Norwegian University of Life Sciences

Philosophiae Doctor Thesis 2010:50

Genomic architecture and complex traits in Norwegian Red cattle

Marte Sodeland

Norwegian University of Life Sciences
Department of Animal and Aquacultural Sciences
P.O. Box 5003
N-1432 Ås
Norway

ISSN: 1503-1667
ISBN: 978-82-575-0965-1

Abstract

The predominant cattle breed in Norway, Norwegian Red cattle (NRF), is an admixed breed formed from local Norwegian breeds and imported animals from other Nordic breeds. The extensive phenotypic records for NRF represent a unique resource for studying genetic factors affecting complex traits of importance for animal production. Phenotypic records of traits related to milk production, meat production and fertility, as well as health traits such as veterinary treated clinical mastitis, are now kept for 96% of Norwegian cattle. Records of veterinary treated clinical mastitis have been kept for most NRF animals for the last three decades.

Mastitis is inflammation of the mammary gland and is the most widespread disease affecting dairy cattle world-wide. Consequences of this disease include animal suffering, reduced milk quality, unwanted use of antibiotics and a more costly production. Main objective of the work described in this doctoral thesis has been to study the genomic architecture of the admixed NRF breed in order to understand the genetics underlying complex traits in cattle, particularly susceptibility to mastitis.

The study was initiated with the genotyping of 2,589 NRF bulls for single-nucleotide polymorphisms (SNPs) using the Bovine Affymetrix 25k MIP SNP array. Construction of linkage maps provided a powerful resource for quality assessment of the bovine genome assembly Btau_4.0 and for investigations of recombination rates and linkage disequilibrium (LD) across the NRF genome. Differences between recent and historic recombination rates were used to identify genomic loci subjected to strong artificial selection in the observed pedigree. Reduced LD was found in NRF compared with other breeds included in the study. The high LD generally reported in cattle facilitates association mapping studies for detection of quantitative trait loci (QTLs) affecting complex traits. In order to detect QTLs affecting susceptibility to mastitis genome-wide association studies with over 17,000 SNPs were performed for occurrence of clinical mastitis (CM) in seven lactational time periods and for lactation average somatic cell score (SCS). Although there is a genetic correlation between CM and SCS, no consistencies were found between SNPs significantly associated with CM and those associated with lactation average SCS. Combined linkage and linkage disequilibrium analysis confirmed quantitative trait loci for CM on bovine chromosomes 2, 6, 14 and 20, with the highest test score for CM being found for a SNP at 90.67Mb on chromosome 6. In addition to the QTL for CM on chromosome 6, a QTL affecting milk protein yield (PY) has been found to coincide with the casein genes around 88Mb on this chromosome. Fine mapping gave highest test scores for PY in and around the casein genes *CSN2* and *CSN1S2* (at 88.33Mb and 88.41Mb), while highest test scores for CM were found within the region 89 to 91Mb. It has been suggested that a haplotype encompassing the casein genes, with a favorable effect on PY and an unfavorable effect on CM, was introduced into the NRF population through importation of a Holstein-Fresian bull (1606 Frasse) in the 1970s. High-throughput re-sequencing allowed for molecular characterization of the long range haplotype from 1606 Frasse and revealed plausible causal polymorphisms in the promoter region of the gene *CSN1S2* and in a known regulatory motif in the 5'-flanking UTR of *CSN1S2*.

Sammendrag

Norsk Rødt fe (NRF) er en syntetisk rase basert på norske raser og importerte dyr fra andre nordiske raser. Fenotypiske egenskaper relatert til melkeproduksjon, kjøttproduksjon og fruktbarhet, i tillegg til helseegenskaper som veterinær behandlet klinisk mastitt, registreres nå for 96 prosent av norsk storfe. Dette materialet utgjør en unik ressurs for studier av komplekse egenskaper av økonomisk viktighet.

Tilfeller av veterinær behandlet klinisk mastitt er blitt registrert for de fleste NRF dyr i over 30 år. Mastitt er en betegnelse på inflammasjon i melkekjertelen og er den vanligste sykdommen i melkekyr på verdensbasis. Resistens mot mastitt i storfe forventes å være påvirket av både genetiske og miljømessige faktorer, og refereres til som en kompleks egenskap. Hovedmålsetningen med dette doktorgradsarbeidet har vært å studere genomets oppbygning og variasjon i NRF og å bruke denne informasjonen til å forstå mer av genetikken bak økonomisk viktige egenskaper i storfe, og da spesielt mastittresistens.

Studien ble innledet med genotyping av 2,589 NRF okser for enkelt-nukleotid polymorfismer (SNPer) fra den bovine Affymetrix 25k MIP SNP arrayen. Koblingskart konstruert i dette materialet ble brukt til kvalitets kontroll av det bovine genomassembliet (Btau_4.0) og for å studere rekombinasjonsrater og grad av koblingsulikevekt i NRF. Forskjeller i nylig og historisk rekombinasjonsrate ble brukt til å identifisere regioner i genomet som kan ha vært utsatt for sterk seleksjon. Redusert koblingsulikevekt ble funnet i NRF sammenlignet med andre storferaser inkludert i studien.

Den generelt høye koblingsulikevekten i storfe er nyttig for deteksjon av områder i genomet som påvirker komplekse egenskaper (QTLer). For å detektere QTLer for mastittresistens ble det gjennomført en helgenom assosiasjonsstudie med over 17,000 SNPer. Registreringer på mastitt ble delt inn i syv tidsperioder i laktasjonen. I tillegg var celletall i melk inkludert som et indirekte mål på sykdommen. Selv om det er genetisk korrelasjon mellom klinisk mastitt og celletall i melk avdekket ikke dette studiet SNPer som var signifikant assosiert med begge egenskapene. QTLer for klinisk mastitt ble identifisert på kromosom 2, 6, 14 og 20, og høyeste testverdi ble funnet for en SNP ved 90.67Mb på kromosom 6.

I tillegg til QTLen for klinisk mastitt rundt 90Mb på kromosom 6 ble det også funnet en QTL for protein mengde i melk ved kasein genene omkring 88Mb. En finkartlegging i området pekte ut kasein genene *CSN2* og *CSN1S2* (ved 88.33Mb og 88.41Mb) som mest sannsynlig QTL område for proteinmengde, mens sterkest assosiasjon med klinisk mastitt ble funnet for SNPer i regionen 89 til 91Mb. Tidligere studier har foreslått at en haplotype som dekker kasein genene, med en positiv effekt på protein mengde i melk og en negativ effekt på mastitt resistens, ble introdusert i NRF populasjonen ved import av en Holstein-Friesian okse (1606 Frasse) i 1970 årene. Storskala resekvensering tillot molekylær karakterisering av haplotypen fra 1606 Frasse, og sannsynlige kausale polymorfismer ble detektert i promotor regionen og i den 5'- flankerende utranslaterte regionen av genet *CSN1S2*.

Acknowledgements

The presented work was carried out at the Department of Animal and Aquacultural Sciences at the Norwegian University of Life Sciences. The work was funded by The Research Council of Norway, GENO Breeding and AI association and BoviBank Ltd.

I would like to thank my main supervisor Sigbjørn Lien and my co-supervisor Matthew P. Kent for their guidance and support, and my co-authors and colleagues for their contributions and collaborations. I would also like to thank my family and friends for their love and patience.

Ås, November 2010

Marte Sodeland

Table of Contents

List of Papers

I Sodeland, M., Kent, M.P., Hayes, B., Grove H. and Lien, S. (2010) Recent and historical recombination in the admixed Norwegian Red cattle breed. *BMC Genomics* (Submitted)

II Sodeland, M., Kent, M.P., Olsen, H.G., Opsal, M.A., Svendsen, M., Sehested, E., Hayes, B. and Lien, S. (2010) Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. *Anim. Genet*. (Accepted)

III Sodeland, M., Grove, H., Kent, M. P, Taylor, S., Svendsen, M., Hayes, B. and Lien, S. Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle. (Manuscript)

Introduction

1. Motivation and main objective

The Norwegian dairy herd recording system include records on phenotypic traits related to milk production, meat production, fertility and health, and records of veterinary treated clinical mastitis have been kept for most Norwegian Red cattle (NRF) animals for the last three decades.

Mastitis is inflammation of the mammary gland and is the costliest and most widespread disease affecting dairy cattle world-wide. Consequences of this disease include animal suffering, reduced milk quality and unwanted use of antibiotics. Susceptibility to bovine mastitis is expected to be affected by a number of genetic factors in addition to environmental factors, and is therefore often referred to as a complex trait.

Association mapping studies enables detection of genomic loci affecting complex traits such as susceptibility to mastitis, and for NRF association mapping for mastitis susceptibility is facilitated by a large number of phenotypic and pedigree records resulting from the extensive national recording system. Recent sequencing of the bovine genome and large-scale detection of genetic variation in cattle have also provided valuable resources for genome research. Main objective of the work described in this doctoral thesis has been to study the genomic architecture of the admixed NRF breed in order to understand the genetics underlying complex traits in cattle, particularly susceptibility to mastitis.

2. Description of population and trait

2.1. Norwegian Red cattle

NRF is an admixed breed formed from local Norwegian breeds and imported animals from other Nordic breeds. There is still some gene flow between NRF and other Scandinavian breeds to ensure diversity and sustainability. NRF is the predominant cattle breed in Norway and main selection goals for NRF include traits related to milk production, meat production, health and fertility [1]. Mastitis resistance has been included in the breeding goal for NRF since 1980, and has for the last twenty years constituted approximately 20% of the breeding goal net merit index. Records on veterinary treated clinical mastitis (CM) have been kept in the Norwegian Cattle Health Recording System for most NRF animals since 1975, yielding an extensive phenotype material for studying genetics of bovine mastitis susceptibility [2].

2.2. Bovine mastitis

Mastitis is inflammation of the mammary gland and symptoms of CM can include changes in milk composition, redness or swelling, pain, fever and loss of appetite. The most common cause of mastitis is bacterial infection, and in Norway the most frequently identified pathogen in inflamed udders is *Staphylococcus aureus* [3]. Susceptibility to mastitis is determined by the ability to avoid or to rapidly recover from disease and this ability is affected by environmental factors such as climate, stress level, hygiene and diet, and depends on both anatomical and molecular defense mechanisms [4-7]. Anatomical components of mammary gland defense include the teat skin and the teat canal. The teat canal is kept close by tight muscle contraction and accumulation of waxy keratin, and thus functions as a barrier against bacterial invasion [7]. If the teat skin is damaged, bacteria can invade the teat after establishing an infection. Bacteria that are able to invade the teat and the mammary gland will be challenged by cellular and soluble defense mechanisms. Important cell types in this defense are neutrophils, macrophages, natural killer cells and lymphocytes, while soluble factors in mammary gland defense include cytokines, complement components and antibodies [8-10]. The constituents of the molecular mammary gland defense work together in integrated

pathways where recognition of invading pathogens and recruitment of cellular and soluble components to the site of infection are crucial steps.
Genetic disposition for mastitis susceptibility could be related to the immune response or other biological influences on disease resistance. For immunological defense components such as neutrophils, antibodies and complement components genetic variability in concentration and functionality have been found [11, 12]. Genetic factors affecting susceptibility to mastitis could also include indirect effects on molecular defense by genetic variability in stress response or increased energy demand at the onset of milk production [9, 13].
Heritability estimates for CM are generally low, ranging between 0.02 and 0.12 for Nordic cattle populations [14-16], and susceptibility to this disease is expected to be affected by a number of genomic loci. Previous studies have reported quantitative trait loci (QTLs) affecting susceptibility to mastitis on several of *Bos Taurus* chromosomes (BTAs); on BTAs 3, 4, 6, 14 and 27 in NRF [17], on BTA9 in a study including three Nordic cattle breeds [18], on BTAs 14 and 18 in Finnish Ayrshire [19] and on BTAs 5, 6, 9, 11, 15 and 26 in Danish Holstein [20].

2.3. Somatic cells in milk

Somatic cells in milk include a number of cell types such as neutrophils, macrophages, lymphocytes and epithelial cells. Macrophages are the dominant cell type in the milk of a healthy udder and upon detection of bacterial components these and other cells release pro-inflammatory mediators [10, 21]. Neutrophil migration from the blood to the mammary gland follows as an important step in early disease resistance, after which neutrophils kill bacteria by phagocytosis and act as a source of antibacterial peptides and pro-inflammatory mediators [7, 21]. In a diseased udder neutrophils can make up over 90% of the somatic cells [14, 22].
A wide range of genetic correlation estimates between CM and somatic cell score (SCS) in cattle populations have been reported, with an average of about 0.7 [23]. A number of studies internationally have used high SCS in milk as an indication of mastitis for QTL mapping [24-26]. SCS varies through lactation and higher correlations have been reported between SCS and occurrence of CM in late lactation than between SCS and occurrence of CM in early lactation [27]. In Norway milk SCS is recorded as lactation means, with samples taken every second month. Such a sampling scheme for SCS will only detect a small fraction of infections and does not necessarily provide a suitable measure of occurrence of mastitis for QTL mapping [28].

2.4. Mastitis in different stages of lactation

Mammary gland defense mechanisms are altered in the periparturient period and animals are more susceptible to infection in this early stage of lactation [5-9]. Production of stress hormones is stimulated in this period, which is believed to impair neutrophil migration, decrease level of lymphocytes in blood, decrease antibody level in mammary secretions and inhibit cytokine production [13]. Number of mature neutrophils in blood and milk is at the lowest around parturition [21].
Heringstad *et al.* [29] found differences in heritability for CM in different stages of lactation in NRF using a threshold model. Their estimates ranged from a heritability of 0.09 in the interval
-30 to 0 days after calving to a heritability of 0.05 in the interval 121 to 300 days after calving. In the same study genetic correlations between susceptibility to CM in different stages of lactation were investigated, and between early and late stages low genetic correlations were found.
Low genetic correlations between mastitis susceptibility in early and late lactation suggest that the genetic factors affecting susceptibility to mastitis change though the lactational stages.

2.5. Milk production and mastitis susceptibility

In NRF there is an undesirable genetic correlation between milk protein yield (PY) and CM, i.e. increased PY coincides with increased occurrence of CM [30]. This relationship could be explained both by existence of pleiotropic effects and by QTLs affecting each trait being closely positioned on bovine chromosomes. Pleiotropic effects could result from competition for energy and nutrients or increased oxidant stress level resulting from elevated milk production or from bacterial infection [6]. Further, high milk production could reduce concentration of molecular defense components in the udder and thereby reduce the ability to avoid or recover from disease. High milk production and milk leakage could also prevent teat closure and thereby the teats ability to act as a barrier against bacterial invasion [21, 31]. Although improvements in both milk production and mastitis resistance have been achieved in NRF over the last two decades it is challenging to improve both traits simultaneously due to the genetic correlation.

3. Methods of assessment

3.1. Bovine genome assembly

The first preliminary assembly of the bovine genome (Btau_1.0) was completed in September 2004 and had a 3x sequence coverage [32, 33]. The third (Btau_3.1) and fourth (Btau_4.0) releases published in 2006 and 2007 had 7.1x coverage. In 2009 high quality finished sequence was incorporated into the draft assembly to result in the fifth release (Btau_4.2). Over 95% of the genomic sequence is expected to be represented in Btau_4.0 and Btau_4.2 [33].

The bovine genome reference sequence is a very valuable resource within bovine genomics and facilitates positioning and annotation of candidate genes, comparative genomics, gene expression profiling and large scale detection of genetic variation. A number of genome sequencing projects have been initiated for other livestock species, including that of horse, chicken, salmon and pig [34-37], and are likely to advance the understanding of genetic factors affecting complex traits important for animal production.

3.2. Single-nucleotide polymorphisms

Genomic loci known to be variant or polymorphic can be used as genetic markers, and the genotype of an individual for a genetic marker is determined by which alleles the individual holds for that loci. Most common genetic markers used today are single-nucleotide polymorphisms (SNPs), which are single base variations in the DNA sequence [38, 39]. SNPs are normally bi-allelic, meaning they each have two alleles or variants for the nucleotide position present in populations where they are polymorphic. A SNP is considered polymorphic in a population if its minor-allele frequency is greater than 1%.

The emergence of high-throughput sequencing technology together with the availability of the bovine genome sequence has provided a powerful approach for SNP detection in cattle by genome re-sequencing [33, 40-42] and for the last few years the number of detected SNPs in cattle has increased dramatically [43-45]. Development of large-scale SNP arrays enables time efficient determination of genotypes, which has contributed to an increase in the application of genetic markers [46, 47].

3.3. Recombination rate and linkage disequilibrium

In diploid species such as mammals each individual has two homologous versions of each autosomal chromosome, one from each of its parents. During meiosis recombination between homologous chromosomes may take place in one or both parents before one chromosome from each parent is transferred to an offspring [48]. Two genomic loci closely positioned on a

chromosome are less likely to have recombination between them than two loci positioned further apart on the same chromosome. Thus, alleles of two genetic markers that are closely positioned on a chromosome are more likely to be transferred together from parent to offspring. Genetic markers are said to be in linkage if the probability of recombination between them during meiosis is less than 0.5, and genetic markers in linkage are part of the same linkage group. Genetic maps for linkage groups can be constructed by calculating genetic distance between adjacent markers from the number of meiotic recombinations occurring between them in an observed pedigree [49-51].

In recent years it has become clear that meiotic recombination tend to occur in regions labeled recombination hotspots [52]. McVean *et al.* [53] found that in humans 50% of recombinations take place in such hotspots, which constitute less than 10% of the genomic sequence. Great variation in density and intensity of recombination hotspots across the human genome has been observed [54].

The term linkage disequilibrium (LD) is used to describe degree of allelic association between genomic loci, which is disrupted by meiotic recombination. Loci with high recombination rate will have reduced LD whilst loci with low recombination rate will have elevated LD. Elevated LD between two loci is generally due to close linkage, but could also be due to selection, genetic drift, gene flow, population substructure, recent admixture or decreased effective population size. Reduced LD could be due to population expansion or gene conversion [55, 56]. Cattle in general have extended LD compared with humans, believed to be caused by low effective population size and strong artificial selection [57-62].

Population recombination rate (ρ), which is inversely related to LD, can be described as a multi-locus LD measure [55, 63-65]. Historic population recombination rate is an estimate of the rate at which chromosomal recombination have occurred in a genomic interval in the history of a population, and recent developments allow incorporation of models accounting for recombination hotspots in the estimation of this parameter [66]. Comparison of historic recombination patterns with recombination patterns in recent generations, obtained from a genetic map, can reveal loci for which there has been an alteration in selection pressure.

3.4. Haplotyping and imputation

Diploid individuals have two alleles for each genetic marker in their genome, and those two alleles make up the individuals genotype for the marker. If the phase for a set of genetic markers in the same linkage group is known for an individual, the two haplotypes for the linkage group for that individual are also known. Here the term haplotype is used in reference to the set of alleles for a chromosome or chromosomal region that was transferred to an individual from one of its parents. A haplotype block can be defined as a set of markers showing strong LD and being closely positioned in the genome. A haplotype block normally contain a limited number of haplotypes, each characterized by its set of co-occurring alleles, and can be a signature of positive selection [56, 67, 68].

