BTA20:10493654, BovineHD2000003312, the TADs were located 475 073 bp from the SNP. The number of genes detected using the UCSC Genome Browser across the defined traits ranged from 11 to 28 and from 4 to 76, based on the human and mouse TAD position respectively.

The list of genes for each trait was analysed using IPA (https://www.qiagenbioinformatics.com/). IPA is a database based on human gene ID, presenting top canonical pathways, upstream regulators and functional networks. In IPA, Fisher's extract-test at 5% level was used for calculation of significant *P*-values.

Results

A total of 36 significant SNPs across all 13 traits were identified by GWAS; 20 of these were unique for only one trait, whereas 16 were significantly associated with more than one trait (Table 2). Significant SNPs for SCM in NR cattle were found on BTA2, -5, -6, -7, -10, -12, -20 and -26 for the alternative SCM traits, whereas for LSCS there were found only on BTA20. The highest number of SNPs was observed on BTA20 (12 SNPs), followed by BTA12 (eight SNPs), BTA26 (six SNPs) and BTA10 (five SNPs). Only one or two SNPs were identified on BTA2, -5, -6 and -7. The SNPs that showed significant association to most traits was BovineHD1000025012 on BTA10, affecting eight of the 13 traits, followed by BovineHD2600010426 on BTA26, affecting seven traits. The rest of the identified SNPs showed significant association with one to six traits. SNPs located on BTA10 affected 10 of the analysed traits, followed by nine on BTA26 and eight on BTA12 and BTA20. Manhattan plots for each trait at 5 and 1% significant chromosome-wide thresholds are presented in Fig. 1.

In total 181 unique genes were identified using the bovine TAD approach. Many genes were associated with more than one SCM trait, and the number of identified associated genes ranged from 15 (LSCS) to 100 (SCM150; SCM200) per trait. Subsequently IPA mapped 168 genes. The genes found by IPA in the top five significant pathways for each trait are presented in Table 3. Additional information on the top five canonical pathways with upstream regulators, *P*-values and gene ID for all the traits are presented in Table S1 and Table S2.

Some of the analysed traits showed highly overlapping IPA results, with the same genes and pathways identified. The results for SCM150 were identical to those for SCM200, and so were results for SCM250 and SCM300. For five traits (SCM250, SCM300, SCM350, SCM400 and D400) acyl coenzyme A (CoA) hydrolysis, part of the metabolism of fatty acids, was identified as the top significant pathway involving the acyl-CoA thioesterase 2 (*ACOT2*) and acyl-CoA thioesterase 4 (*ACOT4*) genes. For three traits (SCM50, SCM100 and LSCS) DNA damage-induced 14-3-3 σ signalling, responsible for control of the biological activity in the cell cycle, was identified as the top pathway involving the checkpoint clamp loader component enzyme (*RAD17*) and cyclin B1 kinase (*CCNB1*) genes. Retinoic acid receptor activation, which may act as transcription factor, was one of the genes in the top pathway detected for SCM150/SCM200. For the SCM200_3 the top pathway was a role of interleukin (IL) 17A in psoriasis with four significant genes from the chemokines family. For the last two traits,

Table 1 Topologically associated domains (TAD) statistics mapped to bovine reference genome *Bos taurus* UMD3.1.1.

Genome	Tissue	Reference assembly	Number of TADs	Percentage of TADs converted successfully	Percentage of the UMD3.1.1 covered by converted TADs
Human[1]	hESC[2]	hg18	3127		
		bostau8	2914	93.2 %	89.1 %
	IMR90[3]	hg18	2349		
		bostau8	2174	92.6 %	87.6 %
Mouse[1]	mESC[4]	mm9	2200		
		bostau8	1893	86.0 %	80.7 %
	Cortex	mm9	1519		
		bostau8	1319	86.8 %	77.4 %

The cell and tissue with input data on TADs from human and mouse genomes were converted to the latest version of the bovine reference genome *Bos taurus* UMD3.1.1. The number and percentage of TADs successfully converted to the bovine genome and percentage of the genome covered by converted TADs are shown.
[1]Human/mouse – data from study by Dixon *et al.* (2012).
[2]hESC – human embryonic stem cells.
[3]IMR90 – foetal lung IMR90 fibroblasts.
[4]mESC – mouse embryonic stem cells.

SCM400_3 and D50, only one gene for each trait was identified within any canonical pathways, hence no significant pathways could be assigned (Table 3 with corresponding abbreviations in List S1). The top networks for the traits (presented with score in Table S3) were mainly associated with developmental disorders (SCM150/SCM200, SCM350), abnormalities (SCM250/SCM300, D50, D400, SCM400) and cellular growth (LSCS) in addition to cell death and survival (SCM50, SCM100) and cell-to-cell signalling (SCM200_3).

