Animal, page 1 of 8 © The Animal Consortium 2019
doi:10.1017/S1751731119000338

animal

The genetic background of clinical mastitis in Holstein-Friesian cattle

J. Szyda[1,2†], M. Mielczarek[1,2], M. Frąszczak[1], G. Minozzi[3], J. L. Williams[4] and K. Wojdak-Maksymiec[5]

[1]*Biostatistics Group, Department of Genetics, Wroclaw University of Environmental and Life Sciences, Kozuchowska 7, 51-631 Wroclaw, Poland;* [2]*Institute of Animal Breeding, Krakowska 1, 32-083 Balice, Poland;* [3]*Department of Veterinary Medicine, Università degli Studi di Milano, Via Giovanni Celoria 10, 20133 Milano, Italy;* [4]*Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, SA 5371, Australia;* [5]*Department of Animal Genetics, West Pomeranian University of Technology, Doktora Judyma 26, 71-466 Szczecin, Poland*

(Received 15 September 2018; Accepted 28 January 2019)

Mastitis is an inflammatory disease of the mammary gland, which has a significant economic impact and is an animal welfare concern. This work examined the association between single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) with the incidence of clinical mastitis (CM). Using information from 16 half-sib pairs of Holstein-Friesian cows (32 animals in total) we searched for genomic regions that differed between a healthy (no incidence of CM) and a mastitis-prone (multiple incidences of CM) half-sib. Three cows with average sequence depth of coverage below 10 were excluded, which left 13 half-sib pairs available for comparisons. In total, 191 CNV regions were identified, which were deleted in a mastitis-prone cow, but present in its healthy half-sib and overlapped in at least nine half-sib pairs. These regions overlapped with exons of 46 genes, among which APP *(BTA1),* FOXL2 *(BTA1),* SSFA2 *(BTA2),* OTUD3 *(BTA2),* ADORA2A *(BTA17),* TXNRD2 *(BTA17) and* NDUFS6 *(BTA20) have been reported to influence CM. Moreover, two duplicated CNV regions present in nine healthy individuals and absent in their mastitis-affected half-sibs overlapped with exons of a cholinergic receptor nicotinic α 10 subunit on BTA15 and a novel gene (ENSBTAG00000008519) on BTA27. One CNV region deleted in nine mastitis-affected sibs overlapped with two neighbouring long non-coding RNA sequences located on BTA12. Single nucleotide polymorphisms with differential genotypes between a healthy and a mastitis-affected sib included 17 polymorphisms with alternate alleles in eight affected and healthy half-sib families. Three of these SNPs were located introns of genes:* MET *(BTA04),* RNF122 *(BTA27) and* WRN *(BTA27). In summary, structural polymorphisms in form of CNVs, putatively play a role in susceptibility to CM. Specifically, sequence deletions have a greater effect on reducing resistance against mastitis, than sequence duplications have on increasing resistance against the disease.*

Keywords: copy number variation, genomic annotation, single nucleotide polymorphism, somatic mutations, whole genome sequence

Implications

We compared whole genome DNA sequence of 13 half-sib pairs, discordant for their clinical mastitis (CM) status, but matched by age, age of calving and lactation season. Results indicated that deletions of genomic regions were more likely to be associated with increased susceptibility to CM than duplications of genomic regions and single nucleotide polymorphisms (SNPs). In total, 191 genomic regions deleted in a mastitis-prone sib and present in a healthy sib, observed in at least nine pairs, overlapped with exons of 46 genes with functions related to immune response and gene expression. Such regions can be used in selection for mastitis resistant cows.

† E-mail: joanna.szyda@upwr.edu.pl

Introduction

Mastitis is an inflammatory disease of the mammary gland that occurs in response to a physical injury or infection by pathogenic microorganisms, such as *Escherichia coli*, *Streptococcus uberis* and *Staphylococcus aureus* (Schukken *et al*., 2011). Following infection of the mammary gland, macrophages and epithelial cells release cytokines that cause the migration of neutrophils, monocytes and other leukocytes from the blood to the site of infection in the mammary tissue. The cost associated with mastitis in Europe, according to current estimates, is 1.55 billion € per year (European Union http://www.sabre-eu.eu/). The frequency and cost of mastitis, and rising public concerns for animal welfare, have made mastitis one of the most important diseases for the dairy sector (Thompson-Crispi et al., 2014).

There have been 292 genetic associations reported for risk of CM distributed across most chromosomes (Animal QTLdb www.animalgenome.org/www.animalgenome.org/QTLdb rel. 36. August 2018; Detilleux, 2009). Recent developments in genome sequencing technologies applied to livestock have facilitated the identification of copy number variations (CNVs) and millions of SNPs in a relatively cost efficient manner. This has enabled researchers to describe the genomic landscape of livestock species (e.g. Choi *et al*., 2014; Szyda *et al*., 2015; Mielczarek *et al*., 2017), and to combine this whole genome sequence data with phenotypic information for genomic prediction (VanRaden *et al*., 2017) or in genome-wide association studies to identify variations associated with various traits (Sanchez *et al*., 2016; Sanchez *et al*., 2017). Combining SNP microarray genotyping with whole genome DNA and RNA sequencing is beginning to disentangle the genomic regulation of economically important traits, such as mastitis susceptibility in dairy cattle (Fang *et al*., 2017).

The current work examined the association between single base pair (SNPs) and structural polymorphisms (CNVs), identified in whole genome DNA sequences, with the incidence of CM. Unlike previous studies, we compared SNP and CNV variation identified in sequence data from 16 paternal half-sibs that were selected to be discordant for susceptibility to CM.