Pedigree information is of great advantage for construction of haplotypes, and there have been substantial developments in haplotyping strategies for large datasets containing related individuals of known complex pedigrees [69-73]. Haplotypes are valuable for a number of applications including determination of genetic relationship between individuals, mapping of genomic loci affecting phenotypic traits, detecting signatures of selection and imputation of untyped genotypes.

Imputation of untyped genotypes builds upon the assumption that only a limited number of haplotypes are present in a population for closely linked loci. For untyped markers information from surrounding markers are used to identify which haplotypes an individual holds for the loci by comparisons with other haplotypes in the population. If haplotypes for the individual can be identified untyped markers can be imputed from genotype information

from individuals holding the same haplotypes.
Imputation provides a very cost effective approach to obtain large genotype datasets and if haplotypes from closely related individuals are available the accuracy of imputation will be greatly improved [73-76]. This approach is particularly attractive in livestock species due to the extensively recorded pedigree information. In cattle genotyping costs may be reduced substantially by combining whole genome sequence data from elite sires with SNP array information from a larger proportion of the population, followed by imputation of untyped genotypes.

3.5. Association mapping in livestock

Association mapping, or LD mapping, is a method for mapping QTLs that take advantage of LD to find association between phenotypic observations and genetic markers. The power of association mapping studies to detect QTLs are dependent upon the density and distribution of genetic markers, the number of individuals being genotyped, the quality of phenotypic records and the genomic architecture of the population [77, 78]. Association mapping in livestock is becoming increasingly popular with the development of large-scale SNP arrays [43, 44, 78], and is often empowered by a high number of available phenotypic records [79-81].
Association mapping depends on LD between genetic markers and loci affecting a phenotypic trait and is in livestock populations empowered by extensive LD [57-61]. An assumption for association mapping is that LD between genetic markers and a polymorphism affecting a phenotypic trait will be due to close linkage. Other causes of LD than linkage can however arise, and a major source of false positives in association mapping is LD due to population structure [55, 82-85]. For livestock populations false positives due to population structure can often be reduced by incorporation of pedigree information [86].
Both single genetic markers and haplotypes can be used for association mapping. An argument for using haplotypes is that bi-allelic markers such as SNPs do not accurately represent parental chromosomes, yielding limited power to describe the genetic relationship structure between individuals for the assessed genomic loci [87]. By using haplotypes more information is gained on the genetic relationship between individuals, but using erroneous haplotypes could introduce errors into the analyses. For livestock populations extensive pedigree records can improve the accuracy of haplotyping and thereby the power of haplotype association mapping [73-75].

4. Summary of papers

4.1. Paper I

The work described in Paper I was initiated by genotyping of 2,589 NRF sires from paternal half-sib families for SNPs from the Affymetrix 25K MIP array [44], followed by construction of a dense genetic map containing over 17,000 SNPs [33]. To detect regions subjected to strong artificial selection in the observed pedigree estimates of historical scaled recombination rate ($\rho(h)$) from LD were compared with recent scaled recombination rate ($\rho(r)$) from the genetic map. A reduced $\rho(r)$ relative to $\rho(h)$ for a genomic region could be an indication of artificial selection. Regions where $\rho(r)$ were most strikingly reduced relative to $\rho(h)$ were for the middle of BTA1 and the middle of BTA20. On BTA1 several QTLs affecting milk production traits have been reported [19, 88-92], and a meta-analysis reported by Khatkar *et al*. [93] indicated presence of three QTLs for milk yield on this chromosome. The BTA20 region centres around a mutation reported to affect protein percentage in the *GHR* gene [94] and Hayes *et al*. [95] reported evidence for strong selection in this region in a study of divergence between dairy cattle and beef cattle.
For comparison of the admixed NRF to other breeds genotypes were retrieved for Holstein,

Finnish Ayrshire, Sided Troender and Nordland Cattle and Icelandic cattle sires. A principal component analysis of the genomic relationship matrix among individuals of different and the same breed was conducted to evaluate genetic distances between breeds [82]. Finnish Ayrshire and NRF animals grouped together but some NRF bulls had high levels of relationship with Holsteins. The analysis also showed increased heterogeneity among NRF animals relative to other breeds.

Genome-wide distributions of LD, measured by r^2, versus inter-marker distance were found from syntenic SNP pairs for all genotyped breeds. Reduced LD was observed in NRF compared to the other breeds, likely reflecting elevated heterogeneity in NRF from historic admixture in combination with recent attempts to maintain a large effective population size by control of inbreeding and gene flow through import of sires from other Nordic countries [1]. For NRF a mean r^2 of 0.5 or more was observed for SNPs positioned less than 10kb apart while a mean r^2 of 0.3 or more was observed for SNPs positioned less than 30kb apart. A report of decline in r^2 with increasing distance between SNPs in Australian Holstein-Friesian cattle [61] showed quite similar results. Reports from other breeds have described similar or more rapid decline in r^2 at short distances than found in NRF [96, 97]. Cattle in general have extended LD compared with humans, believed to be caused by low effective population size and strong artificial selection [57-62].

Finally, estimates of scaled population recombination rate for each interval between adjacent SNPs were used to identify problematic regions in the bovine genome assembly (Btau_4.0) [33]. Positions for 130 previously un-positioned contigs, identified by comparative sequence analysis, were validated by linkage analysis. Of these positions 27% corresponded to extreme values of population recombination rate. An alternative bovine genome assembly (UMD2) was reported by Zimin *et al*. [98], and some of the problematic regions identified in the study described here corresponded to regions identified by sequence alignments to be differing between the UMD2 and the Btau_4.0 bovine genome assemblies.

4.2. Paper II

Association mapping in cattle is facilitated by extensive LD [57-61], which increases power to detect genomic loci affecting phenotypic traits. In Paper II results from genome-wide association studies for CM and lactation average SCS based on a genetic map containing over 17,000 SNPs were presented. Records on veterinary treated CM have been kept in the Norwegian Cattle Health Recording System for the last thirty years and provide a valuable resource for association mapping for this trait in NRF [2]. Genotypes from a total of 2,589 sires with almost 1.4 million daughter records on CM were included in the analysis, and records on occurrence of CM were divided into seven time periods in the three first lactations in order to identify QTLs affecting mastitis susceptibility in particular phases of lactation. None of the QTLs for CM detected in this study were associated with lactation average SCS. Combined linkage disequilibrium and linkage analysis was used to follow up and validate the most convincing results from association mapping for CM, and QTLs were identified for CM in the periparturient period on BTAs 2, 6 and 20 and for CM in late lactation on BTA14. A multiple QTL analysis indicated that none of these QTL regions contained more than one QTL for CM.

Highest test score for CM in the periparturient period from the genome-wide association study was found for a SNP at 90.67Mb on BTA6. This SNP was located near a cluster of genes coding for interleukin 8 and other C-X-C motif chemokines. C-X-C motif chemokines are pro-inflammatory mediators and important constituents in the defence against invading bacteria. Interestingly, SNPs on BTA2 highly associated with CM were located near two genes coding for receptors that have C-X-C motif chemokines as ligands.

4.3. Paper III

A QTL for CM has been detected around 90Mb on BTA6, close to a QTL for PY that coincide with the casein gene cluster around 88Mb [99-103]. Casein proteins constitute the majority of proteins found in bovine milk and have been shown to contain variation associated with milk protein content and protein composition [100-110].

A haplotype covering these two QTLs, with a favorable effect on protein content and an unfavorable effect on mastitis susceptibility, might explain a part of the genetic correlation between PY and CM observed in NRF [30]. This haplotype was introduced into the NRF population through import of a Holstein-Fresian bull (1606 Frasse) in the 1970s [108], and selection for milk production traits has likely increased the frequency of this haplotype in NRF.

In Paper III results from fine mapping in the genomic interval from 86 to 97Mb on BTA6 for CM in the periparturient period of first, second and third lactation, as well as for PY, were presented. Highest test scores for PY were found in and around the casein genes *CSN2* and *CSN1S2* positioned at 88.33Mb and 88.41Mb, while highest test scores for CM were found in the region 89 to 91Mb. The data indicated that the bull 1606 Frasse was homozygote for both these QTLs, and the long range haplotype from this sire was associated with increased PY and increased CM. High-throughput re-sequencing of 1606 Frasse and two of his sons allowed for molecular characterization of this long range haplotype.

A SNP was detected in the promoter region of *CSN1S2*, at -7bp relative to the transcription initiation site. This polymorphism was positioned only three base-pairs downstream of a CCAAT motif, which is a binding site for transcription factors known to regulate the expression of casein genes. A SNP was also detected in a known regulatory motif in the 5'-flanking un-translated region of *CSN1S2*, only -5 bp from the translation initiation codon. It has previously been suggested that variation in this motif might be the cause of the observed variation in translational efficiency between casein genes [111]. Work has been initiated to deduce the effects of the detected polymorphisms on transcription and translation of *CSN1S2*. In order to identify candidate genes and possible causal polymorphisms affecting mastitis susceptibility in this region fine mapping with higher SNP density will be necessary.

5. Concluding remarks and future perspectives

The emergence of whole genome sequences for a number of livestock species is likely to revolutionize the way research is conducted for assessing genetic factors affecting complex traits. Availability of the bovine genomic sequence and advances in high-throughput sequencing have in the past few years led to a dramatic increase in the number of validated SNPs and, with the availability of large-scale SNP arrays, SNPs are increasingly used within genomic research.

In particular, there have been many reports on use of SNP data to detect genomic loci subjected to positive selection, which for livestock species often would co-occur with loci affecting important production traits. One approach to identify loci affected by positive selection is by comparison of recent patterns of recombination to historical patterns of recombination. Shifts in the pattern of recombination could be an indication of alterations in selection pressure, and in this study reduced recent recombination in the bovine genome was found to coincide with some well known QTL regions for milk production traits. With increasing SNP density the power to identify genomic loci subjected to positive selection and the popularity of approaches to detect such loci are expected to increase.

Genome-wide association studies to map QTLs are also becoming increasingly popular with the availability of large-scale SNP arrays. Key elements for identifying and dissecting genetic

factors affecting complex traits are access to comprehensive and reliable phenotypic records, extensive pedigree information and availability of biological samples. Substantial work and investments have been made over the last 30-40 years in organising, developing and maintaining such resources for cattle in Norway. These resources, together with high-throughput genomic technologies, represent a unique opportunity for genetic characterization of economically important traits in cattle. Here genome-wide association studies for CM revealed QTLs on BTA2, BTA6, BTA14 and BTA20. The QTL on BTA6 was located close to a QTL for PY, and fine mapping for both PY and CM was performed for this genomic region. Detection of novel genetic markers for fine mapping was empowered by the availability of the bovine genome assembly (Btau_4.0) and recent advances in high-throughput sequencing technology. Haplotyping provided an efficient way to join genotype datasets through imputation of untyped genotypes, which reduces costs and improves the power of association mapping. Accuracy of haplotyping and imputation in livestock populations is often high due to elevated LD and availability of extensive pedigree records. Moreover, high-throughput genome re-sequencing allowed for complete molecular characterization of long rang haplotypes encompassing the two QTLs for PY and CM on BTA6. With the constant decline in costs for whole genome sequencing the potential that lies in re-sequencing of an increasing number of individuals emerges, as the combination of high-throughput re-sequencing and imputation methods allows for complete characterization of common genetic variation in livestock populations. Such approaches provide the resources necessary to uncover causal polymorphisms affecting complex traits important for animal production and are expected to gain in popularity over the next few years.

List of abbreviations

BTA	*Bos Taurus* chromosome
CM	Clinical mastitis
LD	Linkage disequilibrium
NRF	Norwegian Red cattle
PY	Protein yield
QTL	Quantitative trait locus
SCS	Somatic cell score
SNP	Single-nucleotide polymorphism

References

1. **GENO** [www.geno.no]
2. Østeras O, Solbu H, Refsdal AO, Roalkvam T, Filseth O, Minsaas A: **Results and evaluation of thirty years of health recordings in the Norwegian dairy cattle population**. *J Dairy Sci* 2007, **90**(9):4483-4497.
3. Østeras O, Kruse, H, Sølverød, L, Gjestvang, J, Mørk, T: **Nordic View Concerning Mastitis Pathogen Resistance**. *Proceedings NMC 45th Annual Meeting Tampa, Florida* 2006.
4. Harmon RJ: **Physiology of mastitis and factors affecting somatic cell counts**. *J Dairy Sci* 1994, **77**(7):2103-2112.
5. Detilleux JC: **Genetic factors affecting susceptibility of dairy cows to udder pathogens**. *Vet Immunol Immunopathol* 2002, **88**(3-4):103-110.
6. Rupp R, Boichard D: **Genetics of resistance to mastitis in dairy cattle**. *Vet Res* 2003, **34**(5):671-688.
7. Sordillo LM: **Factors affecting mammary gland immunity and mastitis susceptibility**. *Livestock Production Science* 2005, **98**:89-99.
8. Sordillo LM, Streicher KL: **Mammary gland immunity and mastitis susceptibility**. *J Mammary Gland Biol Neoplasia* 2002, **7**(2):135-146.
9. Waller KP: **Mammary gland immunology around parturition. Influence of stress, nutrition and genetics**. *Adv Exp Med Biol* 2000, **480**:231-245.
10. Rainard P, Riollet C: **Innate immunity of the bovine mammary gland**. *Vet Res* 2006, **37**(3):369-400.
11. Detilleux JC, Kehrli ME, Jr., Stabel JR, Freeman AE, Kelley DH: **Study of immunological dysfunction in periparturient Holstein cattle selected for high and average milk production**. *Vet Immunol Immunopathol* 1995, **44**(3-4):251-267.
12. Kelm SC, Detilleux JC, Freeman AE, Kehrli ME, Jr., Dietz AB, Fox LK, Butler JE, Kasckovics I, Kelley DH: **Genetic association between parameters of inmate immunity and measures of mastitis in periparturient Holstein cattle**. *J Dairy Sci* 1997, **80**(8):1767-1775.
13. Burton JL, Madsen SA, Chang LC, Weber PS, Buckham KR, van Dorp R, Hickey MC, Earley B: **Gene expression signatures in neutrophils exposed to glucocorticoids: a new paradigm to help explain "neutrophil dysfunction" in parturient dairy cows**. *Vet Immunol Immunopathol* 2005, **105**(3-4):197-219.
14. Heringstad B, Klemetsdal G, Ruane J: **Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries** *Livestock Production Science* 2000, **64**(2-3):95-106.
15. Heringstad B, Chang YM, Gianola D, Klemetsdal G: **Genetic analysis of clinical mastitis, milk fever, ketosis, and retained placenta in three lactations of Norwegian red cows**. *J Dairy Sci* 2005, **88**(9):3273-3281.
16. Lund MS, Jensen J, Petersen PH: **Estimation of genetic and phenotypic parameters for clinical mastitis, somatic cell production deviance, and protein yield in dairy cattle using Gibbs sampling**. *J Dairy Sci* 1999, **82**(5):1045-1051.
17. Klungland H, Sabry A, Heringstad B, Olsen HG, Gomez-Raya L, Vage DI, Olsaker I, Odegard J, Klemetsdal G, Schulman N *et al*: **Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle**. *Mamm Genome* 2001, **12**(11):837-842.
18. Sahana G, Lund MS, Andersson-Eklund L, Hastings N, Fernandez A, Iso-Touru T, Thomsen B, Viitala S, Sorensen P, Williams JL *et al*: **Fine-mapping QTL for mastitis resistance on BTA9 in three Nordic red cattle breeds**. *Anim Genet* 2008, **39**(4):354-362.
19. Schulman NF, Viitala SM, de Koning DJ, Virta J, Maki-Tanila A, Vilkki JH: **Quantitative trait Loci for health traits in Finnish Ayrshire cattle**. *J Dairy Sci* 2004, **87**(2):443-449.
20. Lund MS, Guldbrandtsen B, Buitenhuis AJ, Thomsen B, Bendixen C: **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* 2008, **91**(10):4028-4036.

21. Pyorala S: **Mastitis in post-partum dairy cows**. *Reprod Domest Anim* 2008, **43 Suppl 2**:252-259.
22. Kehrli ME, Jr., Shuster DE: **Factors affecting milk somatic cells and their role in health of the bovine mammary gland**. *J Dairy Sci* 1994, **77**(2):619-627.
23. Mrode RA, Swanson GJT: **Genetic and statistical properties of somatic cell count and its suitability as an indirect means of reducing the incidence of mastitis in dairy cattle.** *Animal Breeding Abstracts* 1996, **64**:847–857.
24. Kuhn C, Bennewitz J, Reinsch N, Xu N, Thomsen H, Looft C, Brockmann GA, Schwerin M, Weimann C, Hiendleder S *et al*: **Quantitative trait loci mapping of functional traits in the German Holstein cattle population**. *J Dairy Sci* 2003, **86**(1):360-368.
25. Bennewitz J, Reinsch N, Grohs C, Leveziel H, Malafosse A, Thomsen H, Xu N, Looft C, Kuhn C, Brockmann GA *et al*: **Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle**. *Genet Sel Evol* 2003, **35**(3):319-338.
26. Leyva-Baca I, Schenkel F, Sharma BS, Jansen GB, Karrow NA: **Identification of single nucleotide polymorphisms in the bovine CCL2, IL8, CCR2 and IL8RA genes and their association with health and production in Canadian Holsteins**. *Anim Genet* 2007, **38**(3):198-202.
27. Svendsen M, Heringstad B: **Somatic Cell Count as an Indicator of Subclinical Mastitis. Genetic Parameters and Correlations with Clinical Mastitis.** *Interbull Bulletine* 2006, **35**:12-16.
28. Shook GE, Schutz MM: **Selection on somatic cell score to improve resistance to mastitis in the United States**. *J Dairy Sci* 1994, **77**(2):648-658.
29. Heringstad B, Chang YM, Gianola D, Klemetsdal G: **Multivariate threshold model analysis of clinical mastitis in multiparous norwegian dairy cattle**. *J Dairy Sci* 2004, **87**(9):3038-3046.
30. Heringstad B, Chang YM, Gianola D, Klemetsdal G: **Genetic association between susceptibility to clinical mastitis and protein yield in norwegian dairy cattle**. *J Dairy Sci* 2005, **88**(4):1509-1514.
31. Dingwell RT, Leslie KE, Schukken YH, Sargeant JM, Timms LL, Duffield TF, Keefe GP, Kelton DF, Lissemore KD, Conklin J: **Association of cow and quarter-level factors at drying-off with new intramammary infections during the dry period**. *Prev Vet Med* 2004, **63**(1-2):75-89.
32. Womack JE: **The bovine genome**. *Genome Dyn* 2006, **2**:69-78.
33. Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y *et al*: **Bos taurus genome assembly**. *BMC Genomics* 2009, **10**:180.
34. Ng SH, Artieri CG, Bosdet IE, Chiu R, Danzmann RG, Davidson WS, Ferguson MM, Fjell CD, Hoyheim B, Jones SJ *et al*: **A physical map of the genome of Atlantic salmon, Salmo salar**. *Genomics* 2005, **86**(4):396-404.
35. Wallis JW, Aerts J, Groenen MA, Crooijmans RP, Layman D, Graves TA, Scheer DE, Kremitzki C, Fedele MJ, Mudd NK *et al*: **A physical map of the chicken genome**. *Nature* 2004, **432**(7018):761-764.
36. Wernersson R, Schierup MH, Jorgensen FG, Gorodkin J, Panitz F, Staerfeldt HH, Christensen OF, Mailund T, Hornshoj H, Klein A *et al*: **Pigs in sequence space: a 0.66X coverage pig genome survey based on shotgun sequencing**. *BMC Genomics* 2005, **6**(1):70.
37. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR *et al*: **Genome sequence, comparative analysis, and population genetics of the domestic horse**. *Science* 2009, **326**(5954):865-867.
38. Vignal A, Milan D, SanCristobal M, Eggen A: **A review on SNP and other types of molecular markers and their use in animal genetics**. *Genet Sel Evol* 2002, **34**(3):275-305.
39. Weaver TA: **High-throughput SNP discovery and typing for genome-wide genetic analysis**. *Trends in Genetics* 2000, **December 2000**:36-42.
40. Stratton M: **Genome resequencing and genetic variation**. *Nat Biotechnol* 2008, **26**(1):65-66.