Discussion

Associated SNP positions

QTL located on eight different chromosomes were identified across the 13 analysed SCM traits. Other authors reported QTL for SCC at the same chromosomes BTA5, -6, -7, -10, -12, -20 and -26 for different breeds (Klungland *et al.* 2001; Kuhn *et al.* 2003; Sahana *et al.* 2010; Sodeland *et al.* 2011; Meredith *et al.* 2012; Meredith *et al.* 2013; Raven *et al.* 2014; Sahana *et al.* 2014). However, even if QTL on the same chromosome were reported in other studies and breeds, positions were different in NR cattle. Possible reasons include slightly different traits, different breeds and populations. To the best of our knowledge, only the previous GWAS analysis by Kirsanova *et al.* (2018) for the equally defined SCM50–SCM400 traits based on 35 605 SNP showed similar results. Some false-negative associations are also expected, even if a significant level applied (Korte & Farlow 2013). The GRAMMAR-Gamma function used in this study is a mixed-model-based method that is supposed to estimate unbiased SNP effects (Svischeva *et al.* 2012). Ekine *et al.* (2014) reported the GRAMMAR function to have the lowest type 1 error rate compared with other methods used for GWASs. That method can be recognised as conservative, and the false-positive rate in this case was lower than 0.04.

Analyses of biological processes and pathways underlying the traits

The genes identified in this study were based on TAD positions converted to bovine from mouse and human. Owing to the limited amount of available data and few detected significant genes, TAD data from both these two species were combined. The first study describing TAD positions successfully converted from several species to the bovine genome was Wang *et al.* (2018). Moreover, Wang *et al.* (2018) concluded that TAD positions can be used as a search space for causative regulatory variants and also around significant SNPs. In the current study TAD positions from the human genome mapped better to the bovine genome compared with those from mouse, in agreement with Wang *et al.* (2018).

In Norway *Staphylococcus aureus* is the major bacterium causing SCM (TINE SA 2017). This bacterium is Gram-positive and the cell wall components – lipoteichoic acid or peptidoglycan – are assumed to trigger activation of immune response through expression of interleukin 8 (IL-8; Bannerman *et al.* 2004). In the current study, *IL-8* or C–X–C motif chemokine ligand 8 also known as *CXCL8* was identified as an associated gene for the SCM200_3 trait in all of the five top canonical pathways (Table S1). Moreover, *IL-8* is located within a TAD region and 375 132 bp from a significant SNP identified by GWAS. Also, IPA has defined the *IL-1* gene as a one of the central genes in the top networks for SCM200_3 (Table S3). Moreover, *CXCL1*, *CXCL3* and *CXCL5* genes in addition to *CXCL8* were found for the SNPs placed on BTA6 within the TAD region. The top canonical pathway involving those genes was a role of *IL-17A* in psoriasis. *IL-17A* is a pro-inflammatory cytokine,

Table 2 Significant SNP detected by GWAS for alternative subclinical mastitis (SCM) traits[1].