Material and methods

Animals

Whole genome DNA sequences were produced for 32 cows of the Holstein-Friesian breed that were selected from a dataset of 991 individuals, which had veterinary records for incidence of CM. All the cows were kept in the same barn, managed in the same way, cared for by the same staff, fed the same balanced diet in the form of a total mixed ration, and milked in the same herringbone milking parlour. A potential case of CM was reported to the stud veterinarian by the milking staff. Then a veterinarian made the definitive diagnosis based on the following clinical symptoms: (i) alterations in consistency, colour and smell of milk; (ii) occurrence of redness, swelling, pain, tissue hardening, increased temperature of udder; (iii) deterioration of a cow's general condition manifested by fever, reduced feed intake, significant decrease in milk production and anxiety. In unresolved cases microbiological tests were performed. The 32 cows comprising 16 paternal half-sibs were matched for: (i) age – half-sibs born in the same year and season (the mean age difference between half-sib pairs was 15.36 days), (ii) age at calving, and (iii) year and season of the start of lactation (mean difference ±30.19 days), but discordant in their mastitis status. Mastitis resistant cows had never been diagnosed with mastitis and had been in the herd for a minimum of three full lactations. Mastitis-prone cows had several episodes of CM in the first two full lactations. The incidence of the disease in the study group ranged between 6 and 14 cases, with a mean of 8.00. Cows with mastitis were generally culled, therefore none of those remaining in the herd to the third and subsequent lactations had the highest incidence of mastitis.

Whole genome DNA sequence and bioinformatic pipelines

Whole genome sequences of the 32 cows were obtained using the Illumina HiSeq2000 Next Generation Sequencing platform. A detailed description of the protocol has been described previously (Szyda *et al*., 2015). The total number of raw reads generated for a single animal ranged between 164 984 147 and 472 265 620. Raw reads were aligned to the UMD3.1 reference genome using BWA-MEM software (Li and Durbin, 2009) with the number of aligned reads varying from 155 202 885 (94.07% of UMD3.1) to 454 412 859 (99.57% of UMD3.1) per animal and the average depth of coverage between 5× and 17× (Table 1). Resulting BAM files were processed using a combination of tools from the Picard (http://broadinstitute.github.io/picard/) and SAMtools (Li *et al*., 2009) packages and included re-alignment to the reference sequence as well as removal of PCR duplicates. In order to make use of the half-sib structure of the data set, the variants (SNPs and CNVs) were identified by comparison of healthy and mastitis-prone siblings. In particular, SNPs were identified by comparing the discordant sib pairs using the somatic mutation caller Varscan2 (Koboldt *et al*., 2012). 'Somatic' SNPs were called when a healthy cow was homozygous for a reference allele and the mastitis-prone half-sib was heterozygous or homozygous for the alternative allele. Differences in the number of reads were tested using the Fisher's Exact Test with Bonferroni multiple testing correction. The number of 'somatic' SNPs per half-sib pair ranged from 314 564 to 6 765 038 ($\overline{x} = 4\,932\,444 \pm 1\,615\,483$) (Supplementary figure S1). Copy number variations were identified

Table 1 *Average depth of coverage and per-read quality score for each of the Polish Holstein-Friesian cows half-sib pairs*

Mastitis prone-sibling		Healthy sibling	
Average coverage	Average quality	Average coverage	Average quality
12 ± 53	36.23	17 ± 84	35.00
16 ± 61	34.84	15 ± 62	35.86
10 ± 42	36.64	14 ± 66	35.92
9 ± 37	*36.77*	*17 ± 79*	*34.40*
8 ± 30	*36.76*	*16 ± 73*	*35.32*
13 ± 66	35.90	16 ± 73	35.31
11 ± 61	35.63	17 ± 83	35.06
10 ± 58	36.35	16 ± 78	35.02
17 ± 70	35.13	15 ± 68	35.69
11 ± 61	36.29	14 ± 88	35.53
14 ± 64	35.68	14 ± 76	30.96
16 ± 66	35.36	17 ± 11	33.16
16 ± 74	35.08	15 ± 87	30.43
15 ± 64	35.51	16 ± 84	31.90
16 ± 66	34.95	15 ± 94	30.79
16 ± 72	*34.79*	*5 ± 35*	*25.91*

Half-sibs-pairs removed from the analysis due to a low coverage are shown in italics.

using the CNVnator software (Abyzov *et al.*, 2011). Copy number variations longer than 1 Mbp were removed. To exclude false positive variants being a consequence of artefacts arising from errors in the reference genome, deletions shared by at least 15 cows were removed if they had a 50% or higher overlap with gaps in the reference genome. As CNVnator is based on differences in read depth along the genome, three cows with average depth of coverage below 10 and their respective half-sibs were excluded leaving 13 half-sib pairs which were used to identify CVNs. The total number of CNVs identified for the 26 individuals ranged between 1694 and 5187 ($\overline{x} = 4\,932\,444 \pm 1\,615\,483$) for duplicated sequences and between 9731 and 18 479 ($\overline{x} = 12\,787 \pm 1\,792$) for deleted sequences with lengths ranging between 200 to 724 000 bp for deletions and 200 to 439 300 bp for duplications (Supplementary Figure S2). Deleted sequences covered from 2.71% to 6.09% of the whole genome, while duplications accounted for a 0.51% to 1.58% increase in the total sequence length.