41. Mardis ER: **The impact of next-generation sequencing technology on genetics**. *Trends Genet* 2008, **24**(3):133-141.
42. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S *et al*: **Evaluation of next generation sequencing platforms for population targeted sequencing studies**. *Genome Biol* 2009, **10**(3):R32.
43. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S *et al*: **Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds**. *Science* 2009, **324**(5926):528-532.
44. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS *et al*: **Development and characterization of a high density SNP genotyping assay for cattle**. *PLoS One* 2009, **4**(4):e5350.
45. Eck SH, Benet-Pages A, Flisikowski K, Meitinger T, Fries R, Strom TM: **Whole genome sequencing of a single Bos taurus animal for single nucleotide polymorphism discovery**. *Genome Biol* 2009, **10**(8):R82.
46. Lee JE: **High-throughput genotyping**. *Forum Nutr* 2007, **60**:97-101.
47. Ragoussis J: **Genotyping technologies for genetic research**. *Annu Rev Genomics Hum Genet* 2009, **10**:117-133.
48. Klug WW, Cummings MR, Spencer CA: **Concepts of Genetics**, 8th edition edn: Pearson Education International; 2006.
49. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G *et al*: **A high-resolution recombination map of the human genome**. *Nat Genet* 2002, **31**(3):241-247.
50. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: **Comprehensive human genetic maps: individual and sex-specific variation in recombination**. *Am J Hum Genet* 1998, **63**(3):861-869.
51. Lander ES, Green P: **Construction of multilocus genetic linkage maps in humans**. *Proc Natl Acad Sci U S A* 1987, **84**(8):2363-2367.
52. Jeffreys AJ, Holloway JK, Kauppi L, May CA, Neumann R, Slingsby MT, Webb AJ: **Meiotic recombination hot spots and human DNA diversity**. *Philos Trans R Soc Lond B Biol Sci* 2004, **359**(1441):141-152.
53. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P: **The fine-scale structure of recombination rate variation in the human genome**. *Science* 2004, **304**(5670):581-584.
54. Myers S, Spencer CC, Auton A, Bottolo L, Freeman C, Donnelly P, McVean G: **The distribution and causes of meiotic recombination in the human genome**. *Biochem Soc Trans* 2006, **34**(Pt 4):526-530.
55. Ardlie KG, Kruglyak L, Seielstad M: **Patterns of linkage disequilibrium in the human genome**. *Nat Rev Genet* 2002, **3**(4):299-309.
56. Abecasis GR, Ghosh D, Nichols TE: **Linkage disequilibrium: ancient history drives the new genetics**. *Hum Hered* 2005, **59**(2):118-124.
57. Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M *et al*: **Extensive genome-wide linkage disequilibrium in cattle**. *Genome Res* 2000, **10**(2):220-227.
58. Vallejo RL, Li YL, Rogers GW, Ashwell MS: **Genetic diversity and background linkage disequilibrium in the North American Holstein cattle population**. *J Dairy Sci* 2003, **86**(12):4137-4147.
59. Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM: **Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes**. *J Anim Sci* 2003, **81**(3):617-623.
60. Odani M, Narita A, Watanabe T, Yokouchi K, Sugimoto Y, Fujita T, Oguni T, Matsumoto M, Sasaki Y: **Genome-wide linkage disequilibrium in two Japanese beef cattle breeds**. *Anim Genet* 2006, **37**(2):139-144.
61. Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Cavanagh JA, Barris W, Schnabel RD, Taylor JF, Raadsma HW: **Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel**. *BMC Genomics* 2008, **9**:187.

62. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME: **Novel multilocus measure of linkage disequilibrium to estimate past effective population size**. *Genome Res* 2003, **13**(4):635-643.
63. Pritchard JK, Przeworski M: **Linkage disequilibrium in humans: models and data**. *Am J Hum Genet* 2001, **69**(1):1-14.
64. Mueller JC: **Linkage disequilibrium for different scales and applications**. *Brief Bioinform* 2004, **5**(4):355-364.
65. Hudson RR: **Two-locus sampling distributions and their application**. *Genetics* 2001, **159**(4):1805-1817.
66. Auton A, McVean G: **Recombination rate estimation in the presence of hotspots**. *Genome Res* 2007, **17**(8):1219-1227.
67. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M *et al*: **The structure of haplotype blocks in the human genome**. *Science* 2002, **296**(5576):2225-2229.
68. Wang N, Akey JM, Zhang K, Chakraborty R, Jin L: **Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation**. *Am J Hum Genet* 2002, **71**(5):1227-1234.
69. Sobel E, Lange K: **Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics**. *Am J Hum Genet* 1996, **58**(6):1323-1337.
70. Thomas A, Gutin A, Abkevich V, Bansal A: **Multilocus linkage analysis by blocked Gibbs sampling** *Statistics and Computing* 2000, **10**:259-269.
71. Heath SC, Snow GL, Thompson EA, Tseng C, Wijsman EM: **MCMC segregation and linkage analysis**. *Genet Epidemiol* 1997, **14**(6):1011-1016.
72. Skrivanek Z, Lin S, Irwin M: **Linkage analysis with sequential imputation**. *Genet Epidemiol* 2003, **25**(1):25-35.
73. Gao G, Allison DB, Hoeschele I: **Haplotyping methods for pedigrees**. *Hum Hered* 2009, **67**(4):248-266.
74. Lindholm E, Zhang J, Hodge SE, Greenberg DA: **The reliability of haplotyping inference in nuclear families: misassignment rates for SNPs and microsatellites**. *Hum Hered* 2004, **57**(3):117-127.
75. Druet T, Schrooten C, de Roos AP: **Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle**. *J Dairy Sci* 2010, **93**(11):5443-5454.
76. Anderson CA, Pettersson FH, Barrett JC, Zhuang JJ, Ragoussis J, Cardon LR, Morris AP: **Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms**. *Am J Hum Genet* 2008, **83**(1):112-119.
77. Kruglyak L: **Prospects for whole-genome linkage disequilibrium mapping of common disease genes**. *Nat Genet* 1999, **22**(2):139-144.
78. Goddard ME, Hayes BJ: **Mapping genes for complex traits in domestic animals and their use in breeding programmes**. *Nat Rev Genet* 2009, **10**(6):381-391.
79. Georges M: **Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals**. *Annu Rev Genomics Hum Genet* 2007, **8**:131-162.
80. Sellner EM, Kim JW, McClure MC, Taylor KH, Schnabel RD, Taylor JF: **Board-invited review: Applications of genomic information in livestock**. *J Anim Sci* 2007, **85**(12):3148-3158.
81. Hu X, Gao Y, Feng C, Liu Q, Wang X, Du Z, Wang Q, Li N: **Advanced technologies for genomic analysis in farm animals and its application for QTL mapping**. *Genetica* 2009, **136**(2):371-386.
82. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis**. *PLoS Genet* 2006, **2**(12):e190.
83. Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P *et al*: **An Arabidopsis example of association mapping in structured samples**. *PLoS Genet* 2007, **3**(1):e4.
84. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al*: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness**. *Nat Genet* 2006, **38**(2):203-208.

85. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**(2):945-959.
86. MacLeod IM, Hayes BJ, Savin KW, Chamberlain AJ, McPartlan HC, Goddard ME: **Power of a genome scan to detect and locate quantitative trait loci in cattle using dense single nucleotide polymorphisms**. *J Anim Breed Genet* 2010, **127**(2):133-142.
87. Hauser E, Cremer N, Hein R, Deshmukh H: **Haplotype-based analysis: a summary of GAW16 Group 4 analysis**. *Genet Epidemiol* 2009, **33 Suppl 1**:S24-28.
88. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X *et al*: **Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing**. *Genetics* 1995, **139**(2):907-920.
89. de Koning DJ, Schulmant NF, Elo K, Moisio S, Kinos R, Vilkki J, Maki-Tanila A: **Mapping of multiple quantitative trait loci by simple regression in half-sib designs**. *J Anim Sci* 2001, **79**(3):616-622.
90. Nadesalingam J, Plante Y, Gibson JP: **Detection of QTL for milk production on Chromosomes 1 and 6 of Holstein cattle**. *Mamm Genome* 2001, **12**(1):27-31.
91. Rodriguez-Zas SL, Southey BR, Heyen DW, Lewin HA: **Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle**. *J Dairy Sci* 2002, **85**(11):3081-3091.
92. Viitala SM, Schulman NF, de Koning DJ, Elo K, Kinos R, Virta A, Virta J, Maki-Tanila A, Vilkki JH: **Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle**. *J Dairy Sci* 2003, **86**(5):1828-1836.
93. Khatkar MS, Thomson PC, Tammen I, Raadsma HW: **Quantitative trait loci mapping in dairy cattle: review and meta-analysis**. *Genet Sel Evol* 2004, **36**(2):163-190.
94. Blott S, Kim JJ, Moisio S, Schmidt-Kuntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D *et al*: **Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition**. *Genetics* 2003, **163**(1):253-266.
95. Hayes BJ, Chamberlain AJ, Maceachern S, Savin K, McPartlan H, MacLeod I, Sethuraman L, Goddard ME: **A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle**. *Anim Genet* 2009, **40**(2):176-184.
96. McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C *et al*: **Whole genome linkage disequilibrium maps in cattle**. *BMC Genet* 2007, **8**:74.
97. Sargolzaei M, Schenkel FS, Jansen GB, Schaeffer LR: **Extent of linkage disequilibrium in Holstein cattle in North America**. *J Dairy Sci* 2008, **91**(5):2106-2117.
98. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS *et al*: **A whole-genome assembly of the domestic cow, Bos taurus**. *Genome Biol* 2009, **10**(4):R42.
99. Nilsen H, Olsen HG, Hayes B, Nome T, Sehested E, Svendsen M, Meuwissen TH, Lien S: **Characterization of a QTL region affecting clinical mastitis and protein yield on BTA6**. *Anim Genet* 2009, **40**(5):701-712.
100. Nilsen H, Olsen HG, Hayes B, Sehested E, Svendsen M, Nome T, Meuwissen T, Lien S: **Casein haplotypes and their association with milk production traits in Norwegian Red cattle**. *Genet Sel Evol* 2009, **41**:24.
101. Bovenhuis H, Weller JI: **Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers**. *Genetics* 1994, **137**(1):267-280.
102. Velmala RJ, Vilkki HJ, Elo KT, de Koning DJ, Maki-Tanila AV: **A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle**. *Anim Genet* 1999, **30**(2):136-143.
103. Schopen GC, Koks PD, van Arendonk JA, Bovenhuis H, Visker MH: **Whole genome scan to detect quantitative trait loci for bovine milk protein composition**. *Anim Genet* 2009, **40**(4):524-537.

104. Farrell HM, Jr., Jimenez-Flores R, Bleck GT, Brown EM, Butler JE, Creamer LK, Hicks CL, Hollar CM, Ng-Kwai-Hang KF, Swaisgood HE: **Nomenclature of the proteins of cows' milk--sixth revision**. *J Dairy Sci* 2004, **87**(6):1641-1674.
105. Martin P, Szymanowska M, Zwierzchowski L, Leroux C: **The impact of genetic polymorphisms on the protein composition of ruminant milks**. *Reprod Nutr Dev* 2002, **42**(5):433-459.
106. Caroli AM, Chessa S, Erhardt GJ: **Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition**. *J Dairy Sci* 2009, **92**(11):5335-5352.
107. Schild TA, Geldermann H: **Variants within the 5′-flanking regions of bovine milk-protein-encoding genes. III. Genes encoding the Ca-sensitive caseins αs1, αs2 and β** *Theoretical and Applied Genetics* 1996, **93**:887-893.
108. Lien S, Gomez-Raya L, Steine T, Fimland E, Rogne S: **Associations between casein haplotypes and milk yield traits**. *J Dairy Sci* 1995, **78**(9):2047-2056.
109. Hallen E, Wedholm A, Andren A, Lunden A: **Effect of beta-casein, kappa-casein and beta-lactoglobulin genotypes on concentration of milk protein variants**. *J Anim Breed Genet* 2008, **125**(2):119-129.
110. Szymanowska M, Siadkowska E, Lukaszewicz M, Zwierzchowski L: **Association of nucleotide-sequence polymorphism in the 5'-flanking regions of bovine casein genes with casein content in cow's milk**. *Le Lait* 2004, **84**:579-590.
111. Bevilacqua C, Helbling JC, Miranda G, Martin P: **Translational efficiency of casein transcripts in the mammary tissue of lactating ruminants**. *Reprod Nutr Dev* 2006, **46**(5):567-578.

Paper I

Recent and historical recombination in the admixed Norwegian Red cattle breed

Marte Sodeland[1§], Matthew Kent[1, 2], Ben J. Hayes[2, 3], Harald Grove[1, 2] and Sigbjørn Lien[1, 2]

[1]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway.
[2] Centre for Integrative Genetics, Norwegian University of Life Sciences, N-1432 Aas, Norway. [3] Biosciences Research Division, Department of Primary Industries Victoria, Melbourne, Australia, 3083.

[§]Corresponding author

Abstract

Background

Comparison of recent patterns of recombination derived from linkage maps to historical patterns of recombination from linkage disequilibrium (LD) could help identify genomic regions affected by strong artificial selection, appearing as reduced recent recombination. Norwegian Red cattle (NRF) make an interesting case study for investigating these patterns as it is an admixed breed with an extensively recorded pedigree. NRF have been under strong artificial selection for traits such as milk and meat production, fertility and health.
While measures of LD is also crucial for determining the number of markers required for association mapping studies, estimates of recombination rate can be used to assess quality of genomic assemblies.

Results

A dataset containing more than 17,000 genome-wide distributed SNPs and 2600 animals was used to assess recombination rates and LD in NRF. Although low LD measured by r^2 was observed in NRF relative to some of the breeds from which this breed originates, reports from breeds other than those assessed in this study have described more rapid decline in r^2 at short distances than what was found in NRF. Rate of decline in r^2 for NRF suggested that to obtain an expected r^2 between markers and a causal polymorphism of at least 0.5 for genome-wide association studies, approximately one SNP every 15 kb or a total of 200,000 SNPs would be required. For well known quantitative trait loci (QTLs) for milk production traits on *Bos Taurus* chromosomes 1, 6 and 20, map length based on historic recombination was greater than map length based on recent recombination in NRF.
Further, positions for 130 previously unpositioned contigs from assembly of the bovine genome sequence (Btau_4.0) were found using comparative sequence analysis were validated by linkage analysis, and 28% of these positions corresponded to extreme values of population recombination rate.

Conclusion

While LD is reduced in NRF compared to some of the breeds from which this admixed breed originated, it is elevated over short distances compared to some other cattle breeds. Genomic regions in NRF where map length based on historic recombination was greater than map length based on recent recombination coincided with some well known QTL regions for milk production traits.
Linkage analysis in combination with comparative sequence analysis and detection of regions with extreme values of population recombination rate proved to be valuable for detecting problematic regions in the Btau_4.0 genome assembly.

Background

The historical pattern of recombination in the population of genomes of a species or breed contain an enormous amount of information on history of population size, including expansions and contractions, gene flow between other breeds, and selection [1]. It has also been demonstrated that rate of recombination is not uniform across a chromosomal segment,

rather recombination events tend to occur in recombination hotspots [2, 3]. The pattern of linkage disequilibrium (LD) in the current generation of a species reflects all of these processes. While the pattern of LD therefore contains much information, deciphering the relative contribution of each process to the current pattern of LD is challenging [1, 4-11]. Some additional insight into the relative contribution of each process can be gained from comparing historical patterns of recombination inferred from LD to recent patterns of recombination inferred from genetic maps. One hypothesis would be that in genome regions where large discrepancies are observed between map distances inferred from LD and genetic map distances, strong selection is occurring. Norwegian Red cattle (NRF) was developed mainly through crosses of old Norwegian breeds with other Scandinavian breeds like Swedish Red and White, Black and White Swedish and Finnish Ayrshire. Pedigree data has been recorded since formation of NRF, and the breed has been under strong artificial selection for traits such as milk and meat production, fertility and health. A further attraction of using NRF for this type of study is the extensive pedigree data available, assisting determination of frequency of recombination events between adjacent markers. The extent of LD in cattle has been investigated in a number of studies [7, 12-15]. Relative to humans cattle display elevated LD, which is likely due to small recent effective population size generally observed in livestock populations [7, 12-16]. Previous studies have shown some variation between cattle breeds in rate of decline in LD with increasing distance between genetic markers [15, 17, 18], which is also at least partly attributable to population history.

Another application for recombination rate estimates is within validation of positioning and assembly of genome contigs by linkage analysis. The bovine genome has recently been sequenced by a combined bacterial artificial chromosome and whole-genome shotgun approach [19]. The resulting Btau_4.0 assembly has contig and scaffold N50 sizes of 48.7 kb and 1.9 Mb respectively, and represents 95% of the total genome sequence placed on the 29 autosomes and the X chromosome. Construction of genetic maps in NRF was used to assess quality of the Btau_4.0 assembly and indicated a positional error rate of less than 0.8% [19]. The sequencing and assembly of larger genomes is a complex task with many challenges, and will usually result in imperfect assemblies. The desire to build a complete assembly is often at odds with the application of stringent merging criteria, and a compromise strategy resulting in longer scaffolds containing some assembly errors is usually the end result [20-22].

Aim of this study was to provide maps of historic and recent recombination rate in NRF, and then to attempt to use these to infer aspects of population history. Recombination rate information was also used to assess quality of the Btau_4.0 assembly.