BTA[2]	*N*[3]	SNP	Position (bp)[4]	Minor allele A1[5]	Major allele A2[6]	Trait – see text for further explanations
2	1	BovineHD0200021808	75996043	C	T	SCM400_3
2	2	BovineHD0200024041	84213972	G	A	SCM400_3
5	3	BovineHD0500024021	84931066	G	A	SCM350
6	4	BovineHD0600024697	90184750	T	C	SCM200_3
7	5	BovineHD0700010862	37686316	A	G	D400
10	6	BovineHD1000025012	87958006	T	C	SCM150, SCM200, SCM250, SCM300, SCM350, SCM400, SCM200_3, D400
10	7	BovineHD1000025014	87970294	A	G	SCM400, SCM200_3, D400
10	8	BovineHD1000027284	94432205	T	C	SCM200_3
10	9	ARS-BFGL-NGS-54789	94425952	A	G	SCM200_3
10	10	BovineHD1000027289	94462757	T	C	SCM100, SCM200_3, SCM400_3
12	11	BovineHD1200027006	54131183	T	C	SCM100, SCM150, SCM200, SCM250, SCM300, SCM200_3
12	12	BovineHD1200014960	54126754	C	T	SCM100, SCM150, SCM200, SCM300, SCM200_3
12	13	BovineHD1200014962	54143111	C	T	SCM100, SCM150, SCM200, SCM300, SCM200_3
12	14	BovineHD4100009623	54124043	A	G	SCM100, SCM150, SCM200, SCM200_3
12	15	BovineHD1200006048	20069484	A	C	SCM50, D50
12	16	BovineHD4100009623	19138863	G	A	D50
12	17	BovineHD1200005799	19142767	G	T	D50
12	18	BovineHD1200006023	19980842	C	T	SCM50
20	19	BovineHD2000003312	10493654	C	T	SCM50, SCM100, SCM150, LSCS
20	20	BovineHD2000003308	10486199	A	G	SCM50, SCM100, SCM150, SCM200, LSCS
20	21	BTB-00772821	10486993	C	T	SCM50, SCM100, SCM150, SCM200, LSCS
20	22	BovineHD2000003309	10488085	C	T	SCM50, SCM100, SCM150, SCM200, LSCS
20	23	BovineHD2000003310	10489403	A	G	SCM50, SCM100, SCM150, SCM200, LSCS
20	24	BTB-00772795	10435385	A	C	SCM50, SCM100
20	25	BovineHD2000003311	10491752	T	C	SCM50, SCM100, SCM150, SCM200
20	26	BovineHD4100014652	32962559	A	C	D50
20	27	BovineHD2000009441	32966444	C	T	D50
20	28	BovineHD2000017367	61885151	C	A	SCM400_3
20	29	BovineHD2000017376	61894720	A	G	SCM400_3
20	30	BovineHD2000017384	61902955	T	C	SCM400_3
26	31	BovineHD2600010426	37927026	G	A	SCM100, SCM150, SCM200, SCM250, SCM300, SCM350, SCM400
26	32	BovineHD2600010455	38009502	G	A	SCM400_3
26	33	BovineHD2600010690	39029294	T	C	SCM400_3
26	34	BovineHD2600012556	44579917	G	A	D400
26	35	BovineHD2600012565	44604386	T	C	D400
26	36	BovineHD2600012933	45869170	T	C	D400

[1]Traits – SCM50, -100, -150, -200, -250, -300, -350 and -400, subclinical mastitis above the threshold on two test-days at 50 000, 100 000, 150 000, 200 000, 250 000, 300 000, 350 000 and 400 000 cells/ml respectively; SCM200_3 and SCM400_3, subclinical mastitis traits above the threshold 200 000 and 400 000 on three test-days; D50 and D400, number of days before the first case with SCM50 and SCM400 respectively. LSCS, Lactation-average somatic cell score during 1–3 lactations.
[2]BTA, *Bos Taurus* autosome.
[3]*N*, Number of significant SNP.
[4]Position (bp) –position in base pair.
[5]A1, Minor allele.
[6]A2, Major allele.

known to take part in human chronic skin inflammation, contributing to the inflow of neutrophils, dendritic cells and memory T-cells (Nograles *et al.* 2008; Pfaff *et al.* 2017). This corresponds to the activation of CXC chemokines for SCM200_3, involved in the recruitment of neutrophils to the inflamed or infected tissue (Kobayashi 2006). The *IL-8* gene corresponds with the assumed activation of immune response by the bacterium as described by Bannerman *et al.* (2004). *CXCL1*, also known as *GRO1*, has previously been found to be induced in response to *S. aureus*, together with *CXCL2*, *CXCL3* and *CXCL8* (Rainard *et al.* 2008; Bougarn *et al.* 2010). The last associated gene, *CXCL5*, was previously identified in a microarray analysis by Pareek *et al.* (2005) as an important gene in initiating the innate immune response to *Escherichia coli* infection in bovine mammary epithelial cells. Additionally, for the SCM200_3 trait, *IL-1* was found to be associated. *IL-1* is an important receptor on macrophages, providing signalling for initiation of immunity and the inflammation processes.