Quantifying the association of variants with clinical mastitis
Somatic SNPs were identified by comparing the number of reads supporting an alternative and a reference allele between a healthy and a mastitis-prone half-sib. The 'tumour purity' parameter was set to 0.75 to reflect the genotypic similarity between half-sibs. We focused on CNVs, which were deleted in a mastitis prone cow, but present in its healthy sibling, and on CNVs that were duplicated in a healthy cow, but not duplicated in its mastitis-prone sibling. Such genomic sequences were identified by comparing CNVs in each discordant sib-pair and defined as CNV regions (CNVRs).

Annotation
The genomic annotation of SNPs and CNVs exhibiting a differential polymorphism between a mastitis-prone and a healthy sib was performed using Variant Effect Predictor software (McLaren *et al.*, 2010). Functional, non-coding sequences were annotated based on the miRBase database (http://www.mirbase.org) for micro-RNA and the Noncode database (http://www.noncode.org), Billerey *et al.* (2014) and Koufariotis *et al.* (2015) for long non-coding RNA.

Results

Association of somatic single nucleotide polymorphisms with clinical mastitis
Altogether 672 272 SNPs were identified that showed allele differences between discordant sibs in the number of reads supporting an alternative and a reference allele, which were significant at the 10% type I error level. Among these only 17 SNPs were in common in at least eight half-sib families and were considered in further analysis. Three of these SNPs were located within introns of genes. The most common somatic SNP was found in 10 half-sib families and was located at the position 28 810 758 bp on BTA27 in an intron of *RNF122,* a RefSeq sequence corresponding to the gene coding for ring finger protein 122. Interestingly, ten other somatic SNPs located within this gene with alternate alleles were identified in seven or eight half-sib pairs. *RNF22* encodes a protein involved in basic processes such as protein–protein and protein–DNA interactions. It may therefore be expected to play a role in cell viability, and has been reported to have a role in immune response (Wang *et al.*, 2016). The second most common SNP, present in nine families, was identified on BTA04 at 51 991 924 bp and overlapped with an intron of the proto-oncogene, receptor tyrosine kinase (*MET,* ENSBTAG00000006161), which plays a role in cellular migration and invasion, processes which are related to inflammation. The third SNP, which was common in eight families, was also located on BTA27 (26 370 233 bp) within the *WRN* gene (ENSBTAG00000021592), which is responsible for the Werner syndrome and encodes Werner syndrome RecQ-like helicase protein. Mutations in this gene result in premature aging in humans and an increased susceptibility to infection (Domínguez-Gerpe and Araújo-Vilar, 2008).

Association of copy number variant regions with clinical mastitis
The structural variation examined were (i) genome duplications present predominantly in healthy animals in at least nine half-sib pairs, and (ii) genome deletions present predominantly in mastitis-prone animals, also in at least nine half-sib pairs. The biological hypothesis underlying strategy (i) was that healthy sibs may harbour a duplication in a region functionally important for CM resistance, which may then be associated with a higher level of expression, while in strategy (ii) deletion of functionally important genomic region may increase the risk of CM.

Altogether, 28 755 duplicated CNVRs were present in the mastitis-prone sibs and absent from the healthy sibs, among these only 36 (0.13%) CNVRs were present in at least nine sib-pairs. Two of the regions overlapped with exons of a gene coding for a cholinergic receptor nicotinic alpha 10 subunit on BTA15 and a novel *Bos taurus* gene with Ensembl ID ENSBTAG00000008519 located on BTA27. A more interesting picture emerges when regions deleted in mastitis-prone sibs were compared with healthy sibs. 141 066 such regions were identified including 191 (0.14%) deletions observed in at least nine pairs. Those regions were annotated as exonic sequences of 46 genes (Table 2). 25 of the regions represented novel deletions, while the reminder have already been reported in the Database of Genomic Variants (www.ebi.ac.uk/dgva), albeit not always as deletions. The length of the deleted regions varied between 100 and 15 200 bp and covered part of a single exon through to several exons. Seven of the affected genes potentially influence the incidence of mastitis (Figure 1). Genes with deletions found in the mastitis-prone sibs have been implicated in immune response and variations in somatic cell score, which is an

Table 2 *Copy number variant regions (CNVRs) within coding sequences of the* Bos taurus *genome, represented by duplications present in at least nine healthy cows, but absent in their mastitis-prone half-sibs and deletions present in at least nine mastitis-prone cows, but absent in their healthy half-sibs*