Results and discussion

A total of 2,480 paternal half-sib NRF sires and 109 founding NRF sires were genotyped using the Affymetrix 25K MIP array. The final male genetic map contained 17,347 SNPs distributed on the 29 *Bos Taurus* chromosomes (BTAs) [19], and distributions of spacing between adjacent SNPs and minor-allele frequency (MAF) for the SNPs are presented in additional files 1 and 2. In order to examine the relationships between NRF and cattle breeds that have contributed to the development of NRF, 53 Holstein, 40 Finnish Ayrshire, 19 Sided Troender and Nordland Cattle and 39 Icelandic bulls were also genotyped. Icelandic cattle are

believed to have been derived from old Norwegian breeds approximately 1000 years ago. Genetic distances between breeds were investigated using a principal component analysis of the genomic relationship matrix among individuals of different and the same breed [23]. Principal component 1 (PC1), PC2 and PC3 are plotted in Figure 1. For PC1 and PC2, the Finnish Ayrshires and NRF animals group together, likely reflecting the high level of contribution of Finnish Ayrshire to NRF. Icelandic cattle appear genetically distinct, perhaps reflecting the 1000 years of genetic isolation of this breed from the other breeds. PC3 separates Holsteins from the other breeds. The principal component analysis also clearly demonstrates heterogeneity in composition among NRF. For example, some NRF animals have higher than average levels of relationship to Finnish Ayrshires, while other NRF animals have high levels of relationship with Holsteins.

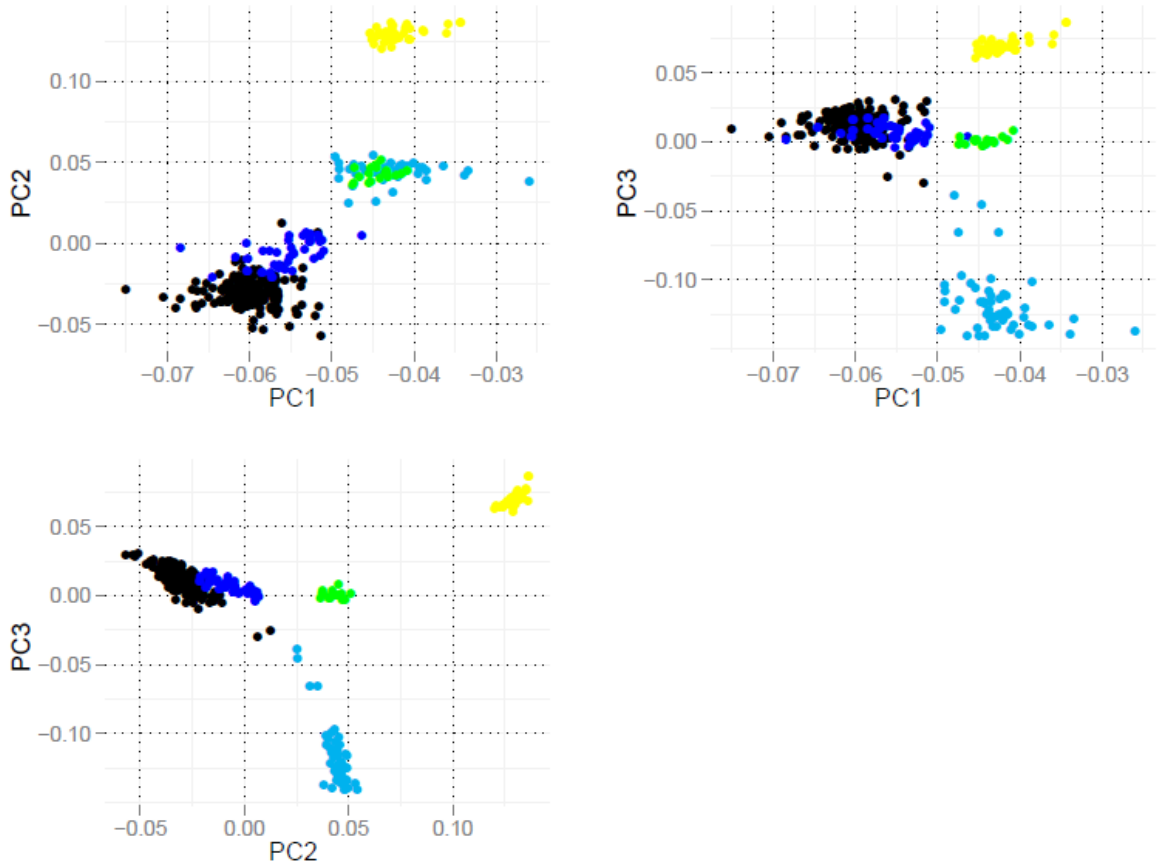


Figure 1 - Principal component analysis

Principal component (PC) analysis of genomic relationships among Norwegian Red cattle (black), Holsteins (light blue), Sided Troender and Nordland Cattle (green), Finnish Ayrshires (dark blue), and Icelandic Cattle (yellow). Plots are PC2 versus PC1, PC3 versus PC1 and PC3 versus PC2.

The extent of LD in each breed was assessed by average r^2 for pairs of markers binned by distance between them (Figure 2). At short distances (<100kb) Icelandic cattle had highest LD, likely reflecting small effective population size. NRF had lower levels of LD at comparable distances, especially distance greater than 100kb, than any of the other breeds. The low levels of LD observed in NRF relative to the other breeds is likely due to elevated heterogeneity in NRF from historic admixture, recent attempts to control inbreeding and gene flow through import of sires from other Nordic countries [24].
For NRF the highest and lowest chromosomal mean values for r^2 were found on BTA22 and BTA1, and highest and lowest mean values of r^2 for inter marker distances less than 10Mb were found for BTA5 and BTA19. Chromosomal mean values for NRF for r^2, and for r^2 for inter marker distances less than 10Mb, for all chromosomes are presented in Additional file 3. At very short inter-marker distances, the level of LD in NRF was high (Table 1). A mean r^2 of 0.5 or more was observed for SNPs positioned less than 10kb apart while a mean r^2 of 0.3 or more was observed for SNPs positioned less than 30kb apart. The results suggest that to obtain an expected r^2 between markers and a causal polymorphism of at least 0.5 for genome-wide association studies, approximately one SNP every 15 kb or a total of 200,000 SNPs would be required for the 2.87Gb genome. A report of decline in r^2 with increasing distance between SNPs in Australian Holstein-Friesian cattle [15] describes quite similar results as for NRF at these short distances. Reports from other breeds have described similar or more rapid decline in r^2 at short distances than what was found in NRF [17, 18]. However, long range LD (Figure 2) is lower in NRF than in these other breeds.

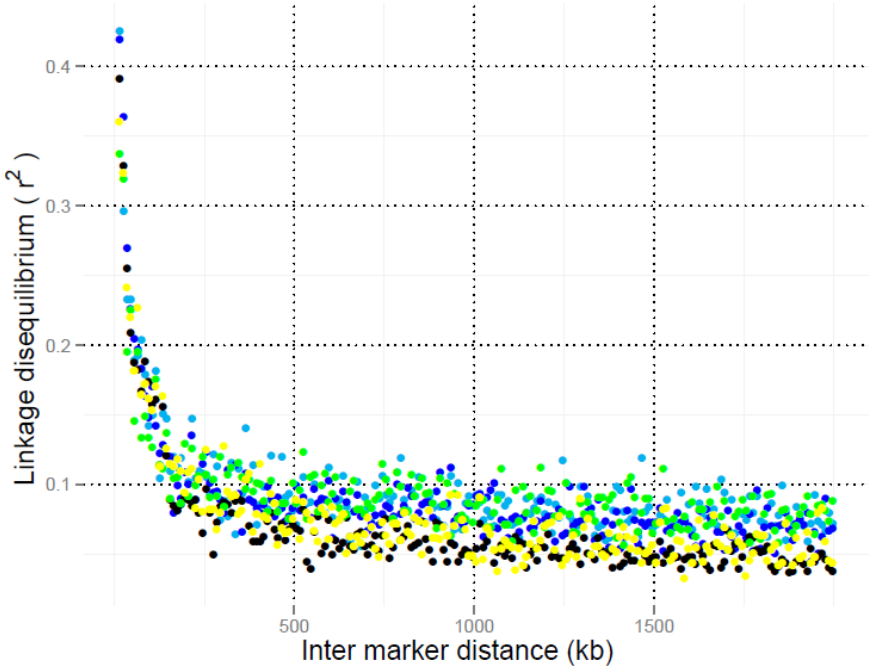


Figure 2 - Extent of linkage disequilibrium
Extent of linkage disequilibrium (r^2) in Norwegian Red cattle (black), Holsteins (light blue), Sided Troender and Nordland Cattle (green), Finnish Ayrshires (dark blue), and Icelandic Cattle (yellow).

Table 1- Expected linkage disequilibrium by inter-marker distance

Whole-genome mean r^2 for bins of short inter-marker distances (0 to 100kb) between syntenic SNPs in Norwegian Red cattle.

Distance (kb)	r^2 mean	r^2 sd
0-1	0.7497	0.34175332
1-5	0.5960	0.35871376
5-10	0.4770	0.36559827
10-20	0.3524	0.37987201
20-30	0.2680	0.39594648
30-40	0.2211	0.40066317
40-50	0.2187	0.3976346
50-100	0.1543	0.36471807

To investigate recombination patterns across genomic regions, maps describing historic LD levels were constructed for each chromosome based on population recombination rate in the NRF data. By the method presented by Auton and McVean [25], estimates of scaled population recombination rate ($\rho=4cN_e$) [9] were found for each interval between adjacent SNPs for all 29 BTAs with the LDhat software [26]. Following Pritchard and Przeworski [8], historical scaled recombination rate ($\rho(h)$) was compared with recent scaled recombination rate ($\rho(r)$) calculated from the genetic map by plotting their cumulative values against physical position (Figure 3).

Correlation between total cumulative $\rho(h)$ and $\rho(r)$ over all chromosomes was found to be 0.84. A reduced $\rho(r)$ relative to $\rho(h)$ for a genomic region could be an indication that animals in the observed pedigree have been under strong artificial selection for traits affected by polymorphisms in that particular region. Regions where $\rho(r)$ was most strikingly reduced relative to $\rho(h)$ were in the middle of BTA1 and in the middle of BTA20. Reduced $\rho(r)$ relative to $\rho(h)$ was also found on BTAs 6, 10,15, 16, 18, 19, 27 and 29, while elevated $\rho(r)$ relative to $\rho(h)$ was found on BTAs 3, 4, 7, 9, 11, 14, and 17.

On BTA1 several QTLs affecting milk production traits have been reported [27-32], and a meta-analysis reported by Khatkar *et al*. [33] indicated presence of three QTLs for milk yield on this chromosome. The BTA20 region centres around a mutation reported to affect protein percentage in the *GHR* gene [34]. Hayes *et al*. [35] reported evidence for strong selection in this region in a study of divergence between dairy cattle and beef cattle. On BTA6 two QTLs affecting milk production traits have been reported in NRF [36, 37] and signatures of strong selection have been detected [38].

Elevated population recombination rate may be due to population expansion or gene conversion, while reduced recombination rate may be due to directional selection, genetic drift, gene flow, population substructure or low effective population size [1]. Regions under strong selection in both historic and recent generations might not show differences between $\rho(h)$ and $\rho(r)$.

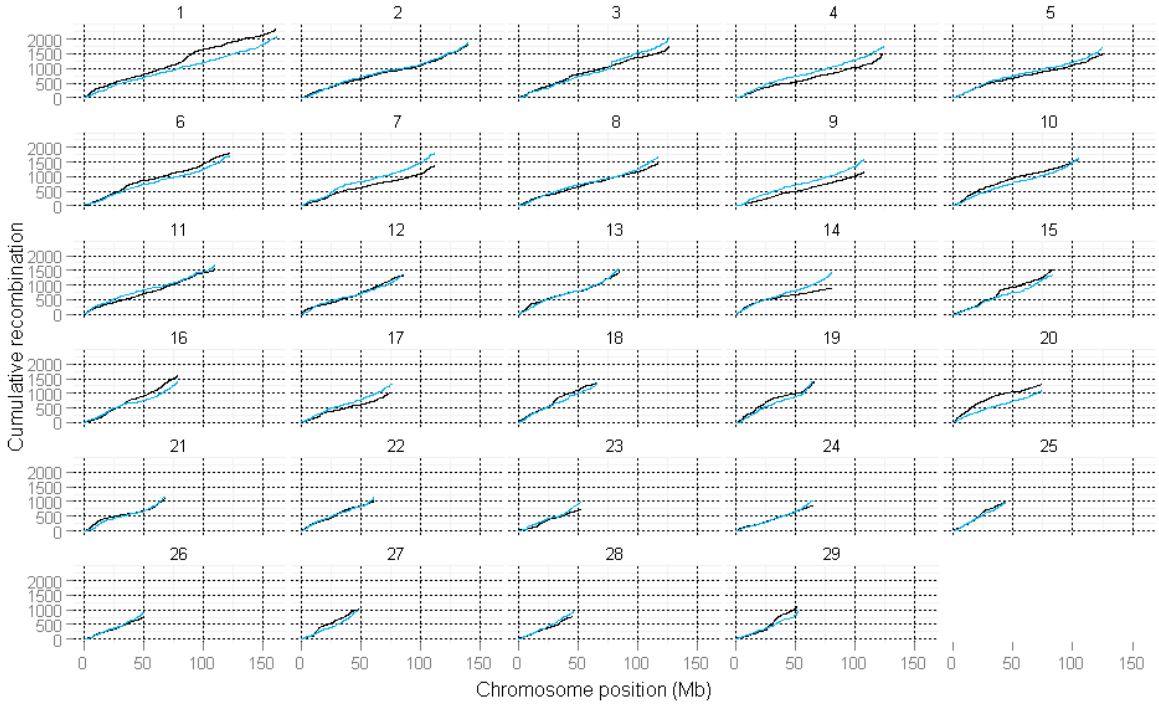


Figure 3 - Recent and historic recombination rate

Cumulative values of recent recombination rate (blue) and historic recombination rate (black) across each bovine autosomal chromosome is plotted against physical chromosome position (Mb). Chromosome numbers are indicated above each plot.

It was also investigated whether recombination rates per physical distance were related to chromosomal region or chromosome size. Telomeres showed significantly higher values for both recent and historic recombination rate per physical distance than the genome average (p-values 5.72e-12 and 2.97e-8). A negative correlation between recombination rate and chromosome length is expected [10, 39-41], and for ρ(h) and ρ(r) correlations of -0.6 and -0.83 was found between length of genetic map and physical chromosome length. Identification of chromosomes with unexpectedly elevated or reduced recombination rate could be identified by looking at outliers deviating from the expected linear relationship between recombination rate and chromosome length. In Figure 4 total cumulative values of recombination rate for ρ(h) and ρ(r) are plotted against physical chromosome length for each bovine chromosome. Expected recombination map lengths relative to chromosome lengths with 95% confidence intervals (CIs) are also indicated in the figure.
It can be seen that ρ(h) is elevated for BTAs 1, 13, 15, 16, 18, 19, 20 and 29, while ρ(r) is elevated for BTAs 3, 7, 13, 18 and 19. Further, reduced ρ(h)is observed for BTAs 4, 5, 7, 8, 9, 14, 17, 23, 24 and 26, while ρ(r)is reduced for BTAs 20 and 24. Consistently elevated ρ(h) and ρ(r) are found for BTAs 13, 18 and 19 and consistently reduced ρ(h) and ρ(r) are found for BTA24. In accordance with results described above, ρ(r) is significantly reduced relative to ρ(h) for BTA1, BTA6 and BTA20.

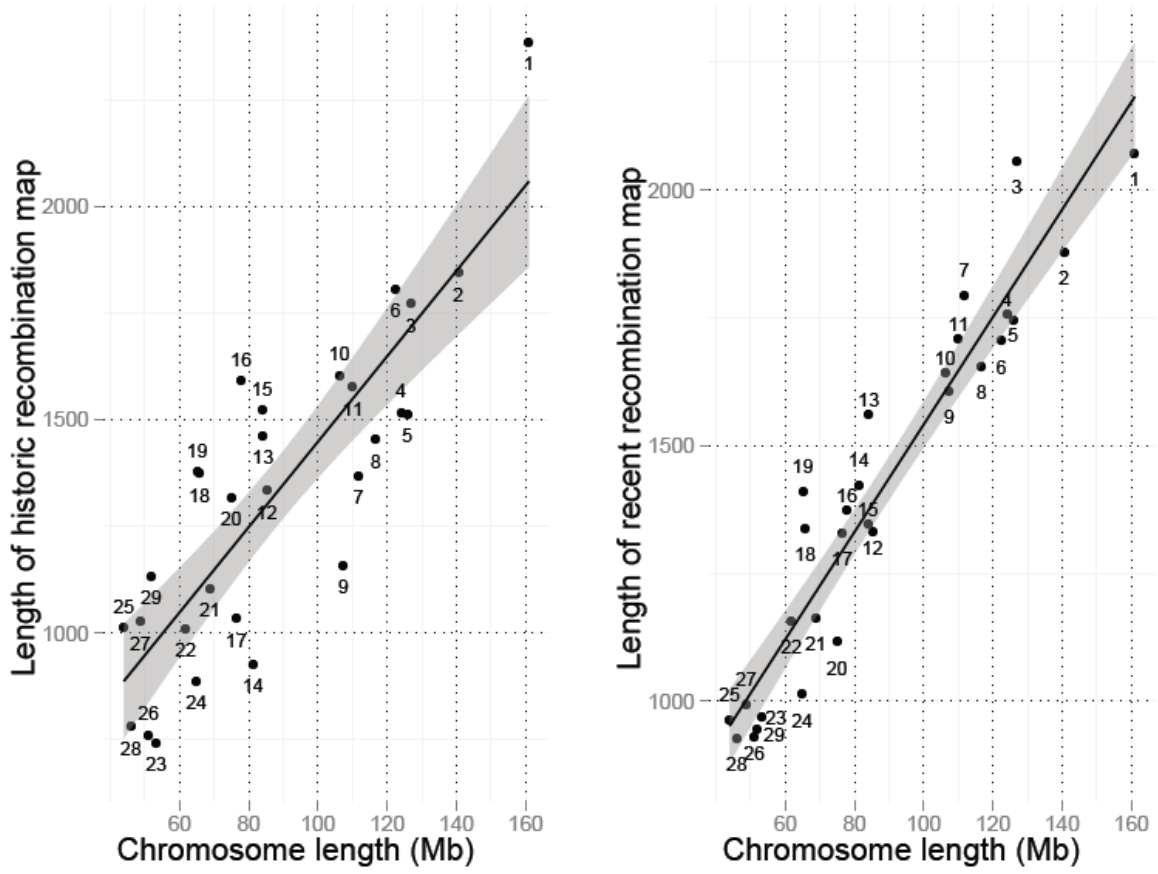


Figure 4 – Recombination maps versus physical chromosome length
Length of historic recombination map (left) and recent recombination map (right) are plotted against physical chromosome length for each bovine autosomal chromosome. The straight black lines indicate expected recombination map lengths relative to chromosome length and the darker grey regions their 95% confidence intervals.