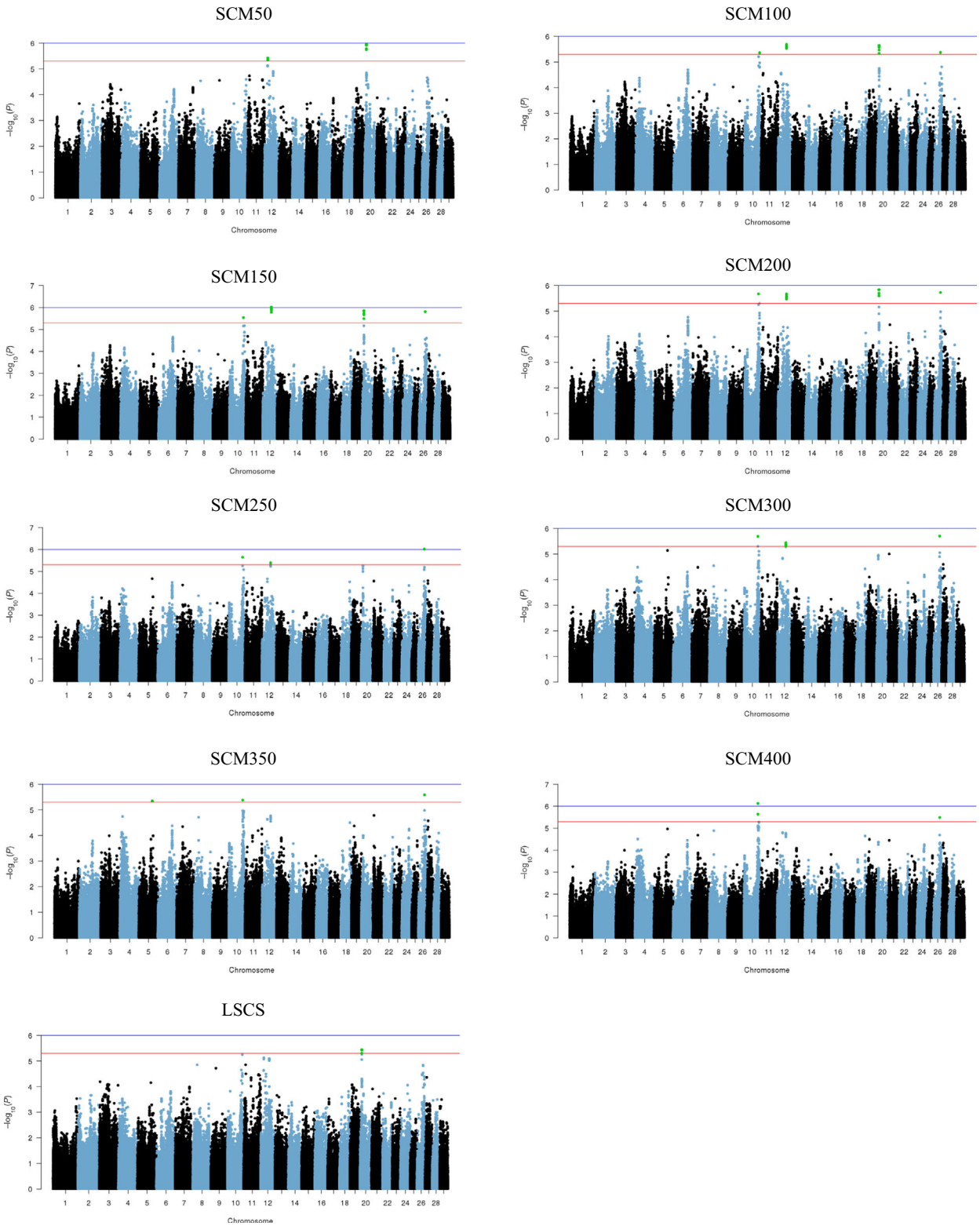

Figure 1 Manhattan plots from GWAS for alternative subclinical mastitis traits showing significant SNPs with genome-wide corrected *P*-values. The red horizontal line indicates defined chromosome-wide significance threshold on −log10(*P*) = 5.30103 ($P = 5 \times 10^{-6}$) at 5%. The blue horizontal line indicates threshold on −log10(*P*) = 6.0 ($P = 1 \times 10^{-6}$) at 1%. Number of significant SNPs – see Table 1 for further explanations.