Chromosome	CNVR type	Begin (bp)	End (bp)	No. of differential sib pairs	Gene
1	Deletion	9 607 201	9 608 900	9	*APP*
1	Deletion	131 241 401	131 243 600	9	*FOXL2*
2	Deletion	14 751 101	14 751 200	9	*SSFA2*
2	Deletion	133 415 901	133 416 100	10	*OTUD3*
3	Deletion	16 287 401	16 291 900	9	*ATP8B2*
3	Deletion	34 270 401	34 270 900	9	*CELSR2*
4	Deletion	27 855 001	27 855 100	9	*TWIST1*
4	Deletion	69 321 501	69 324 300	9	*HOXA10*
4	Deletion	69 332 701	69 334 000	9	ENSBTAG00000011476
4	Deletion	69 340 101	69 342 800	9	*HOXA7*
4	Deletion	69 353 401	69 355 400	9	*HOXA5, HOXA6*
4	Deletion	77 241 501	77 255 800	9	*CCM2, NACAD*
5	Deletion	82 468 401	82 469 500	9	*KLHL42*
5	Deletion	90 959 701	90 961 000	9	*AEBP2*
5	Deletion	101 471 901	101 472 700	9	*PHC1*
6	Deletion	109 355 101	109 357 700	9	*UVSSA*
8	Deletion	102 987 401	102 987 700	9	*UGCG*
10	Deletion	32 861 601	32 862 000	9	*MEIS2*
10	Deletion	102 885 601	102 885 700	9	ENSBTAG00000045849
11	Deletion	100 038 001	100 038 100	9	ENSBTAG00000019513
11	Deletion	105 837 501	105 838 600	9	*TOR4A*
12	Deletion	32 305 701	32 305 701	9	*URAD*
12	Deletion	88 665 401	88 667 100	9	*IRS2*
13	Deletion	39 202 401	39 202 500	9	*SLC24A3*
13	Deletion	54 345 301	54 350 900	9	*SAMD10*
14	Deletion	45 828 001	45 828 200	9	*ZBTB10*
15	Duplication	52 193 201	52 196 000	9	*CHRNA10*
15	Deletion	53 582 301	53 582 500	9	*ARHGEF17*
16	Deletion	19 500 201	19 500 500	9	*KCTD3*
17	Deletion	68 419 101	68 423 200	9	*TFIP11, SRRD*
17	Deletion	73 578 301	73 586 000	9	*ADORA2A*
17	Deletion	74 708 501	74 708 600	10	*MRPL40*
17	Deletion	74 708 601	74 708 900	9	*MRPL40*
17	Deletion	74 891 701	74 892 600	9	*TXNRD2*
18	Deletion	54 208 601	54 208 700	9	*DACT3*
18	Deletion	54 209 401	54 209 800	9	*DACT3*
20	Deletion	70 990 101	70 991 000	9	*NDUFS6*
22	Deletion	1 877 601	1 877 800	9	*SLC4A7*
24	Deletion	5 357 201	5 358 300	9	*NETO1*
24	Deletion	40 768 401	40 768 800	9	*PTPRM*
25	Deletion	478 401	479 800	9	*CAPN15*
26	Deletion	14 446 201	14 461 400	10	*CYP26A1, CYP26C1*
26	Deletion	25 928 401	25 929 100	10	ENSBTAG00000046499
26	Deletion	25 929 101	25 929 700	9	ENSBTAG00000046499
27	Duplication	4 799 101	4 799 400	9	ENSBTAG00000008519
29	Deletion	44 573 901	44 575 200	9	*OVOL1*

ADORA2A = Adenosine A2a receptor; *AEBP2* = AE binding protein 2; *APP* = Amyloid β precursor protein; *ARHGEF17* = ρ guanine nucleotide exchange factor 17; *ATP8B2* = ATPase phospholipid transporting 8B2; *CAPN15* = Calpain 15; *CCM2* = CCM2 scaffold protein; *CELSR2* = Cadherin, EGF LAG seven-pass G-type receptor 2; *CHRNA10* = Cholinergic receptor nicotinic α 10 subunit; *CYP26A1* = Cytochrome P450, family 26, subfamily A, polypeptide 1; *CYP26C1* = Cytochrome P450 family 26 subfamily C member 1; *DACT3* = Dishevelled-binding antagonist of β catenin 3; *FOXL2* = Forkhead box L2; *HOXA5* = Homeobox A5; *HOXA6* = Homeobox A6; *HOXA7* = Homeobox A7; *HOXA10* = Homeobox A10; *IRS2* = Insulin receptor substrate 2; *KCTD3* = Potassium channel tetramerization domain containing 3; *KLHL42* = Kelch-like family member 42; *MEIS2* = Meis homeobox 2; *MRPL40* = Mitochondrial ribosomal protein L40; *NACAD* = NAC α domain containing; *NDUFS6* = *NADH*:ubiquinone oxidoreductase subunit S6; *NETO1* = Neuropilin and tolloid like 1; *OTUD3* = OTU deubiquitinase 3; *OVOL1* = Ovo-like transcriptional repressor 1; *PHC1* = Polyhomeotic homolog 1; *PTPRM* = Protein tyrosine phosphatase, receptor type M; *SAMD10* = Sterile α motif domain containing 10; *SLC24A3* = Solute carrier family 24 member 3; *SLC4A7* = Solute carrier family 4 member 7; *SRRD* = SRR1 domain containing; *SSFA2* = ITPR interacting domain containing 2; *TFIP11* = Tuftelin-interacting protein 11; *TOR4A* = Torsin family 4 member A; *TWIST1* = Twist family bHLH transcription factor 1; *TXNRD2* = Thioredoxin reductase 2; *UGCG* = UDP-glucose ceramide glucosyltransferase; *URAD* = Ureidoimidazoline (2-oxo-4-hydroxy-4-carboxy-5-) decarboxylase; *UVSSA* = UV stimulated scaffold protein A; *ZBTB10* = Zinc finger and BTB domain containing 10