In addition to chromosome region and chromosome size, sex-specific differences in recombination rates have been reported [42-44]. In this study male genetic maps for NRF were used, which might differ from female genetic maps. The patterns of recombination described here might also be sex-specific.

Moreover, rate of recombination is not uniform across a chromosomal segment and recombination events tend to occur in recombination hotspots [2, 3]. It has been estimated that these hotspots occur on average every 50 – 100kb in the human genome [45, 46]. Although the procedure applied here for estimation of $\rho(h)$ incorporates a model accounting for variable recombination rates across chromosomes [26], a SNP density higher than obtained in this study would be required in order to detect such fine-scale recombination hotspots in the bovine genome.

Quality assessment of Btau_4.0

Approximately 5% of the genomic sequence is expected to be missing in Btau_4.0 [19], and positioning of previously un-positioned contigs could aid completion of the assembly by pointing towards regions of special interest for re-sequencing efforts. Here a comparative analysis of the Btau_4.0 assembly with the human genome Build 19 allowed 4,276 previously un-positioned bovine contigs to be given putative genome positions. Determining recombination events between adjacent markers in an extensive pedigree can be used to construct dense genetic maps, and sufficient information was available from our NRF linkage analysis to validate the positions of 321 of these contigs [19]. Comparative analysis and linkage analysis identified 130 new contig positions as being less than 5Mb apart (Additional file 4). Even though large synteny blocks exist between species comparative analysis will yield spurious positions. Here 40% of positions identified by comparative analysis were validated by linkage analysis.

To further assess assembly accuracy, population recombination rates between adjacent SNPs were assessed. Regions of putative problematic assembly were identified as extreme values of scaled population recombination rate (ρ) relative to inter-marker distance between adjacent SNP pairs. Extreme values of ρ would be expected for intervals where assembled inter-marker distance was shorter than actual inter-marker distance. In Figure 5 ρ is plotted for BTAs 5, 6, 13 and 25. Contig positions predicted by comparative analysis are indicated in light grey, contig positions predicted by linkage analysis are indicated in dark grey and contig positions validated by similar positions from both comparative analysis and linkage analysis are indicated in light blue. Plots for all chromosomes are given in Additional file 5. From Figure 5 it can be seen that several ρ peaks were located near validated positions (indicated in light blue in fig. 5) for previously un-positioned contigs, which further highlights these regions as erroneous in the original assembly. Examples are the three ρ peak regions on BTA6 around 10Mb, 35Mb and 100Mb, the ρ peak region on BTA5 around 120Mb, the ρ peak region on BTA13 around 10Mb and the ρ peak region on BTA25 around 35Mb. Of the putative contig positions found by comparative analysis (indicated in light grey in fig. 5) 24% lay within 1Mb of an extreme value for population recombination rate ($\rho > 10$) and of positions found by linkage analysis (indicated in dark grey in fig. 5) 27% lay within 1Mb of an extreme value for population recombination rate. For the 130 contigs given similar positions by comparative analysis and linkage analysis 28% of validated positions (indicated in light blue in fig. 5) lay within 1Mb of an extreme value for population recombination rate.

Putative contig positions not coinciding with elevated ρ may be due to failure to detect regions with elevated recombination rate, incorrect positioning of un-positioned contigs or un-positioned contigs containing sequence overlap with already assembled contigs. Inability to detect regions with elevated recombination rate could result from surrounding SNPs not containing enough information or from un-positioned contigs being too short to detectably affect ρ. Incorrect positioning of un-positioned contigs could be due to repeat sequence mapping to similar but not equal genomic sequence or random mapping to the wrong position by linkage analysis. Incorrect positioning of un-positioned contigs could also explain why not all of positions found by linkage analysis for the 321 contigs validated positions found by comparative analysis.

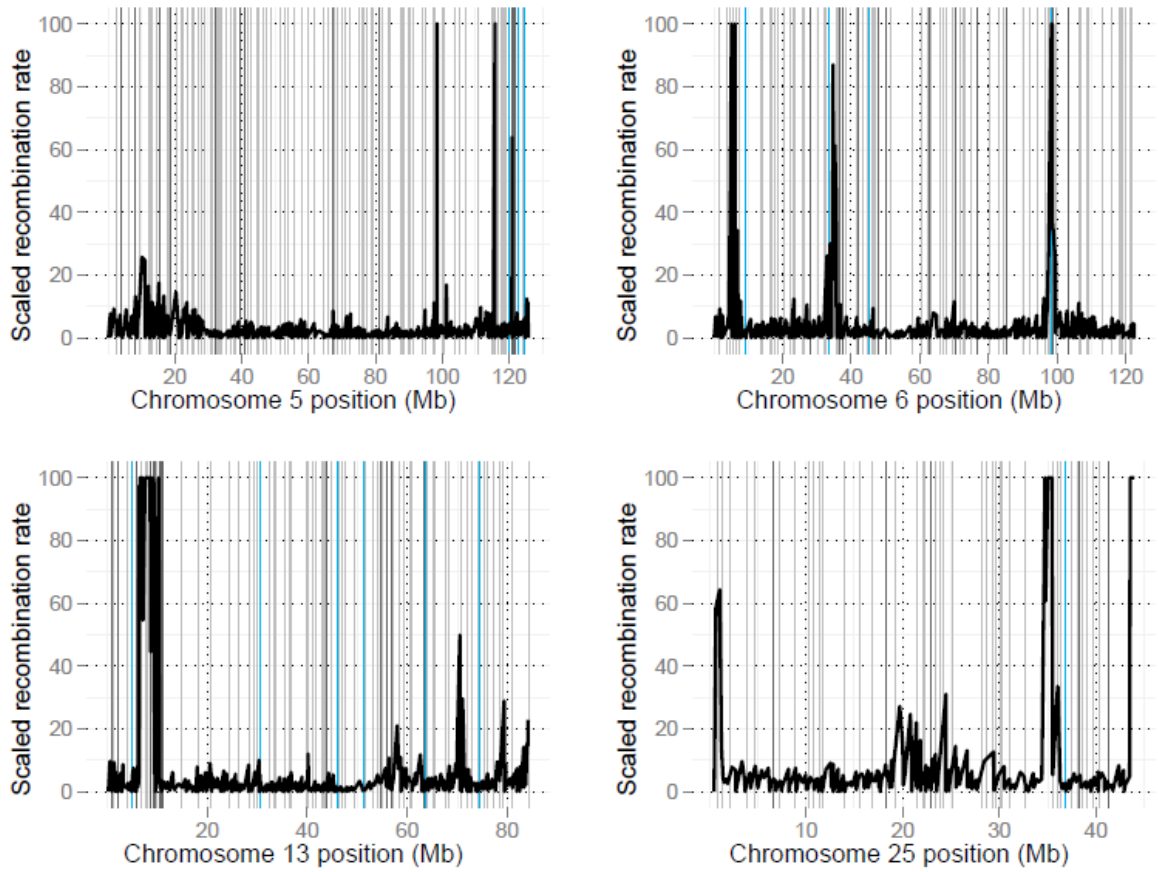


Figure 5 – Quality assessment of the bovine genome assembly

Scaled recombination rate versus physical distance (kb) is plotted for chromosomes 5 (top left), 6 (top right), 13 (bottom left) and 25 (bottom right). Contig positions predicted by comparative sequence analysis are indicated in light grey and contig positions predicted by linkage analysis are indicated in dark grey. Contigs given similar positions by both methods are indicated in light blue.

An alternative *Bos Taurus* genome assembly (UMD2) was reported by Zimin *et al*. [47]. This assembly had 95% identity with the Btau_4.0 assembly but had more genomic sequence placed on the bovine chromosomes. Regions differing between UMD2 and Btau_4.0 were identified by sequence alignments [47] and some of these regions coincide with regions identified as problematic here. Some examples are the region around 105Mb on BTA7, the region around 70Mb on BTA12, the region around 10Mb on BTA13, the region around 30Mb on BTA18 and the proximal ends of BTA20, BTA21 and BTA28 (Additional file 5). Genotyping of SNPs positioned on un-positioned contigs in a large pedigree and consequent linkage analyses as described here provide useful information for improving the current bovine genome assembly (Btau_4.0). The approach will gain even higher power and more accurate predictions as denser genetic maps become available. Likewise comparative sequence analysis would be a good supplement for correct contig or scaffold positioning.

Conclusions

Low levels of LD were observed in NRF relative to some of the breeds from which this breed originates. This is likely due to elevated heterogeneity in NRF from historic admixture, recent attempts to maintain a large effective population size through control of inbreeding and gene flow through import of sires from other Nordic countries. Reports from breeds other than those assessed in this study have described more rapid decline in r^2 at short distances [17, 18] than was found in NRF. The results suggested that to obtain an expected r^2 between markers and a causal polymorphism of at least 0.5 for genome-wide association studies in NRF, approximately one SNP every 15 kb or a total of 200,000 SNPs would be required for the 2.87Gb genome.

For well known QTL regions for milk production on BTA1, BTA6 and BTA20, map length based on historic recombination was greater than map length based on recent recombination in NRF. Selective sweeps have previously been identified for the QTL regions on BTA20 [35] and BTA6 [38]. Reduced $\rho(r)$ relative to $\rho(h)$ was also found on BTAs 10,15, 16, 18, 19, 27 and 29, while elevated $\rho(r)$ relative to $\rho(h)$ was found on BTAs 3, 4, 7, 9, 11, 14, and 17.

While over 95% of the total genome sequence is included in bovine genomic assembly Btau_4.0, problematic regions exists and should be identified to facilitate assembly completion. Here such regions were identified by combining comparative sequence analysis, linkage analyses and detection of regions with elevated population recombination rate.

Methods

Genotyping and initial filtering

The Affymetrix 25K MIP array [48] was used to genotype 2,589 NRF sires with paternal half-sib pedigree structure. In addition, 53 Holstein, 40 Finnish Ayrshire, 19 Sided Troender and Nordland Cattle and 39 Icelandic sires were genotyped. Genotypes were filtered for discordants (<2.5%), MAF (>0.025) and genotyped percentage (>75%). After initial filtering 17,483 SNPs remained. MAF were calculated for these SNPs with the Haploview 4.1 software [49].

Genetic map construction

Genetic maps for each of the 29 BTAs were constructed by use of the CRI-MAP 2.4 package [50]. The map file created by use of the CRI-MAP *fixed* option was checked for elevated recombination rates between adjacent SNPs. Elevated recombination rates could be an indication of a wrongly positioned contig in the assembly. SNPs with genetic distance >6 cM between its two flanking SNPs or a genetic distance >4cM between itself and one of its flanking SNPs were identified as suspicious and temporarily taken out of the genetic map. The CRI-MAP *chrompic* option was used to identify double recombinants. Double recombinants were manually inspected and corrected. SNPs showing up as double recombinants in more than 30 animals from 5 or more families were identified as suspicious and temporarily taken out of the genetic map. The *fixed* and *chrompic* procedures were

repeated until no SNPs showed a genetic distance >6cM between its two flanking SNPs or a genetic distance >4cM between itself and one of its flanking SNPs. The SNPs temporarily taken out of the genetic map were attempted repositioned by use of the CRI-MAP *twopoint* option. SNPs mapping more strongly to positions within 2.5Mb of their original positions were not repositioned. The genetic maps has previously been reported in Liu *et al*. [19].

Haplotypes and missing genotypes

The PHASE software [51] and the locally developed CRIHAP package were applied to utilize both linkage and LD information for determining haplotypes and impute missing genotypes.

Principal component analysis

A principal component analysis of the genomic relationship matrix among individuals of different and the same breed [23] was conducted to evaluate genetic distances between breeds.

Recombination rate distribution

Recombination rates in telomeric regions were compared to average recombination rates across chromosomes. Mean recombination rate per bp in a 10Mb telomeric region for all autosomes was compared to overall mean recombination rate per bp. A t-test was used to compare means.

Linkage disequilibrium

Estimates of pair-wise linkage disequilibrium measure r^2 were calculated with the Haploview 4.1 software [49].

Population recombination rate

Scaled population recombination rate (ρ) between adjacent markers relative to inter marker distances was estimated using the LDhat 2.1 software [26] with haplotypes from 17,347 SNPs distributed on the 29 BTAs. The LDhat 2.1 software incorporates a model which allows for variable recombination rates across chromosomes [26]. The reversible-jump markov chain monte carlo (rjMCMC) chain was run for 10,000,000 iterations, performing 5000 iterations between each sample. A block penalty of 5 was applied and the first 500,000 iterations were discarded as burn-ins. Some extreme values of ρ were observed, which could to be due to wrongly assembled contigs or other assembly artefacts. To determine historical recombination rate (ρ(h)) extreme values were corrected by replacing extreme values of ρ by a maximum value (max interval ρ =10). Less than 5% of intervals between adjacent SNPs had ρ higher than this maximum value. To determine values of recent scaled population recombination rate from the observed pedigree (ρ(r)), pedigree based estimates of recombination (c) was scaled by a factor corresponding to $4N_e$ (from $\rho=4cN_e$) [3, 8]. The applied scaling factor $\rho(h)/c = 4N_e$ was found by taking average values of total cumulative ρ(h) and total cumulative c for all autosomes. Cumulative values of recombination over each interval was used because we were interested in comparing a recent population recombination map based on the observed pedigree with a historic population recombination map, rather than comparing point estimates of recombination rate per distance. Cumulative interval values for ρ(r) and ρ(h)

were found by multiplying recombination rate per length unit for each interval with interval length and then calculating cumulative values across each autosomal chromosome.

Positioning previously un-positioned contigs

Positioning of previously un-positioned contigs from the bovine genome sequencing (Btau_4.0) [19] was done both by comparative sequence analysis with the human genome and by linkage analysis with already positioned SNPs in the bovine genome (Btau_4.0). In the comparative sequence analysis positioning of un-positioned bovine contigs was performed by combining MegaBLAST [52] searches for un-positioned contigs against the human genome Build 19 followed by MegaBLAST searches of hits in the human genome against the bovine genome (Btau_4.0). The first search revealed which areas of the human genome was most similar to each unknown contig, while the second search mapped those areas in the human genome sequence back to the bovine genome sequence. When two human sequences from one chromosomal region gave MegaBLAST hits against two sequences of a bovine chromosomal region it was assumed that the sequence between those human sequences on the human chromosome would also have similarity to the bovine genomic sequence between the two hits on the bovine chromosome. Bovine positions were predicted for 4,276 previously un-positioned contigs by this comparative method. Moreover, linkage analysis was conducted to position 568 SNPs distributed on 321 of the un-positioned contigs. The *twopoint* option in CRI-MAP 2.4 [50] was used to map these 568 SNPs to SNPs already positioned in the bovine genome assembly. The results from linkage analysis were also presented in Liu *et al.* [19].

Authors' contributions

MS construction of genetic maps, linkage analyses, linkage disequilibrium analyses for Norwegian Red cattle, recombination rate analyses and writing of manuscript. MK molecular genetic studies and initial data cleaning. BJH principal component analysis, linkage disequilbrium analysis for breed comparison and assistance in drafting the manuscript. HG comparative sequence analysis. SL initial data cleaning, construction of genetic maps, conceived of the study, coordination and assistance in drafting the manuscript. All authors helped finalize the manuscript and read and approved of the final version.

Acknowledgements

Thanks to Hanne Hamland and Monica Aasland Opsal for sample processing and Torfinn Nome for bioinformatics assistance. Thanks also to Johanna Vilkki and Emma Eyþórsdóttir for proving samples of Finish Ayshire and Icelandic Cattle, respectively. This project has been funded by The Research Council of Norway, GENO Breeding and AI association and BoviBank Ltd.

References

1. Ardlie KG, Kruglyak L, Seielstad M: **Patterns of linkage disequilibrium in the human genome**. *Nat Rev Genet* 2002, **3**(4):299-309.
2. Jeffreys AJ, Holloway JK, Kauppi L, May CA, Neumann R, Slingsby MT, Webb AJ: **Meiotic recombination hot spots and human DNA diversity**. *Philos Trans R Soc Lond B Biol Sci* 2004, **359**(1441):141-152.
3. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P: **A fine-scale map of recombination rates and hotspots across the human genome**. *Science* 2005, **310**(5746):321-324.
4. Slatkin M: **Linkage disequilibrium in growing and stable populations**. *Genetics* 1994, **137**(1):331-336.
5. Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O *et al*: **A composite of multiple signals distinguishes causal variants in regions of positive selection**. *Science* 2010, **327**(5967):883-886.
6. Patin E, Barreiro LB, Sabeti PC, Austerlitz F, Luca F, Sajantila A, Behar DM, Semino O, Sakuntabhai A, Guiso N *et al*: **Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes**. *Am J Hum Genet* 2006, **78**(3):423-436.
7. Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM: **Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes**. *J Anim Sci* 2003, **81**(3):617-623.
8. Pritchard JK, Przeworski M: **Linkage disequilibrium in humans: models and data**. *Am J Hum Genet* 2001, **69**(1):1-14.
9. Hudson RR: **Two-locus sampling distributions and their application**. *Genetics* 2001, **159**(4):1805-1817.
10. Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW *et al*: **Comparison of human genetic and sequence-based physical maps**. *Nature* 2001, **409**(6822):951-953.
11. Mueller JC: **Linkage disequilibrium for different scales and applications**. *Brief Bioinform* 2004, **5**(4):355-364.
12. Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M *et al*: **Extensive genome-wide linkage disequilibrium in cattle**. *Genome Res* 2000, **10**(2):220-227.
13. Vallejo RL, Li YL, Rogers GW, Ashwell MS: **Genetic diversity and background linkage disequilibrium in the North American Holstein cattle population**. *J Dairy Sci* 2003, **86**(12):4137-4147.
14. Odani M, Narita A, Watanabe T, Yokouchi K, Sugimoto Y, Fujita T, Oguni T, Matsumoto M, Sasaki Y: **Genome-wide linkage disequilibrium in two Japanese beef cattle breeds**. *Anim Genet* 2006, **37**(2):139-144.
15. Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Cavanagh JA, Barris W, Schnabel RD, Taylor JF, Raadsma HW: **Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel**. *BMC Genomics* 2008, **9**:187.
16. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME: **Novel multilocus measure of linkage disequilibrium to estimate past effective population size**. *Genome Res* 2003, **13**(4):635-643.
17. McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C *et al*: **Whole genome linkage disequilibrium maps in cattle**. *BMC Genet* 2007, **8**:74.
18. Sargolzaei M, Schenkel FS, Jansen GB, Schaeffer LR: **Extent of linkage disequilibrium in Holstein cattle in North America**. *J Dairy Sci* 2008, **91**(5):2106-2117.
19. Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y *et al*: **Bos taurus genome assembly**. *BMC Genomics* 2009, **10**:180.