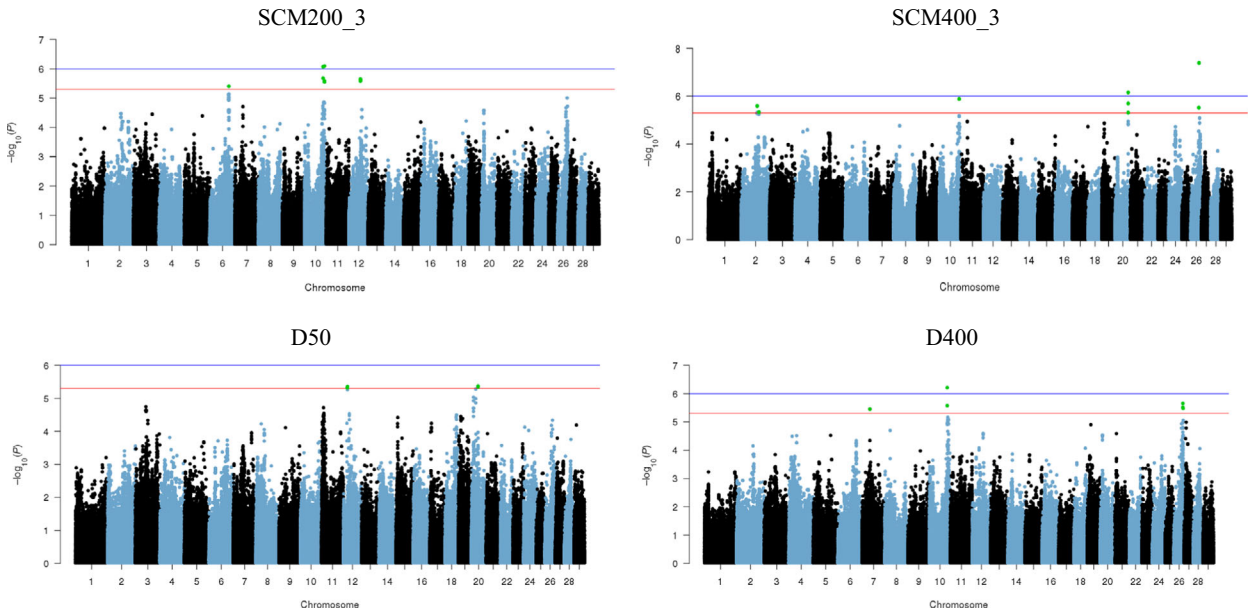

Figure 1 (Continued)

Interesting results were obtained for almost all of the alternative SCM traits with the thresholds higher than 250 000 cells/ml, which definitely indicate SCM. Acyl-CoA, a main associated canonical pathway for five of the analysed traits (SCM250, SCM300, SCM350, SCM400 and D400) involves the *ACOT2* and *ACOT4* genes, both placed 2.5 Mb from the significant SNP on BTA10. These genes, found to have an important role in bovine milk fat synthesis regulation during lactation (Strillacci *et al.* 2014), may be associated with decreased milk production in chronic SCM cases. For other traits with SCC threshold lower than 100 000 cells/ml (SCM50 and SCM100) and for LSCS, the canonical pathway of DNA damage-induced 14-3-3 σ signalling appears with two associated genes, *RAD17* and *CCNB1*, on BTA20 at 193 and 52 kb respectively, away from the significant SNP. Both genes are involved in regulation of the cell cycle process (Griffiths *et al.* 1995). *RAD17* is required for repair in response to DNA damage, which is important for correct duplication of DNA before cell division, whereas activation of the *CCNB1* gene decides that mitosis will be activated (http://www.ncbi.nlm.nih.gov/). Moreover, DNA repair genes are known to be involved in the inflammatory response during infections (Roth *et al.* 2014; Spanou *et al.* 2017). The phosphoinositide 3-kinase (*PI3K*) complex is a central molecule in the networks identified to be associated to LSCS, SCM50, SCM100 and SCM350. *PI3K* is most known in association with cancer, but it is a part of cellular function, e.g. cell growth, survival, motility and migration, which also corresponds to the increasing number of cells during udder inflammation (Harmon 1994). Nuclear factor kappa B (*NFkB*) was shown to be one of the central molecules in the identified networks for SCM100 as well as for SCM400_3 and D50. *NFkB* is a part of the classic macrophages activated by *S. aureus* (Lewandowska-Sabat *et al.* 2013). By entering the macrophages the bacterium initiates activation of *NFkB*, which activates the pro-inflammatory cytokines and the following inflammation response. Boutet *et al.* (2007) confirmed that inflammatory responses in bovine mammary epithelial cells occur via *NFkB* activation. For the traits SCM250, SCM400, SCM200_3 and D400, the most central molecules were found to be extracellular signal-regulated kinase 1 and 2 (*ERK 1/2*). *ERK* are members of MAPK1 (*ERK2*) and MAPK3 (*ERK1*), which are involved in regulation of the *NFkB* (complex) and *FOS* genes, in addition to having a role in survival, migration and cell death. Yonezawa *et al.* (2008) reported the importance of *ERK 1/2* genes in combination with *Akt* gene (serine/threonine kinase) in the bovine mammary epithelial cells via unsaturated fatty acid proliferation. In the current study, the *Akt* gene was also identified in one of the networks for the SCM200_3 and SCM250 traits as well as in a combination of *Akt*, *FOS* and *NFkB* as central genes in the network detected by IPA for D400. Moreover, *ERK 1/2* is activated by *S. aureus* infection (Ellington *et al.* 2001). Finally, for SCM150/200 the *FOS* gene appears as a main component of the top network and as one of the associated genes for SCM250/300, SCM350, SCM400, SCM200_3 and D400. *FOS* is placed at 86 887 kb on BTA10, which is more than 1 Mb from the significant SNP position (87 958 kb). The *FOS* gene is a transcription factor, and has previously been found in several Ingenuity canonical pathways associated with *S. aureus* infection in macrophages (Lewandowska-Sabat *et al.* 2013). This gene was significant for

Table 3 Alternative subclinical mastitis (SCM) traits[1] with detected associated genes by ingenuity pathway analysis.