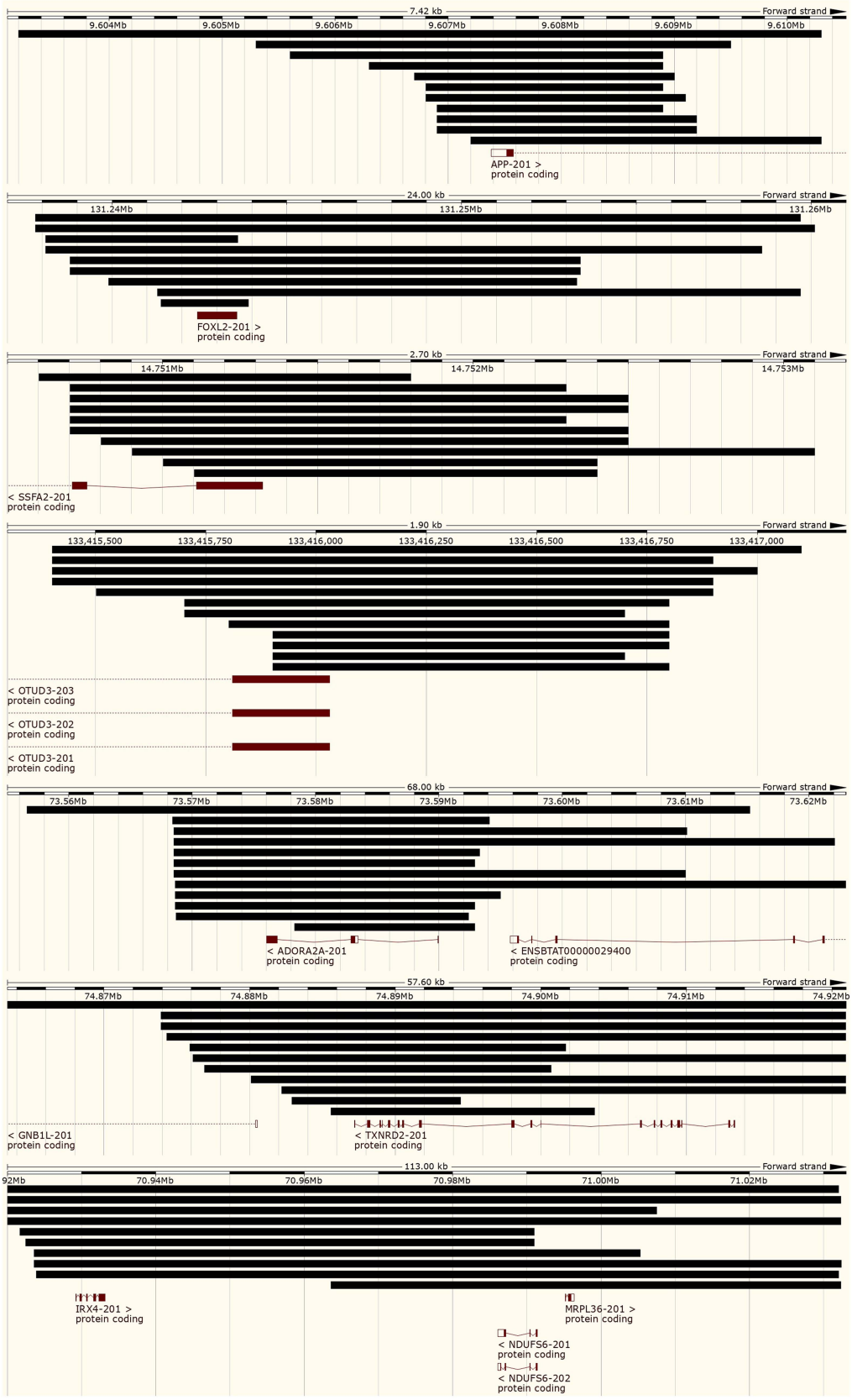

Figure 1 Positions of copy number variant regions deleted in mastitis-prone siblings in *Bos taurus* genome covering exons of *APP* and *FOXL2* (chromosome 1), *SSFA2* and *OTUD3* (chromosome 2), *ADORA2A* and *TXNRD2* (chromosome 17), *NDUFS6* (chromosome 20) genes. *APP* = Amyloid β precursor protein; *FOXL2* = Forkhead box L2; *SSFA2* = ITPR interacting domain containing 2; *OTUD3* = OTU deubiquitinase 3; *ADORA2A* = Adenosine A2a receptor; *TXNRD2* = Thioredoxin reductase 2; *NDUFS* = NADH:ubiquinone oxidoreductase subunit S6.