20. Blakesley RW, Hansen NF, Gupta J, McDowell JC, Maskeri B, Barnabas BB, Brooks SY, Coleman H, Haghighi P, Ho SL *et al*: **Effort required to finish shotgun-generated genome sequences differs significantly among vertebrates**. *BMC Genomics* 2010, **11**:21.
21. Chain PS, Grafham DV, Fulton RS, Fitzgerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C *et al*: **Genomics. Genome project standards in a new era of sequencing**. *Science* 2009, **326**(5950):236-237.
22. Phillippy AM, Schatz MC, Pop M: **Genome assembly forensics: finding the elusive mis-assembly**. *Genome Biol* 2008, **9**(3):R55.
23. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis**. *PLoS Genet* 2006, **2**(12):e190.
24. **GENO** [www.geno.no]
25. Auton A, McVean G: **Recombination rate estimation in the presence of hotspots**. *Genome Res* 2007, **17**(8):1219-1227.
26. McVean G, Auton A: **LDhat 2.1: A package for the population genetic analysis of recombination**. *Department of Statistics, Oxford, OX1 3TG, UK* 2007.
27. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X *et al*: **Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing**. *Genetics* 1995, **139**(2):907-920.
28. de Koning DJ, Schulmant NF, Elo K, Moisio S, Kinos R, Vilkki J, Maki-Tanila A: **Mapping of multiple quantitative trait loci by simple regression in half-sib designs**. *J Anim Sci* 2001, **79**(3):616-622.
29. Nadesalingam J, Plante Y, Gibson JP: **Detection of QTL for milk production on Chromosomes 1 and 6 of Holstein cattle**. *Mamm Genome* 2001, **12**(1):27-31.
30. Rodriguez-Zas SL, Southey BR, Heyen DW, Lewin HA: **Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle**. *J Dairy Sci* 2002, **85**(11):3081-3091.
31. Viitala SM, Schulman NF, de Koning DJ, Elo K, Kinos R, Virta A, Virta J, Maki-Tanila A, Vilkki JH: **Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle**. *J Dairy Sci* 2003, **86**(5):1828-1836.
32. Schulman NF, Viitala SM, de Koning DJ, Virta J, Maki-Tanila A, Vilkki JH: **Quantitative trait Loci for health traits in Finnish Ayrshire cattle**. *J Dairy Sci* 2004, **87**(2):443-449.
33. Khatkar MS, Thomson PC, Tammen I, Raadsma HW: **Quantitative trait loci mapping in dairy cattle: review and meta-analysis**. *Genet Sel Evol* 2004, **36**(2):163-190.
34. Blott S, Kim JJ, Moisio S, Schmidt-Kuntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D *et al*: **Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition**. *Genetics* 2003, **163**(1):253-266.
35. Hayes BJ, Chamberlain AJ, Maceachern S, Savin K, McPartlan H, MacLeod I, Sethuraman L, Goddard ME: **A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle**. *Anim Genet* 2009, **40**(2):176-184.
36. Olsen HG, Nilsen H, Hayes B, Berg PR, Svendsen M, Lien S, Meuwissen T: **Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle**. *BMC Genet* 2007, **8**:32.
37. Nilsen H, Olsen HG, Hayes B, Sehested E, Svendsen M, Nome T, Meuwissen T, Lien S: **Casein haplotypes and their association with milk production traits in Norwegian Red cattle**. *Genet Sel Evol* 2009, **41**(1):24.
38. Hayes BJ, Lien S, Nilsen H, Olsen HG, Berg P, Maceachern S, Potter S, Meuwissen TH: **The origin of selection signatures on bovine chromosome 6**. *Anim Genet* 2008, **39**(2):105-111.

39. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M: **A second-generation linkage map of the human genome**. *Nature* 1992, **359**(6398):794-801.
40. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G *et al*: **A high-resolution recombination map of the human genome**. *Nat Genet* 2002, **31**(3):241-247.
41. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: **Comprehensive human genetic maps: individual and sex-specific variation in recombination**. *Am J Hum Genet* 1998, **63**(3):861-869.
42. Poissant J, Hogg JT, Davis CS, Miller JM, Maddox JF, Coltman DW: **Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep**. *BMC Genomics* 2010, **11**(1):524.
43. Paigen K, Szatkiewicz JP, Sawyer K, Leahy N, Parvanov ED, Ng SH, Graber JH, Broman KW, Petkov PM: **The recombinational anatomy of a mouse chromosome**. *PLoS Genet* 2008, **4**(7):e1000119.
44. Mank JE: **The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals**. *Genet Res* 2009, **91**(5):355-363.
45. Myers S, Spencer CC, Auton A, Bottolo L, Freeman C, Donnelly P, McVean G: **The distribution and causes of meiotic recombination in the human genome**. *Biochem Soc Trans* 2006, **34**(Pt 4):526-530.
46. Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P: **Human recombination hot spots hidden in regions of strong marker association**. *Nat Genet* 2005, **37**(6):601-606.
47. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS *et al*: **A whole-genome assembly of the domestic cow, Bos taurus**. *Genome Biol* 2009, **10**(4):R42.
48. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS *et al*: **Development and characterization of a high density SNP genotyping assay for cattle**. *PLoS One* 2009, **4**(4):e5350.
49. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps**. *Bioinformatics* 2005, **21**(2):263-265.
50. Green P, K. Falls and S. Crooks.: **Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St. Louis, Mo., USA**. 1990.
51. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data**. *Am J Hum Genet* 2001, **68**(4):978-989.
52. Zhang Z, Schwartz S, Wagner L, Miller W: **A greedy algorithm for aligning DNA sequences**. *J Comput Biol* 2000, **7**(1-2):203-214.

Additional files

Additional file 1- Inter-marker distance distribution

Additional file 2 – Minor allele frequency

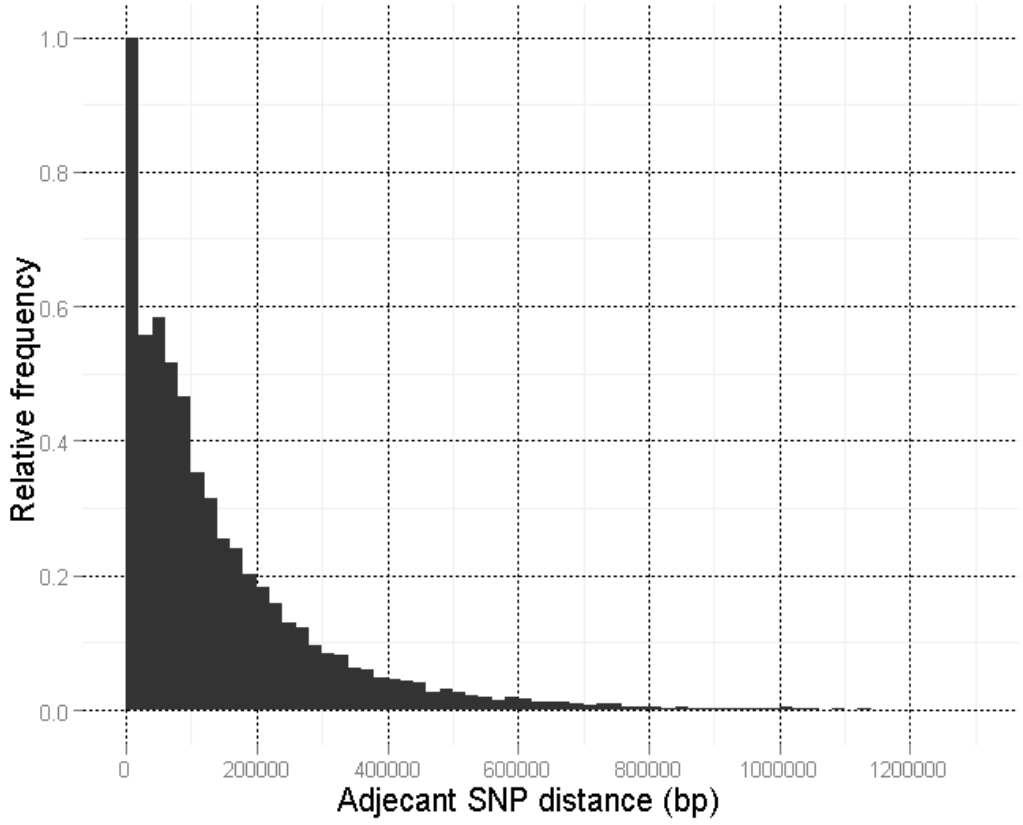
Additional file 3 – Chromosomal linkage disequilibrium

Additional file 4 – Positioning unpositioned contigs

Additional file 5 - Quality assessment of the bovine genome assembly Btau_4.0

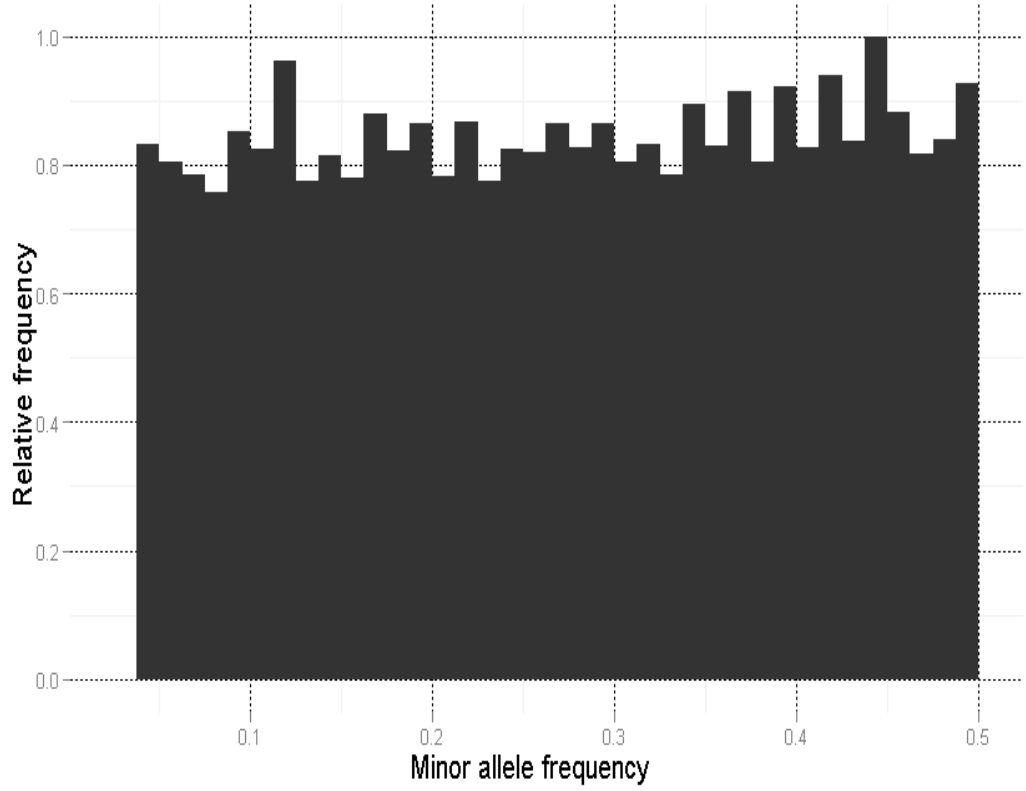
Additional file 1- Inter-marker distance distribution

Figure A1- Inter-marker distance distribution
Genome-wide distribution of distance (bp) between adjacent SNPs.



Additional file 2 – Minor allele frequency

Figure A2 – Minor allele frequency
Genome-wide distribution of minor allele frequencies after filtering (>0.025).



Additional file 3 – Chromosomal linkage disequilibrium

Table A3 – Chromosomal linkage disequilibrium

Number of SNPs, number of SNP pairs, mean chromosomal r^2 and mean r^2 for inter-marker distances <10Mb for the 29 BTAs.

BTA	Number of SNPs	Pairs	r^2	SD(r^2)	$r^2_{0\text{-}10Mb}$	SD($r^2_{0\text{-}10Mb}$)
1	1105	582660	0.00838	0.02921	0.0357	0.07459
2	972	445096	0.00988	0.03332	0.0359	0.07899
3	855	337431	0.00949	0.03248	0.03505	0.07482
4	871	350703	0.01107	0.03744	0.04255	0.08482
5	775	271216	0.01243	0.03844	0.04397	0.08511
6	859	355746	0.01055	0.03523	0.03835	0.07904
7	704	232903	0.01009	0.03551	0.03514	0.07779
8	788	290703	0.012	0.03595	0.04035	0.07735
9	676	215496	0.01207	0.03811	0.03871	0.08003
10	718	243951	0.01069	0.03705	0.03714	0.08009
11	776	285390	0.01003	0.03258	0.0344	0.07056
12	618	182710	0.01212	0.03866	0.03553	0.07509
13	610	172578	0.01219	0.04267	0.03735	0.08337
14	583	159330	0.01422	0.04577	0.04072	0.08693
15	534	131328	0.01373	0.04425	0.0385	0.08447
16	549	143380	0.01158	0.03989	0.03092	0.07479
17	566	150975	0.0139	0.04143	0.03636	0.07573
18	457	97903	0.01459	0.04818	0.03532	0.0829
19	439	87571	0.01131	0.0375	0.02773	0.06492
20	563	147696	0.01393	0.04265	0.03619	0.07924
21	420	81810	0.01564	0.04146	0.03639	0.07027
22	446	94395	0.01634	0.04917	0.03558	0.08181
23	379	67896	0.01609	0.04696	0.03557	0.07478
24	466	97020	0.01317	0.03624	0.03033	0.06263
25	287	39621	0.01552	0.04192	0.02948	0.06131
26	361	59685	0.01521	0.04719	0.0326	0.07409
27	332	52650	0.01517	0.04653	0.03085	0.07187
28	317	48205	0.01423	0.0424	0.02824	0.06491
29	321	47586	0.01567	0.04818	0.03273	0.07685
all	17357	4890974	0.01188	0.03854	0.03625	0.07768

Additional file 4 – Positioning unpositioned contigs

Table A4 - Positioning unpositioned contigs

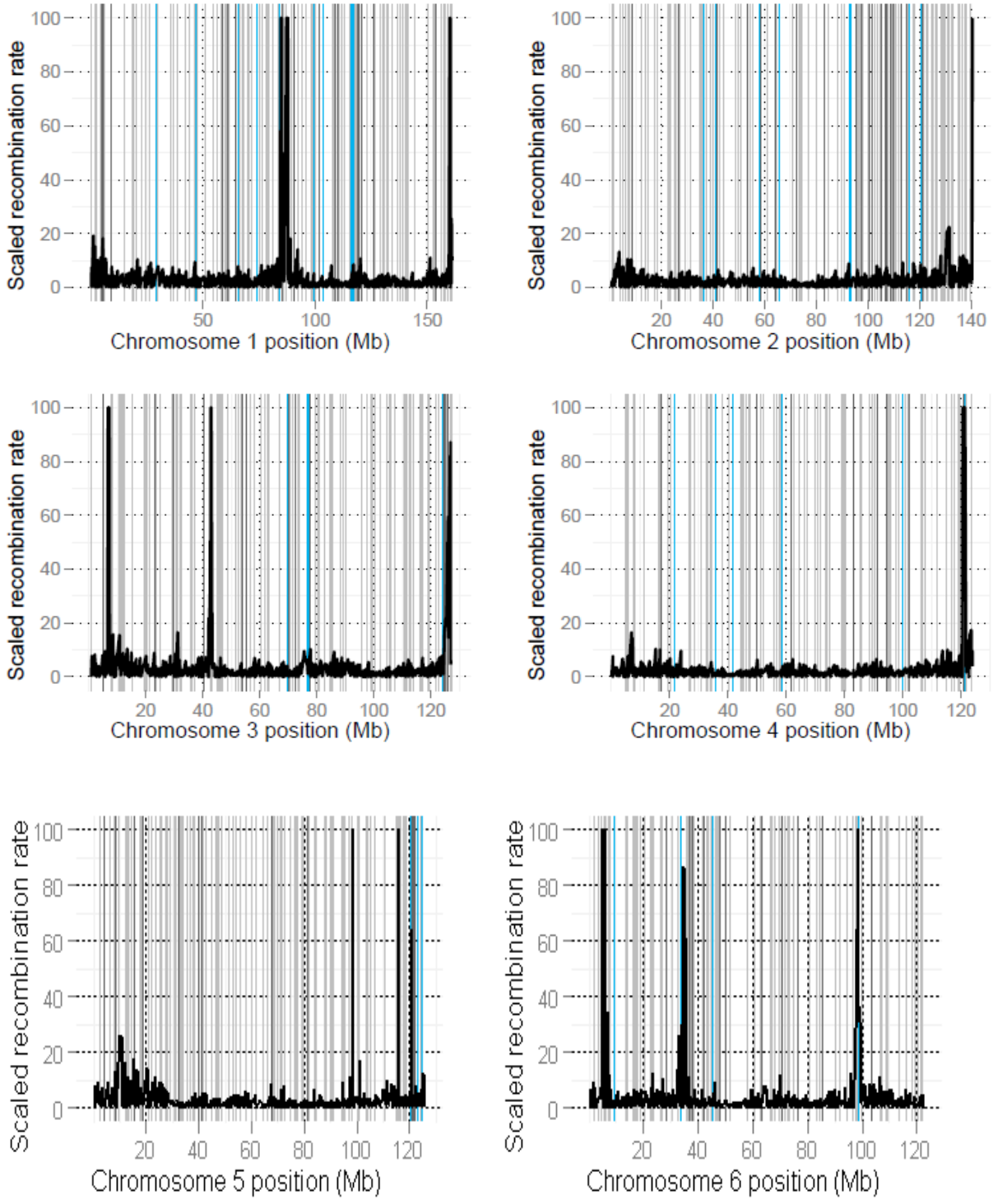
Comparative sequence analysis (CSA) contig positions were compared with the positions predicted by linkage analysis (LA) presented in Liu *et al*. [19]. The table shows 130 contigs unpositioned in the genome assembly (Btau_4.0) for which contig positions from these two prediction methods are less than 5Mb apart. Contig, BTA, position given by CSA and position given by LA is presented.