N	Trait	Gene ID
1	SCM50	RAD17, CCNB1, TAF9, CDK7, KPNA3, PIK3R1
2	SCM100	RAD17, CDK7, CCNB1, TAF9, GRK5
3[2]	SCM150/200	FOS, PIK3R1, CDK7, TGFB3, SNW1, ACOT2, ACOT4, RAD17, CCNB1, ARHGEF12, PSEN1
4[2]	SCM250/300	ACOT2, ACOT4, NUMB, PSEN1, FOS, TGFB3, ALDH6A1, EDNRB, PGF
5	SCM350	ACOT2, ACOT4, FOS, ITPR2, KRAS, TGFB3, PGF
8	SCM400	ACOT2, ACOT4, NUMB, PSEN1, FOS, TGFB3, ALDH6A1, SNW1
9	LSCS	RAD17, CCNB1, CDK7, TAF9, PIK3R1
10	SCM200_3	CXCL8, CXCL3, CXCL1, CXCL5, CXCL2, PF4, FOS
11	SCM400_3	GLI2, SLC18A2, SEL1L, GLI2, STAT4
12	D50	NADP
13	D400	ACOT2, ACOT4, NUMB, PSEN1, FOS, TGFB3, ALDH6A1, SNW1

[1]Traits – SCM50, -100, -150, -200, -250, -300, -350, -400, subclinical mastitis above the threshold in two test-days at 50 000, 100 000, 150 000, 200 000, 250 000, 300 000, 350 000 and 400 000 cells/ml respectively; LSCS, lactation-average somatic cell score during 1–3 lactations; SCM200_3 and SCM400_3, subclinical mastitis traits above the threshold 200 000 and 400 000 on three test-days; D50 and D400, number of days before the first case with SCM50 and SCM400 respectively.
[2]3/4, Traits with identical results.

eight of the 13 analysed traits in the current study, which makes it a promising candidate involved in regulation of chronic SCM.

However, IPA is a database primary designed for the human pharmaceutical industry with specific weight on cancer, owing to a high level of activity in this area. Hence, it can be difficult to interpret the IPA results in relation to bovine chronic SCM. In addition, some of the bovine genes identified in the UCSC Genome Browser were unmapped by IPA and have not been included in the analysis.

Conclusion

Possible causative genes affecting chronic SCM in NR cattle were identified. The most significant genes were not the same across traits, which illustrates the complexity of chronic SCM. Alternatively defined traits can cover additional genetic information to improve genetic evaluation and selection for improved udder health in NR cattle. Moreover, TAD regions proved to make sense as alternative search space for candidate genes linked to significant SNP.

Acknowledgements

The Geno Breeding and AI Association and the Norwegian Dairy Herd Recording System are acknowledged for providing access to data. Dr Christos Dadousis is acknowledged for help with GWAS analyses. We thank Arne B. Gjuvsland for help on the SNP and TAD analysis strategies. The current study was funded by the Research Council of Norway project no. 233778.

Conflict of interest

The authors declare that they have no conflict of interest.

Availability of data

The phenotypes and SNP data used in this study are the property of Geno SA, the breeding organisation of Norwegian Red cattle. The phenotypes and SNP data will be made available by Geno SA after signing a Material Transfer Agreement, in which Geno SA can impose reasonable conditions, such as confidentiality.

References

Aulchenko Y.S., Ripke S., Isaacs A., van Duijn C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–6.

Bannerman D.D., Paape M.J., Lee J.W., Zhao X., Hope J.C., Rainard P. (2004) Escherichia coli and Staphylococcus aureus elicit differential innate immune responses following intramammary infection. *Clinical and Diagnostic Laboratory Immunology* **11**, 463–72.

Bougarn S., Cunha P., Harmache A., Fromageau A., Gilbert F.B., Rainard P. (2010) Muramyl dipeptide synergizes with Staphylococcus aureus lipoteichoic acid to recruit neutrophils in the mammary gland and to stimulate mammary epithelial cells. *Clinical and Vaccine Immunology* **17**, 1797–809.

Boutet P., Sulon J., Closset R., Detilleux J., Beckers J.F., Bureau F., Lekeux P. (2007) Prolactin-induced activation of nuclear factor kappaB in bovine mammary epithelial cells: role in chronic mastitis. *Journal of Dairy Science* **90**, 155–64.

Devlin B., Risch N. (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**, 311–22.

Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Hu M., Liu J.S., Ren B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–80.

Ekine C.C., Rowe S.J., Bishop S.C., de Koning D.J. (2014) Why breeding values estimated using familial data should not be used for genome-wide association studies. *G3 (Bethesda)* **4**, 341-7.