indicator trait for mastitis. *FOXL2* coding Forkhead box L2 protein (ENSBTAG00000031277), located on BTA1, had a 2 200 bp deleted region in the only exon of this gene, which was observed in nine of the mastitis-prone sibs but was not present in the healthy half-sibs. The forkhead transcription factor, encoded by *FOXL2*, plays a role in inflammation processes (Moumné *et al*., 2008). Sperm specific antigen 2 (*SSFA2*; ENSBTAG00000000937), located on BTA2 had a short deletion of 100 bp in the first exon in nine mastitis-prone sibs. Variations in this gene are associated with somatic cell score, which is an indicator trait for mastitis (Strillacci *et al*., 2014). *SSFA2* also overlaps with a QTL for bovine immunoglobulin G level (QTL ID: 20470). NADH-ubiquinone oxidoreductase subunit S6 protein gene (*NDUFS6*; ENSBTAE00000106646) is located on BTA20 and had a 900 bp deletion that removes a whole exon 2 and has been reported previously (Boussaha *et al*., 2015). This genomic region harbours a quantitative trait locus for somatic cell score (Durán Aguilar *et al*., 2016; Animal QTL database ID: 122069). OTU Deubiquitinase 3 protein gene (*OTUD3*; ENSBTAG00000017108), located on BTA2 contained a 200 bp deletion in 10 mastitis-prone compared with healthy sibs. This deletion removes a part of the first exon and has also been reported previously (Keel *et al*., 2017). *OTUD3* plays a role in inflammation and is associated with inflammatory bowel disease in humans (e.g. McGovern *et al*., 2010). Adenosine A2a receptor gene (*ADORA2A*; ENSBTAG00000016944), located on BTA17 had a deletion of 7700 bp covering the whole exon 2 in nine mastitis-prone sibs. This deletion was previously reported by Hou *et al*. (2011). *ADORA2A* is involved in inflammatory diseases by modulating tissue response to inflammation (Salmon *et al*., 1993) and is assigned to the GO terms: inflammatory response biological process (GO:0006954) and a negative regulation of inflammatory response biological process (GO:0050728). *ADORA2A* is expressed at high levels in the mammary gland (Yue *et alet al*., 2014). Two other genes with deletions have been associated with susceptibility to bacterial infection. Thioredoxin reductase 2, mitochondrial precursor protein gene (*TXNRD2*; ENSBTAG00000043581) is located on BTA17 and had a 900 bp deletion in nine mastitis-prone sibs, that removes a part of the 12th exon. A CNV in this region was also reported by Hou *et al*. (2011). *TXNRD2* is a candidate gene influencing susceptibility to *S. aureus*, which is one of the most common causes of mastitis, in mice (Ye *et al*., 2014), humans (Nelson *et al*., 2014) and cattle (Ghorbani *et al*., 2015). Amyloid β precursor protein gene (*APP*; ENSBTAG00000017753) located on BTA1, had a 1700 bp deletion in first exon, which was observed in nine sib-pair comparisons. The gene product is known to have bactericidal and antifungal activities and amyloid proteins have been shown to be associated with sub-CM in ruminants (Miglio *et al*., 2013). One CNVR, deleted in nine mastitis-prone sibs overlapped two neighbouring long non-coding RNA sequences located on BTA12 (90 641 959 to 90 643 802 bp and 90 644 418 to 90 646 031 bp) reported by Koufariotis *et al*. (2015).

Discussion

Our comparison of the genomes of mastitis resistant and mastitis-prone cows suggested that deletions of genomic regions were more likely to be associated with increased susceptibility to the disease. The function of genes in the affected regions was enriched in terms related to immune response and gene expression (RNA polymerase II core promoter proximal region sequence-specific DNA binding GO:0000978 and transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding GO:0000982). Similar functions were reported in the analysis of genes that were up-regulated after an artificially induced intra-mammary infection with lipopolysaccharide (Fang *et al*., 2017). These included, among others, genes related to RNA processing, the regulation of gene expression and the inflammatory response. There has been strikingly low reproducibility across other studies carried out to disentangle the genetic background of mastitis resistance in dairy cattle (e.g. Sodeland *et al*., 2011; Sahana *et al*., 2014; Olsen *et al*., 2016; Wang *et al*., 2016). The only common feature emerging were genes related to immune response. This disparity reflects the multifactorial nature of the disease, which makes identification of genes involved difficult. In our study, we reduced variation in management by working within a single cohort at the same farm and minimised the genetic variation by studying half-sib pairs, which share the same allele IBD with 0.5 probability. Moreover, the advances in technologies for whole genome sequencing have increased the proportion of variants identified in a genome, which can then be tested for association traits, by the availability of whole genome sequences with potentially (ignoring technical error in variant detection) all point and structural variants present in a genome. In this study we identified variants differing between healthy and mastitis-prone sibs and identified putative causal polymorphisms and candidate genes for CM. However, production of whole genome sequence still has a high cost and the management of the large data sets poses logistical problems. As a result studies using whole genome sequence generally have a small sample size that do not permit rigorous statistical testing of associations, so that the number of false associations (type I errors) cannot be determined. This is the case for the small data set used here, and also means that some true associations could have been missed. Data quality also presents problems, as some regions could have been missed or have a sequence depth that was insufficient to determine variants. Nevertheless, the results presented here provide information on regions putatively involved in mastitis susceptibility and are a starting point for further studies.

Conclusions

The present study identified genomic regions and genes, which potentially harbour causal mutations contributing to mastitis susceptibility. Many of these regions contain structural polymorphisms in the form of CNVs within genes

involved in immune response, that are likely to play a role in disease susceptibility. Interestingly sequence deletions were most frequently found in the mastitis-prone sibs, suggesting that loss of function reduces resistance against CM, rather than sequence duplication, increases susceptibility to mastitis. An interesting further extension of these findings will be estimating effects of the polymorphisms at a population-wide level.

Acknowledgements

The research was supported by the European Union Seventh Framework Programme through the NADIR (FP7-228394) project, by the Polish National Science Centre (NCN) grant 2014/13/B/NZ9/02016, and by The Leading National Research Centre (KNOW) programme for 2014–2018. Computations were carried out at the Poznan Supercomputing and Networking Centre.

Declaration of interest

The authors declare that they have no competing interests

Ethics statement

Not applicable.

Software and data repository resources

Deoxyribonucleic acid sequences of the 32 cows are available from the NCBI BioProject database under the accession ID PRJNA359667.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731119000338

References

Abyzov A, Urban AE, Snyder M and Gerstein M 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Research 21, 974–984.

Billerey C, Boussaha M, Esquerré D, Rebours E, Djari A, Meersseman C, Klopp C, Gautheret D and Rocha D 2014. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. BMC Genomics 15, 499.