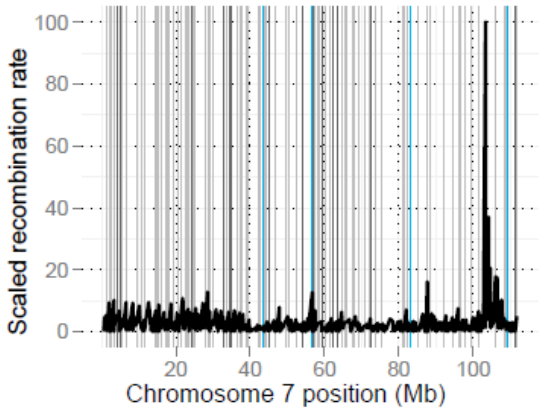
Contig	BTA	CSA pos (bp)	LA pos (bp)	Contig	BTA	CSA pos (bp)	LA pos (bp)
ChrUn.004.768	1	26281956	28879298	ChrUn.004.181	6	98157363	98421268
ChrUn.004.85	1	50891360	47153415	ChrUn.004.2758	7	44286170	43289003
ChrUn.004.1001	1	61874192	65957389	ChrUn.004.7	7	58031854	56735107
ChrUn.004.1144	1	74227536	74099262	ChrUn.004.538	7	83240720	83283438
ChrUn.004.371	1	84130992	83842131	ChrUn.004.47	7	111786467	109447065
ChrUn.004.321	1	95731182	99481432	ChrUn.004.2458	8	22646403	22996780
ChrUn.004.11	1	107740683	103459113	ChrUn.004.1528	8	77992562	77818564
ChrUn.004.1062	1	115643895	115969634	ChrUn.004.1112	8	85426569	85064076
ChrUn.004.29	1	117040459	116820290	ChrUn.004.1769	8	96160463	97868552
ChrUn.004.747	1	117040459	116934834	ChrUn.004.1350	8	105033113	107991032
ChrUn.004.766	2	36098554	36426908	ChrUn.004.310	9	59473956	56206860
ChrUn.004.765	2	41317799	40975539	ChrUn.004.22	9	63692465	62848329
ChrUn.004.297	2	61251434	58063626	ChrUn.004.4916	9	98928035	99397801
ChrUn.004.1449	2	64218868	65535887	ChrUn.004.1768	9	99823180	101723746
ChrUn.004.3074	2	88458990	92842235	ChrUn.004.360	10	293535	234438
ChrUn.004.382	2	90856067	93320532	ChrUn.004.3037	10	50630272	49803491
ChrUn.004.1701	2	111956989	116100367	ChrUn.004.519	10	52653988	51216078
ChrUn.004.404	2	120789554	120578545	ChrUn.004.1250	10	83209908	79864903
ChrUn.004.1480	3	66838524	69718616	ChrUn.004.114	10	95632496	95372032
ChrUn.004.3871	3	79358875	76728899	ChrUn.004.1844	10	99977821	103295043
ChrUn.004.5148	3	79358875	76997488	ChrUn.004.1919	11	2441949	3612595
ChrUn.004.25	3	125654871	124219358	ChrUn.004.705	11	3435759	4838605
ChrUn.004.3881	4	21574657	21635538	ChrUn.004.704	11	12876305	12891182
ChrUn.004.761	4	33910203	35636863	ChrUn.004.475	11	65530279	65893071
ChrUn.004.982	4	44630601	41683742	ChrUn.004.4462	11	90267750	89018683
ChrUn.004.700	4	61880729	58473419	ChrUn.004.135	11	109420358	110120453
ChrUn.004.712	4	100751169	99881861	ChrUn.004.4	12	3622333	219151
ChrUn.004.2120	4	119106981	121214561	ChrUn.004.2425	12	25695359	29298576
ChrUn.004.46	5	120128110	120200555	ChrUn.004.1654	12	36223108	35745307
ChrUn.004.101	5	124029455	122807308	ChrUn.004.23	12	55825173	57534765
ChrUn.004.152	5	124029455	124690330	ChrUn.004.51	12	60698388	59088156
ChrUn.004.1978	6	7421145	9256262	ChrUn.004.8	13	8129958	5056004
ChrUn.004.14	6	36811562	33498581	ChrUn.004.3218	13	30096787	30551956
ChrUn.004.688	6	47739165	45406895	ChrUn.004.3124	13	47000000	46102567

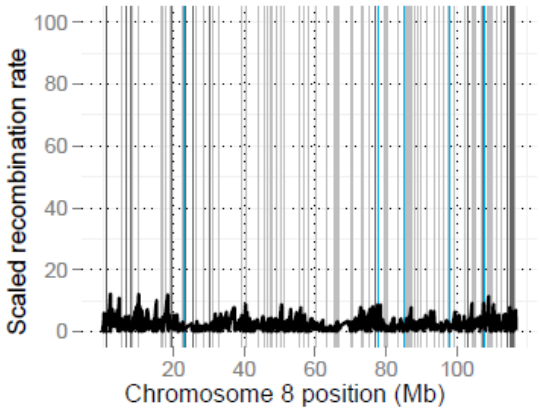
Contig	BTA	CSA pos (bp)	LA pos (bp)	Contig	BTA	CSA pos (bp)	LA pos (bp)
ChrUn.004.256	13	53922982	51389627	ChrUn.004.1047	20	21515643	19217974
ChrUn.004.288	13	64065293	63708121	ChrUn.004.1321	20	39352014	36458435
ChrUn.004.1967	13	73027223	74341206	ChrUn.004.374	20	59872532	60203096
ChrUn.004.3650	13	75876281	74341206	ChrUn.004.816	21	16878708	21334087
ChrUn.004.209	14	1888669	1647400	ChrUn.004.673	21	34453544	30105866
ChrUn.004.1	14	35153552	30580221	ChrUn.004.909	21	52240388	50972300
ChrUn.004.402	14	49666812	47064609	ChrUn.004.582	21	59001500	61410494
ChrUn.004.2216	14	52288903	49652936	ChrUn.004.2235	21	60354669	61410494
ChrUn.004.2679	14	70992550	70750938	ChrUn.004.177	21	68643427	64050679
ChrUn.004.1073	15	7126997	4707548	ChrUn.004.201	21	68830449	67177207
ChrUn.004.423	15	6467493	5727843	ChrUn.004.886	22	23071612	23043793
ChrUn.004.28	15	7126997	10819187	ChrUn.004.187	22	50136329	51965281
ChrUn.004.289	15	44694489	42584302	ChrUn.004.340	23	48042797	46627945
ChrUn.004.1936	15	58220038	59000656	ChrUn.004.1225	23	52000000	47444533
ChrUn.004.2	16	20886394	18185763	ChrUn.004.3014	24	38715301	38708428
ChrUn.004.3	16	34310295	30983359	ChrUn.004.894	24	49263691	50125408
ChrUn.004.721	16	32323091	33295121	ChrUn.004.4000	24	55804015	58376977
ChrUn.004.481	16	77473196	75179404	ChrUn.004.242	25	38641412	36887600
ChrUn.004.108	16	76018320	77581778	ChrUn.004.5408	25	38039057	36887600
ChrUn.004.492	17	55350140	54752610	ChrUn.004.1167	26	23996164	23755539
ChrUn.004.226	17	61858842	57827509	ChrUn.004.377	26	46864751	47322086
ChrUn.004.428	17	69725288	71543454	ChrUn.004.240	26	50194359	49465313
ChrUn.004.1536	17	68540063	72151921	ChrUn.004.2193	26	51734544	51054953
ChrUn.004.794	18	10433035	10168415	ChrUn.004.276	28	298870	3035821
ChrUn.004.660	18	15139873	16858980	ChrUn.004.5489	28	2492789	3220800
ChrUn.004.354	18	26457691	23383717	ChrUn.004.261	28	510799	3460323
ChrUn.004.2172	19	41115255	42058156	ChrUn.004.432	29	33931870	34967060
ChrUn.004.706	19	53560899	54452705	ChrUn.004.65	29	47533875	46824437
ChrUn.004.241	19	60822297	58035657	ChrUn.004.171	29	51580000	47525062
ChrUn.004.331	19	63091394	61096438	ChrUn.004.137	29	51580000	51539390
ChrUn.004.1236	20	4577209	4745147	ChrUn.004.163	29	51580000	51539390

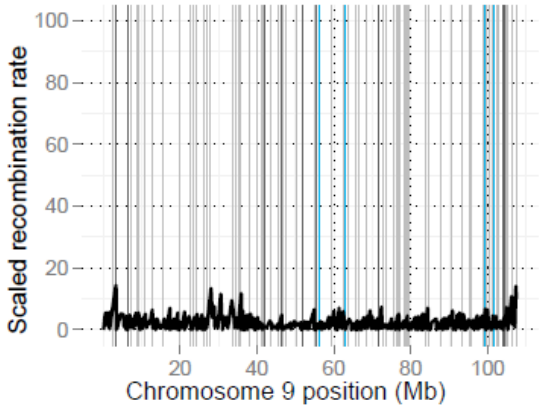
Additional file 5 - Quality assessment of the bovine genome assembly Btau_4.0

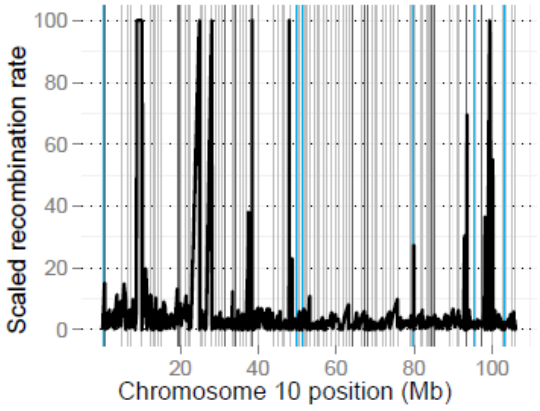
Figure A5 - Quality assessment of the bovine genome assembly Btau_4.0
Scaled recombination rate versus physical distance (kb) is plotted for all 29 autosomal bovine chromosomes. Contig positions predicted by comparative sequence analysis are indicated in light grey and contig positions predicted by linkage analysis are indicated in dark grey. Contigs given similar positions by both methods are indicated in light blue.

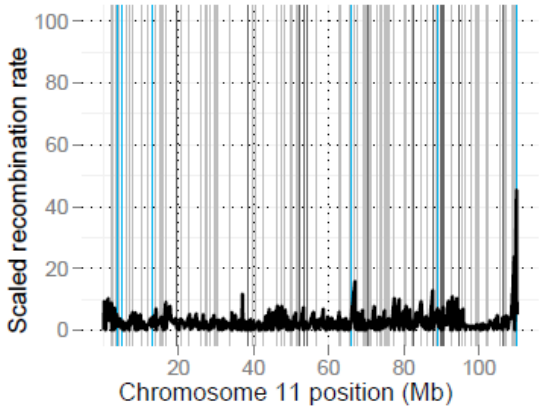


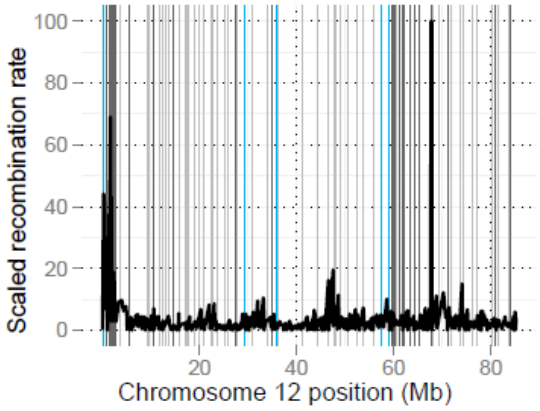

Scaled recombination rate
0
20
40
60
80
100
20
40
60
80
100
Chromosome 7 position (Mb)

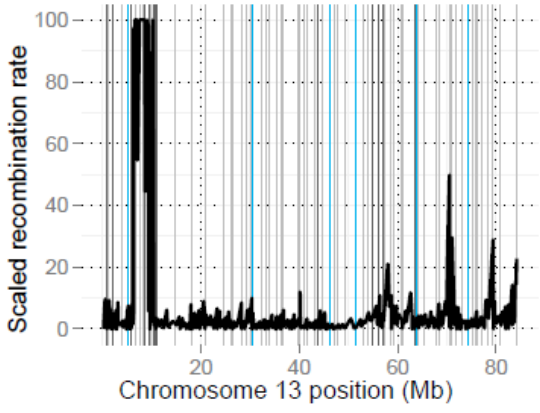

Scaled recombination rate
0
20
40
60
80
100
20
40
60
80
100
Chromosome 8 position (Mb)

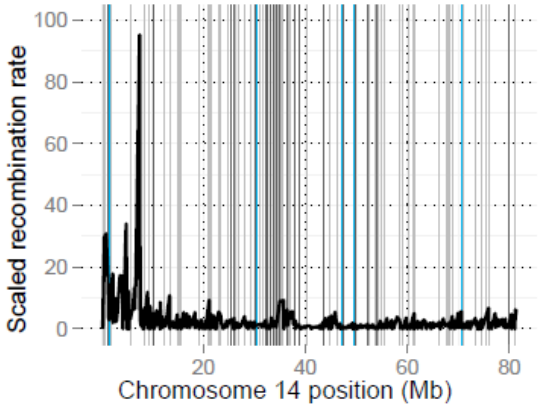

Scaled recombination rate
0
20
40
60
80
100
20
40
60
80
100
Chromosome 9 position (Mb)

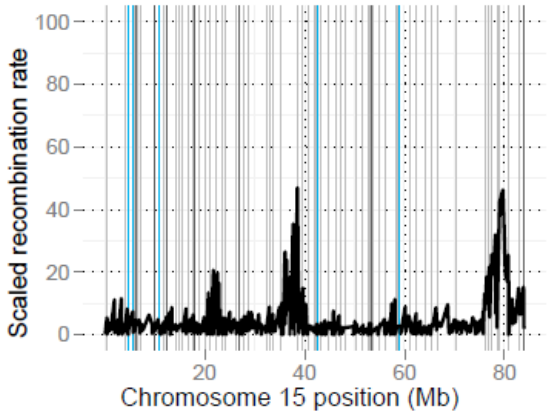

Scaled recombination rate
0
20
40
60
80
100
20
40
60
80
100
Chromosome 10 position (Mb)

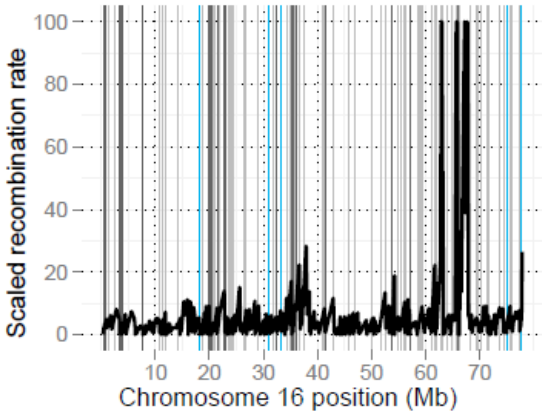

Scaled recombination rate
0
20
40
60
80
100
20
40
60
80
100
Chromosome 11 position (Mb)

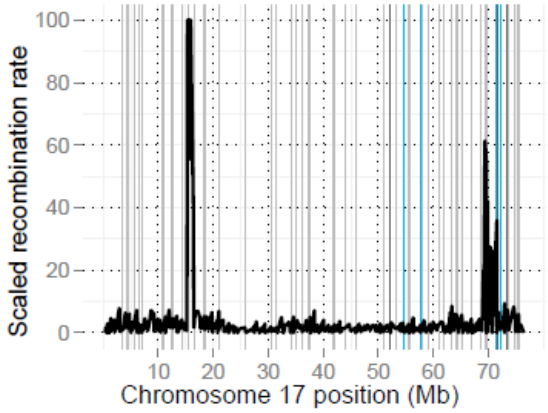

Scaled recombination rate
0
20
40
60
80
100
20
40
60
80
Chromosome 12 position (Mb)

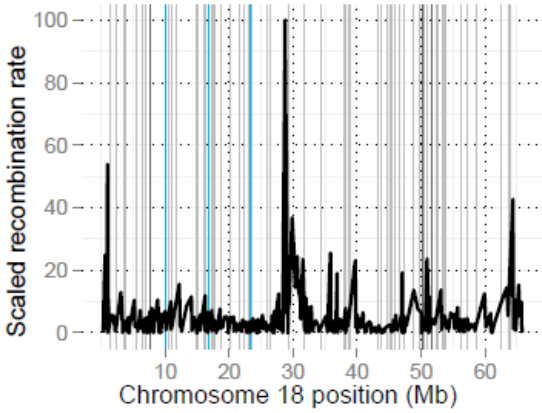

Scaled recombination rate
0
20
40
60
80
100
20
40
60
80
Chromosome 13 position (Mb)

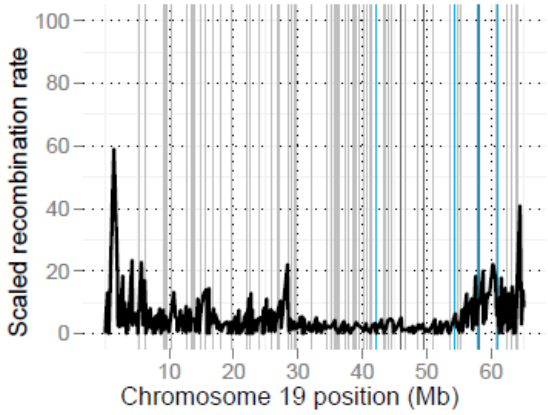

Scaled recombination rate
0
20
40
60
80
100
20
40
60
80
Chromosome 14 position (Mb)

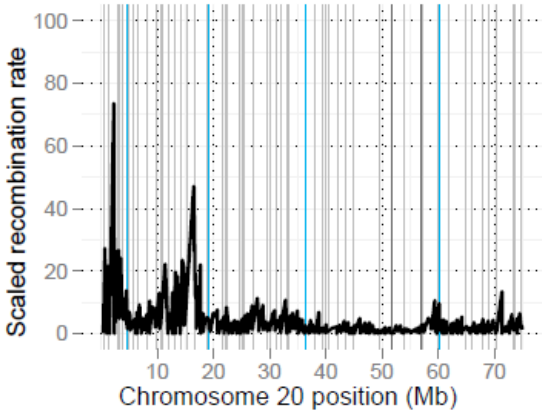

Scaled recombination rate
100
80
60
40
20
0
20
40
60
80
Chromosome 15 position (Mb)

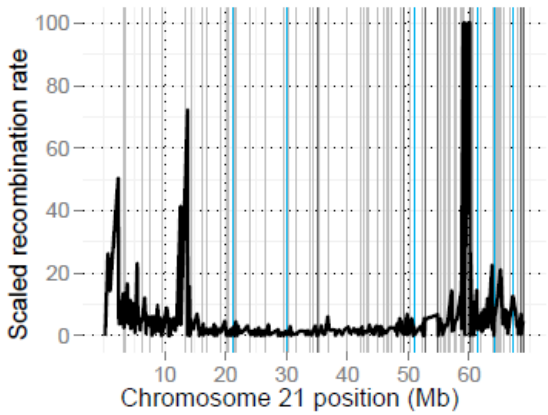

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
60
70
Chromosome 16 position (Mb)

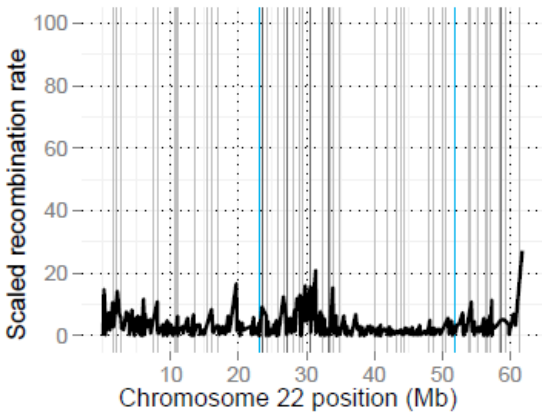

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
60
70
Chromosome 17 position (Mb)