Ellington J.K., Elhofy A., Bost K.L., Hudson M.C. (2001) Involvement of mitogen-activated protein kinase pathways in Staphylococcus aureus invasion of normal osteoblasts. *Infection and Immunity* **69**, 5235–42.

Griffiths D.J., Barbet N.C., McCready S., Lehmann A.R., Carr A.M. (1995) Fission yeast rad17: a homologue of budding yeast RAD24 that shares regions of sequence similarity with DNA polymerase accessory proteins. *The EMBO Journal* **14**, 5812–23.

Harmon R.J. (1994) Symposium - mastitis and genetic evaluation for somatic-cell count - physiology of mastitis and factors affecting somatic-cell counts. *Journal of Dairy Science* **77**, 2103–12.

Haugaard K., Heringstad B., Whist A.C. (2013) Genetic associations between somatic cell score and pathogen-specific subclinical mastitis in Norwegian Red cows. *Journal of Animal Breeding and Genetics* **130**, 98–105.

Kirsanova E., Heringstad B., Lewandowska-Sabat A., Olsaker I. (2018) *Genome-wide Association Study for Alternative Subclinical Mastitis Traits in Norwegian Red*. World Congress on Genetics Applied to Livestock Production, http://www.wcgalp.org/system/files/proceedings/2018/genome-wide-association-study-alternative-subclinical-mastitis-traits-norwegian-red.pdf.

Kirsanova E., Heringstad B., Lewandowska-Sabat A., Olsaker I. (2019) Alternative subclinical mastitis traits for genetic evaluation in dairy cattle. *Journal of Dairy Science*;**102**:5323-9.

Klungland H., Sabry A., Heringstad B. *et al*. (2001) Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mammalian Genome* **12**, 837–42.

Kobayashi Y. (2006) Neutrophil infiltration and chemokines. *Critical reviews in immunology* **26**, 307–16.

Korte A., Farlow A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29.

Krefting J., Andrade-Navarro M.A., Ibn-Salem J. (2018) Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biology* **16**, 87.

Kuhn C., Bennewitz J., Reinsch N. *et al*. (2003) Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *Journal of Dairy Science* **86**, 360–8.

Kuhn R.M., Haussler D., Kent W.J. (2013) The UCSC genome browser and associated tools. *Briefings in Bioinformatics* **14**, 144–61.

Lewandowska-Sabat A.M., Boman G.M., Downing A., Talbot R., Storset A.K., Olsaker I. (2013) The early phase transcriptome of bovine monocyte-derived macrophages infected with Staphylococcus aureus in vitro. *BMC Genomics* **14**, 891.

Madsen P. (2012) *DmuTrace. A program for extracting the pedigree for a subset of animals from a larger pedigree. Version 2*. Aarhus University.

Madsen P., Jensen J. (2013) *A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models. Version 6, release 5.2*. Aarhus University.

Matukumalli L.K., Lawley C.T., Schnabel R.D. *et al*. (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* **4**, e5350.

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

Meredith B.K., Kearney F.J., Finlay E.K., Bradley D.G., Fahey A.G., Berry D.P., Lynn D.J. (2012) Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genetics* **13**, 21.

Nograles K.E., Zaba L.C., Guttman-Yassky E. *et al*. (2008) Th17 cytokines interleukin (IL)-17 and IL-22 modulate distinct inflammatory and keratinocyte-response pathways. *British Journal of Dermatology* **159**, 1092–102.

Olsen H.G., Knutsen T.M., Lewandowska-Sabat A.M. *et al*. (2016) Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production. *Genetics Selection Evolution* **48**, 79.

Pareek R., Wellnitz O., Van Dorp R., Burton J., Kerr D. (2005) Immunorelevant gene expression in LPS-challenged bovine mammary epithelial cells. *Journal of Applied Genetics* **46**, 171–7.

Pfaff C.M., Marquardt Y., Fietkau K., Baron J.M., Luscher B. (2017) The psoriasis-associated IL-17A induces and cooperates with IL-36 cytokines to control keratinocyte differentiation and function. *Scientific Reports* **7**, 15631.

Pope B.D., Ryba T., Dileep V. *et al*. (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–5.

Quinlan A.R., Hall I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–2.

Rainard P., Fromageau A., Cunha P., Gilbert F.B. (2008) Staphylococcus aureus lipoteichoic acid triggers inflammation in the lactating bovine mammary gland. *Veterinary Research* **39**, 52.

Rao S.S.P., Huntley M.H., Durand N.C. *et al*. (2014) A 3D Map of the human genome at Kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–80.