Boussaha M, Esquerré D, Barbieri J, Djari A, Pinton A and Letaief R 2015. Genome-wide study of structural variants in Bovine Holstein, Montbéliarde and Normande dairy breeds. PLoS One 10, 8.

Choi J-W, Liao X, Stothard P, Chung WH, Jeon HJ, Miller SP, Choi S, Lee JK, Yang B, Lee KT, Han KJ, Kim HC, Jeong D, Oh JD, Kim N, Kim TH, Lee HK and Lee SJ 2014. Whole-genome analyses of Korean native and Holstein cattle breeds by massively parallel sequencing. PLoS One 9, 101127.

Detilleux JC 2009. Genetic factors affecting susceptibility to udder pathogens. Veterinary Microbiology 134, 157–164.

Domínguez-Gerpe L and Araújo-Vilar D 2008. Prematurely aged children: molecular alterations leading to Hutchinson-Gilford Progeria and Werner syndromes. Current Aging Science 1, 202–212.

Durán Aguilar M, Román Ponce SI, Ruiz López FJ, González Padilla E, Vásquez Peláez CG, Bagnato A and Strillacci MG 2016. Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. Journal of Animal Breeding and Genetics 134, 49–59.

Fang L, Sahana G, Su G, Yu Y, Zhang S, Lund MS and Sørensen P 2017. Integrating sequence-based GWAS and RNA-Seq provides Novel insights into the genetic basis of mastitis and milk production in dairy cattle. Scientific Reports 7, 45560.

Ghorbani S, Tahmoorespur M, Masoudi Nejad A, Nasiri M and Asgari Y 2015. Analysis of the enzyme network involved in cattle milk production using graph theory. Molecular Biology Research Communications 4, 93–103.

Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM, Sonstegard TS and Van Tassell CP 2011. Genomic characteristics of cattle copy number variations. BMC Genomics 12, 127.

Keel BN, Keele JW and Snelling WM 2017. Genome-wide copy number variation in the bovine genome detected using low coverage sequence of popular beef breeds. Animal Genetics 48, 141–150.

Koboldt D, Zhang Q, Larson D, Shen D, McLellan M, Lin L, Miller C, Mardis E, Ding L and Wilson R 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research 22, 568–576.

Koufariotis LT, Chen YP, Chamberlain A, van der Jagt C and Hayes BJ 2015. A catalogue of novel bovine long noncoding RNA across 18 tissues. PLoS One 10, e0141225.

Li H and Durbin R 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078–2079.

McGovern DP, Gardet A, Törkvist L, Goyette P, Essers J, Taylor KD, Neale BM, Ong RT, Lagacé C, Li C, Green T, Stevens CR, Beauchamp C, Fleshner PR, Carlson M, D'Amato M, Halfvarson J, Hibberd ML, Lördal M, Padyukov L, Andriulli A, Colombo E, Latiano A, Palmieri O, Bernard EJ, Deslandres C, Hommes DW, de Jong DJ, Stokkers PC, Weersma RK, NIDDK IBD, Consortium G, Sharma Y, Silverberg MS, Cho JH, Wu J, Roeder K, Brant SR, Schumm LP, Duerr RH, Dubinsky MC, Glazer NL, Haritunians T, Ippoliti A, Melmed GY, Siscovick DS, Vasiliauskas EA, Targan SR, Annese V, Wijmenga C, Pettersson S, Rotter JI, Xavier RJ, Daly MJ, Rioux JD and Seielstad M 2010. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nature Genetics 42, 332–337.

McLaren W, Pritchard B, Rios D, Chen Y, Flicek P and Cunningham F 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. Bioinformatics 26, 2069–2070.

Mielczarek M, Fraszczak M, Giannico R, Minozzi G, Williams JL, Wojdak-Maksymiec K and Szyda J 2017. Analysis of copy number variations in Holstein-Friesian cow genomes based on whole genome sequence data. Journal of Dairy Science 100, 5515–5525.

Miglio A, Moscati L, Fruganti G, Pela M, Scoccia E, Valiani A and Maresca C 2013. Use of milk amyloid A in the diagnosis of subclinical mastitis in dairy ewes. Journal of Dairy Research 80, 496–502.

Moumné L, Batista F, Benayoun B, Nallathambi J, Fellous M, Sundaresan P and Veitia RA 2008. The mutations and potential targets of the forkhead transcription factor FOXL2. Molecular and Cellular Endocrinology 282, 2–11.

Nelson CL, Pelak K, Podgoreanu MV, Ahn SH, Scott WK, Allen AS, Cowell LG, Rude TH, Zhang Y, Tong A, Ruffin F, Sharma-Kuinkel BK and Fowler VG Jr 2014. A genome-wide association study of variants associated with acquisition of Staphylococcus aureus bacteremia in a healthcare setting. BMC Infectious Diseases 14, 83.

Olsen HG, Knutsen TM, Lewandowska-Sabat AM, Grove H, Nome T, Svendsen M, Arnyasi M, Sodeland M, Sundsaasen KK, Dahl SR, Heringstad B, Hansen HH, Olsaker I, Kent MP and Lien S 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.

Sahana G, Guldbrandtsen B, Thomsen B, Holm LE, Panitz F, Brondum RF, Bendixen C and Lund MS 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, 7258–7275.

Salmon JE, Brogle N, Brownlie C, Edberg JC, Kimberly RP, Chen BX and Erlanger BF 1993. Human mononuclear phagocytes express adenosine A1 receptors.

A novel mechanism for differential regulation of Fc gamma receptor function. Journal of Immunology 151, 2775–2785.