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

Roth S., Rottach A., Lotz-Havla A.S. *et al*. (2014) Rad50-CARD9 interactions link cytosolic DNA sensing to IL-1beta production. *Nature Immunology* **15**, 538–45.

Rudan M.V., Barrington C., Henderson S., Ernst C., Odom D.T., Tanay A., Hadjur S. (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports* **10**, 1297–309.

Sahana G., Guldbrandtsen B., Bendixen C., Lund M.S. (2010) Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Animal Genetics* **41**, 579–88.

Sahana G., Guldbrandtsen B., Thomsen B., Holm L.E., Panitz F., Brondum R.F., Bendixen C., Lund M.S. (2014) Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. *Journal of Dairy Science* **97**, 7258–75.

Sahana G., Guldbrandtsen B., Thomsen B., Lund M.S. (2013) Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. *Animal Genetics* **44**, 620–6.

Shen Y., Yue F., McCleary D.F. *et al*. (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–20.

Sodeland M., Kent M.P., Olsen H.G., Opsal M.A., Svendsen M., Sehested E., Hayes B.J., Lien S. (2011) Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. *Animal Genetics* **42**, 457–65.

Spanou E., Kalisperati P., Pateras I.S., Papalampros A., Barbouti A., Tzioufas A.G., Kotsinas A., Sougioultzis S. (2017) Genetic Variability as a regulator of TLR4 and NOD signaling in response to bacterial driven DNA damage response (DDR) and inflammation: focus on the gastrointestinal (GI) tract. *Frontiers in Genetics* **8**, 65.

Strillacci M.G., Frigo E., Canavesi F. *et al*. (2014) Quantitative trait loci mapping for conjugated linoleic acid, vaccenic acid and (9) -desaturase in Italian Brown Swiss dairy cattle using selective DNA pooling. *Animal Genetics* **45**, 485–99.

Svendsen M., Heringstad B. (2006) *Somatic cell count as an indicator of sub-clinical mastitis*. Genetic parameters and correlations with clinical mastitis. *Interbull Bulletin*, 12.

Svishcheva G.R., Axenovich T.I., Belonogova N.M., van Duijn C.M., Aulchenko Y.S. (2012) Rapid variance components-based method for whole-genome association analysis. *Nature Genetics* **44**, 1166–70.

TINE (2017) *TINE Rådgivning*. Statistikksamling 2017 [TINE Annual Satistics 2017]. Accessed 12 October 2018, Available at: URL (in Norwegian, with English Tables). https://medlem.tine.no/fagprat/helse/_attachment/427411?_ts=16167bec8ab.

Turner S.D. (2014) *qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots*. *bioRxiv* 10.1101/005165. doi: 10.1101/005165

VanRaden P.M., Wiggans G.R. (1991) Derivation, calculation, and use of national animal model information. *Journal of Dairy Science* **74**, 2737–46.

Wang M., Hancock T.P., Chamberlain A.J., Vander Jagt C.J., Pryce J.E., Cocks B.G., Goddard M.E., Hayes B.J. (2018) Putative bovine topological association domains and CTCF binding motifs can reduce the search space for causative regulatory variants of complex traits. *BMC Genomics* **19**, 395.

Wu X., Lund M.S., Sahana G., Guldbrandtsen B., Sun D., Zhang Q., Su G. (2015) Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. *Genetics Selection Evolution* **47**, 50.

Yonezawa T., Haga S., Kobayashi Y., Katoh K., Obara Y. (2008) Unsaturated fatty acids promote proliferation via ERK1/2 and Akt pathway in bovine mammary epithelial cells. *Biochemical Biophysical Research Communications* **367**, 729–35.

Zimin A.V., Delcher A.L., Florea L. *et al.* (2009) A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biology* **10**, R42.

Ødegard J., Heringstad B., Klemetsdal G. (2004) Short communication: bivariate genetic analysis of clinical mastitis and somatic cell count in Norwegian dairy cattle. *Journal of Dairy Science* **87**, 3515–7.

Ødegård J., Klemetsdal G., Heringstad B. (2003) Variance components and genetic trend for somatic cell count in Norwegian Cattle. *Livestock Production Science* **79**, 135–44.

Østeras O., Solbu H., Refsdal A.O., Roalkvam T., Filseth O., Minsaas A. (2007) Results and evaluation of thirty years of health recordings in the Norwegian dairy cattle population. *Journal of Dairy Science* **90**, 4483–97.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Top ingenuity canonical pathways.

Table S2 Gene ID detected based on topologically associated domains (TAD) positions.

Table S3 Top networks with score.

Appendix S1. Coordinates of converted TADs and percentage of the UMD3.1.1 genome covered by converted TADs.

List S1 List of abbreviations.