Sanchez M-P, Govignon-Gion A, Croiseau P, Fritz S, Hozé C, Miranda G, Martin P, Barbat-Leterrier A, Letaïef R, Rocha D, Brochard M, Boussaha M and Boichard D 2017. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. Genetics Selection Evolution 49, 68.

Sanchez M-P, Govignon-Gion A, Ferrand M, Gelé M, Pourchet D, Amigues Y, Fritz S, Boussaha M, Capitan A, Rocha D, Miranda G, Martin P, Brochard M and Boichard D 2016. Whole-genome scan to detect quantitative trait loci associated with milk protein composition in 3 French dairy cattle breeds. Journal of Dairy Science 99, 8203–8215.

Schukken YH, Günther J, Fitzpatrick J, Fontaine MC, Goetze L, Holst O, Leigh J, Petzl W, Schuberth HJ, Sipka A, Smith DG, Quesnell R, Watts J, Yancey R, Zerbe H, Gurjar A, Zadoks RN and Seyfert HM 2011. Host-response patterns of intramammary infections in dairy cows. Veterinary Immunology and Immunopathology 144, 270–289.

Sodeland M, Kent MP, Olsen HG, Opsal MA, Svendsen M, Sehested E, Hayes BJ and Lien S 2011. Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. Animal Genetics 42, 457–465.

Strillacci MG, Frigo E, Schiavini F, Samoré AB, Canavesi F, Vevey M, Cozzi MC, Soller M, Lipkin E and Bagnato A 2014. Genome-wide association study for somatic cell score in Valdostana Red Pied cattle breed using pooled DNA. BMC Genetics 15, 106.

Szyda J, Frąszczak M, Mielczarek M, Giannico R, Minozzi G, Nicolazzi EL, Kamiński S and Wojdak-Maksymiec K 2015. The assessment of inter-individual variation of whole-genome DNA sequence in 32 cows. Mammalian Genome 26, 658–665.

Thompson-Crispi K, Atalla H, Miglior F and Mallard BA 2014. Bovine mastitis: frontiers in immunogenetics. Frontiers in Immunology 7, 493.

VanRaden PM, Tooker ME, O'Connell JR, Cole JB and Bickhart DM 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. Genetics Selection Evolution 49, 32.

Wang W, Jiang M, Liu S, Zhang S, Liu W, Ma Y, Zhang L, Zhang J and Cao X 2016. RNF122 suppresses antiviral type I interferon production by targeting RIG-I CARDs to mediate RIG-I degradation. PNAS 113, 9581–9586.

Wang XG, Ju ZH, Hou MH, Jiang Q, Yang CH, Zhang Y, Sun Y, Li RL, Wang CF, Zhong JF and Huang JM 2016. Deciphering transcriptome and complex alternative splicing transcripts in mammary gland tissues from cows naturally infected with Staphylococcus aureus mastitis. PLoS One 11, e0159719.

Ye Z, Vasco DA, Carter TC, Brilliant MH, Schrodi SJ and Shukla SK 2014. Genome wide association study of SNP-, gene-, and pathway-based approaches to identify genes influencing susceptibility to Staphylococcus aureus infections. Frontiers in Genetics 5, 125.

Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, Shen Y, Pervouchine DD, Djebali S, Thurman RE, Kaul R, Rynes E, Kirilusha A, Marinov GK, Williams BA, Trout D, Amrhein H, Fisher-Aylor K, Antoshechkin I, DeSalvo G, See LH, Fastuca M, Drenkow J, Zaleski C, Dobin A, Prieto P, Lagarde J, Bussotti G, Tanzer A, Denas O, Li K, Bender MA, Zhang M, Byron R, Groudine MT, McCleary D, Pham L, Ye Z, Kuan S, Edsall L, Wu YC, Rasmussen MD, Bansal MS, Kellis M, Keller CA, Morrissey CS, Mishra T, Jain D, Dogan N, Harris RS, Cayting P, Kawli T, Boyle AP, Euskirchen G, Kundaje A, Lin S, Lin Y, Jansen C, Malladi VS, Cline MS, Erickson DT, Kirkup VM, Learned K, Sloan CA, Rosenbloom KR, Lacerda de Sousa B, Beal K, Pignatelli M, Flicek P, Lian J, Kahveci T, Lee D, Kent WJ, Ramalho Santos M, Herrero J, Notredame C, Johnson A, Vong S, Lee K, Bates D, Neri F, Diegel M, Canfield T, Sabo PJ, Wilken MS, Reh TA, Giste E, Shafer A, Kutyavin T, Haugen E, Dunn D, Reynolds AP, Neph S, Humbert R, Hansen RS, De Bruijn M, Selleri L, Rudensky A, Josefowicz S, Samstein R, Eichler EE, Orkin SH, Levasseur D, Papayannopoulou T, Chang KH, Skoultchi A, Gosh S, Disteche C, Treuting P, Wang Y, Weiss MJ, Blobel GA, Cao X, Zhong S, Wang T, Good PJ, Lowdon RF, Adams LB, Zhou XQ, Pazin MJ, Feingold EA, Wold B, Taylor J, Mortazavi A, Weissman SM, Stamatoyannopoulos JA, Snyder MP, Guigo R, Gingeras TR, Gilbert DM, Hardison RC, Beer MA and Ren B, Mouse ENCODE Consortium 2014. A comparative encyclopedia of DNA elements in the mouse genome. Nature 515, 3553–3564.