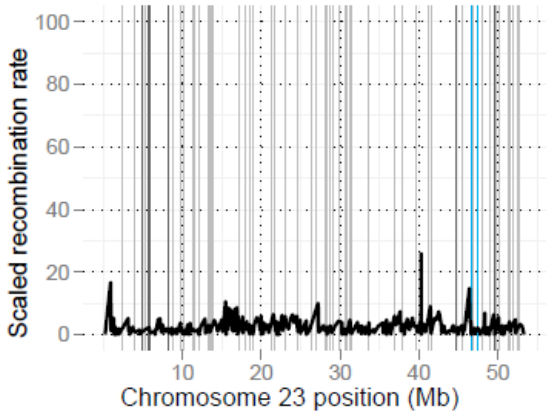

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
60
Chromosome 18 position (Mb)

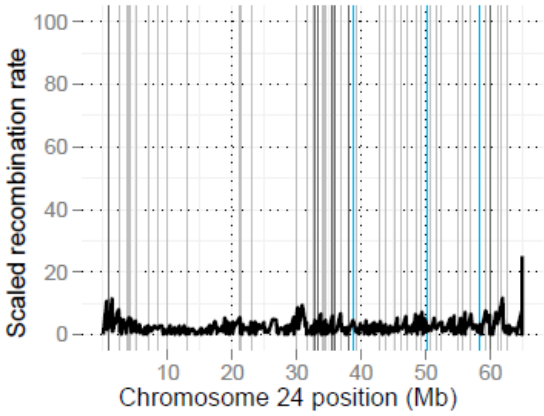

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
60
Chromosome 19 position (Mb)

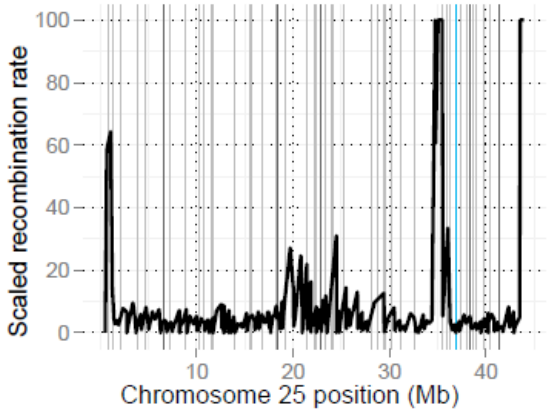

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
60
70
Chromosome 20 position (Mb)

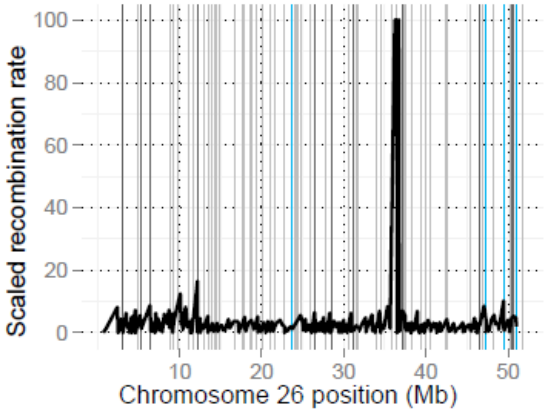

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
60
Chromosome 21 position (Mb)

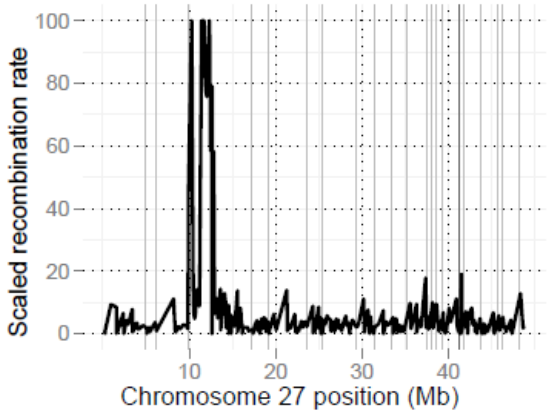

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
60
Chromosome 22 position (Mb)

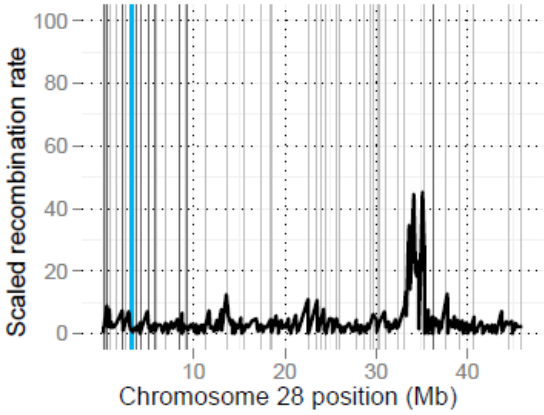

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
Chromosome 23 position (Mb)

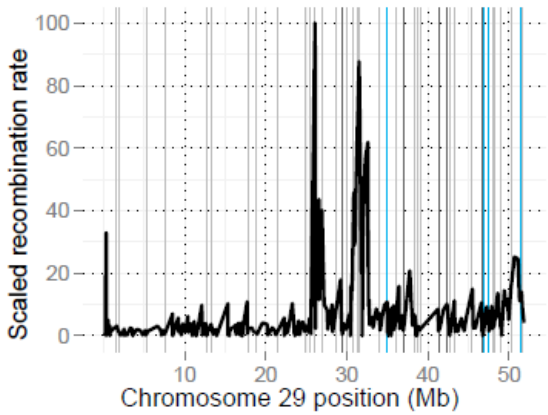

Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
60
Chromosome 24 position (Mb)


Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
Chromosome 25 position (Mb)


Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
Chromosome 26 position (Mb)


Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
Chromosome 27 position (Mb)


Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
Chromosome 28 position (Mb)


Scaled recombination rate
100
80
60
40
20
0
10
20
30
40
50
Chromosome 29 position (Mb)

Paper II

Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle

M. Sodeland, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway.
For correspondence and reprint requests:
E-mail: marte.sodeland@umb.no Telephone: +47 64 96 60 35 Fax: +47 64 96 51 01

M.P. Kent, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway. Centre for Integrative Genetics, Norwegian University of Life Sciences, N-1432 Aas, Norway.

H. G. Olsen, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway. Centre for Integrative Genetics, Norwegian University of Life Sciences, N-1432 Aas, Norway.

M. A. Opsal, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway.

M. Svendsen, Geno Breeding and AI organization, Norwegian University of Life Sciences, Box 5003, N-1432 Aas, Norway.

E. Sehested, Geno Breeding and AI organization, Norwegian University of Life Sciences, Box 5003, N-1432 Aas, Norway.

B.J. Hayes, Biosciences Research Division, Department of Primary Industries Victoria, Melbourne, Australia, 3083.
Centre for Integrative Genetics, Norwegian University of Life Sciences, N-1432 Aas, Norway.

S. Lien, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway.
Centre for Integrative Genetics, Norwegian University of Life Sciences, N-1432 Aas, Norway.

Summary
Mastitis is the most frequent and costly disease in dairy production and solutions leading to a reduction in the incidence of mastitis are highly demanded. Here genome-wide association studies were performed to identify polymorphisms affecting susceptibility to mastitis. Genotypes for more than 17,000 SNPs distributed across the 29 bovine autosomal chromosomes from over 2,500 sires with almost 1.4 million daughters with records on clinical mastitis were included in the analysis. Records of occurrence of clinical mastitis were divided into seven time periods in the three first lactations in order to identify quantitative trait loci affecting mastitis susceptibility in particular phases of lactation. The most convincing results from the association mapping were followed up and validated by a combined linkage disequilibrium and linkage analysis. The study revealed quantitative trait loci affecting occurrence of clinical mastitis in the periparturient period on chromosomes 2, 6 and 20, and a quantitative trait loci affecting occurrence of clinical mastitis in late lactation on chromosome 14. None of the quantitative trait loci for clinical mastitis detected in the study seem to affect lactation average of somatic cell score. SNPs highly associated with clinical mastitis lie near both the gene coding for interleukin 8 on chromosome 6 and the genes coding for the two interleukin 8 receptors on chromosome 2.

Background
Mastitis has a substantial impact on the dairy industry. The disease affects approximately 25 percent of Norwegian dairy cows each year (Østeras 2006) and estimated annual losses for dairy farmers caused by mastitis are $2 billion in the US and £300 million in the UK (Viguier *et al.* 2009). Infections imply great costs to farmers, animal suffering and use of antibiotics. In Norwegian Red cattle (NRF) *Staphylococcus aureus* is identified in the inflamed udder quarter of 55 percent of cows with clinical mastitis (CM) (Østeras 2006).
Bacterial infections are usually cleared by host defence mechanisms or antibiotic treatment within few days. If host defence and antibiotic treatment are not successful the outcome may be chronic infection, mammary gland tissue damage or death (Waller 2000; Sordillo 2005; Strandberg *et al.* 2005; Lahouassa *et al.* 2007).
Immunological defences of the mammary gland include anatomical features, cells, soluble molecules and receptors. Efficiency of this defence changes through the stages of lactation. Susceptibility to mastitis increases during the periparturient period (-15 to 30 days postpartum) and two thirds of mastitis incidents occur in the two first months of lactation (Syvajarvi *et al.* 1986; Waller 2000; Sordillo 2005). During the periparturient period the mammary gland goes through a transition to initiate milk production. The transition requires both hormonal changes and higher energy demand. Anatomical, cellular and soluble defences against infection are all altered or impaired during this period (Waller 2000; Sordillo 2005; Østeras 2006). In NRF patterns of strong genetic correlation for CM in the periparturient period, as well for CM in late lactation, have been found between the three first lactations (Svendsen & Heringstad 2006a). Since defence mechanisms change through stages of lactation it is reasonable to treat occurrence of CM in the different stages of lactation as different traits when attempting to map and characterise quantitative trait loci (QTLs) affecting susceptibility to CM. Previous (linkage mapping) studies have reported QTLs affecting CM on a number of *Bos Taurus* chromosomes (BTAs). Klungland *et al.* (2001) reported a QTL for CM on BTAs 3, 4, 6, 14 and 27 in NRF, Sahana *et al.* (2008) reported a QTL for CM on BTA9 at 73,9Mb in three Nordic cattle breeds and Schulman *et al.* (2004) reported QTLs for mastitis on BTAs 14 and 18 in Finnish Ayrshire.

Several authors have used high somatic cell score (SCS) in milk as an indication of clinical mastitis in QTL mapping (Bennewitz *et al.* 2003; Kuhn *et al.* 2003; Leyva-Baca *et al.* 2007). A wide range of values for genetic correlation between CM and SCS in cattle populations have been reported, with an average of about 0.7 (Mrode & Swanson 1996). In a NRF study of genetic correlations between SCS and occurrence of CM in different stages of lactation, higher correlations were found between SCS and occurrence of CM in late lactation than between SCS and occurrence of CM in the periparturient period (Svendsen & Heringstad 2006b). While a high value of SCS is an indicator of disease a low value might not be such a good indicator of udder status (Heringstad *et al.* 2000).
In NRF occurrence of CM is also correlated with protein yield in milk (Heringstad *et al.* 2005). This correlation could be due to linkage between QTLs influencing milk production traits and QTLs influencing CM or due to presence of pleiotropic effects. Pleiotropic effects of increased milk production on occurrence of CM might be due to increased energy demand or increased strain on anatomical features of the mammary gland.
Detection of QTLs in NRF and other cattle populations is facilitated by the bovine genome sequence (Liu *et al.* 2009) and the availability of large scale bovine SNP-arrays (Gibbs *et al.* 2009; Matukumalli *et al.* 2009). NRF is an ideal model breed for genome-wide association studies (GWAS) for CM as very large numbers of veterinary reported clinical mastitis (VRCM) records are available for this breed (Østeras *et al.* 2007).

Materials and methods
Animals and experimental design NRF is a mixed breed formed from local Norwegian breeds, Ayrshire and Swedish Red and White, but also with some influence of Holstein. Norway has a nationwide recording system for health data from dairy cattle. The national based Norwegian Dairy Herd Recording System has included VRCM since 1975 and records on SCS since 1978 (Østeras *et al.* 2007). Here a design with genotypes from NRF sires and phenotypic records from their daughters for VRCM or SCS was used. Such a design benefits from the large number of records per sire, which gives a marked decrease in variance due to environmental effects compared with other designs (Weller *et al.* 1990). Records of VRCM were retrieved as a binary trait for a total of 1,389,776 daughters of 2,086 paternal half-sib sires from 109 families. Number of daughters per sire ranged from 45 to 5,793 with a median value of 285. VRCM records were divided into occurrence of CM in each of seven categorical time periods, treating occurrences of CM in the different time periods as different traits. First lactation was divided into three time periods, whereas second and third lactation were divided into two time periods each. Time periods are described in Table 1. Daughter -yield-deviations (DYD) were calculated for each sire for CM and SCS based on daughter records for each sire. DYD for CM were calculated for each of the seven time periods since CM in the different time periods were treated as different traits in our analyses. Records on SCS were retrieved as lactation averages for 2,791,524 daughter lactations of 2,118 paternal half-sib sires from 109 families. Number of daughter lactations per sire ranged from 11 to 22,516 with a median value of 558.

Table 1
Records on occurrence of clinical mastitis in granddaughters were divided into seven time periods (CM1, CM2, CM3, CM4, CM5, CM6 and CM7) as described by the table. Time period (TP), lactation, days postpartum, number of records (N) and standard deviation (SD) for DYD of the trait are given.

TP	Lactation	Days postpartum	N	SD (DYD)
CM1	1	-15 to 30	1,389,776	0.02922895
CM2	1	31 to 120	1,375,776	0.01474941
CM3	1	121 to 305	1,283,469	0.01943932
CM4	2	-15 to 30	989,525	0.02871001
CM5	2	31 to 305	885,345	0.04020668
CM6	3	-15 to 30	632,262	0.03633787
CM7	3	31 to 305	543,408	0.04852926

Genotyping, linkage maps and phase inference Two-thousand-four-hundred and eighty paternal half-sib sires and 109 founding sires were genotyped with the bovine Affymetrix 25K MIP array (Matukumalli *et al.* 2009). A SNP filtering process considering discordants (<2.5%), minor-allele frequency (>0.025) and genotyped percentage (>75%) for each SNP was performed, which reduced the number of usable assays to 17,483. These SNPs were used to construct male linkage maps for the 29 bovine autosomal chromosomes that in turn were used to assist assembly of the bovine genome sequence (Liu *et al.* 2009). Locally developed software was further used to determine phased chromosomes and impute missing genotypes. This software used information generated from a modified version of the CRI-MAP 2.4 (Green 1990) and PHASE (Stephens *et al.* 2001) programs. Final genetic maps contained 17,347 SNPs.

Genome-wide association study GWAS for CM and SCS were performed to estimate marker effects. Genotype and phenotype information for 2,086 sires for CM and for 2,118 sires for SCS were included in the analysis. Data for both CM and SCS were divided into two datasets which were analysed independently. This division was done by listing the families by grandsire identity number, subscribing every second family to subset 1 and remaining families to subset 2 for each of the two traits. GWAS was performed on the two datasets and on the combined dataset for all 17,347 SNPs for both CM and SCS. The mixed model was:

$$P_i = Xg_j + Ya_i + Zm_k + e_{ijk}$$

Here phenotypic value P is DYD of sire i weighted by number of daughters, g is fixed effect of grandsire j, a i s random effect of sire i where co-variance structure between sires is determined from pedigree relationships, m is random effect of genetic marker k and e is an error term. The polygenic component (a) was fitted to remove the effect of population stratification, for example due to large half sib families. MacLeod *et al.* (2010) demonstrated that including effect of sire based on pedigree relationships reduces the number of false positives due to population stratification in a genome scan.

Combined linkage disequilibrium and linkage analysis Linkage disequilibrium linkage analysis (LDLA) implements historic recombination events in addition to recombination events within genotyped families to estimate haplotype effects. Phased chromosomes from all 2,589 genotyped sires were included in the analysis along with phenotype information for 2,086 sires for CM. The analysis was performed using the GridQTL LDLA software in April 2009 (Hernandez-Sanchez *et al.* 2009). Effective male and female population sizes were set to 200 and 10,000 respectively. Number of discrete generations since population foundation was set to 100 (e.g. Meuwissen and Goddard 2001). IBD probabilities were calculated based on 5-marker haplotypes for positions separated by 1Mb intervals along each chromosome. The mixed model was:

$$P_i = Xb_{ij} + Zh_k + e_{ijk}$$

Here phenotypic value P is DYD of sire i, b contains fixed effects for sire i and grandsire j, h is random effect of haplotype k and e is an error term. For further details see Meuwissen and Goddard (2001).

SNP effects For the detected QTLs the effect of alleles and genotypes of the most significant SNP on DYD for CM was calculated. A model including the same fixed effects as those used for the association analyses was used. In addition genotype was included as a fixed effect. Effects were found by dividing predicted value relative to a mean of zero by DYD standard deviation, giving predicted standard deviations from mean DYD. Standard deviations for CM DYDs in the seven time periods are given in Table 1.

Test for multiple QTLs Multiple QTL analysis was performed to find out if any of the QTL regions for CM contained more than one QTL. SNPs showing the highest test score within each QTL region were modelled as fixed effects to see if this influenced the test scores of the other markers in the region. The mixed model was:

$$P_i = Vs + Xg_j + Ya_i + Zm_k + e_{ijk}$$

Here s is fixed effect of the SNP with the strongest trait-marker association in the QTL region. Remaining terms are described above for the GWAS.

Test score The likelihood ratio test (LRT) was used for hypothesis testing. LRT scores were calculated as two times the log-likelihood (LogL) ratio. LogL ratio values were obtained with the ASREML software (Gilmour 2000) for each SNP as the difference between the log-likelihood of a model containing the SNP- or haplotype effect and the log-likelihood of a model not containing this effect. LRT scores were expected to be distributed as a mixture of two χ^2 distributions with 0 and 1 degree of freedom. For GWAS on the two independent datasets a LRT score larger than 2.7 (p-value ≤ 0.05) in both datasets was considered a significant trait-marker association, while for GWAS on the combined dataset a LRT score larger than 13.81 (p-value ≤ 0.0001) was considered significant trait-marker association. A logarithm of odds (LOD) score >3, corresponding to a LRT score > 13.81, is an indication of genome-wide significance (Lander & Botstein 1989). The GridQTL LDLA software analysis was expected to give higher on-average LRT scores than single-marker association tests (Hernandez-Sanchez *et al.* 2009). An association was considered confirmed by LDLA if a LRT score above 20 from the LDLA analysis was found for a position within 10Mb of a significant

association from GWAS. The LOD drop-off method (Lander & Botstein 1989) was used to find approximate confidence intervals (CIs) from LDLA LRT scores for confirmed QTL regions. The CIs were defined as including positions within 4.6 LRT of the maximum on both sides of the LRT peak for LDLA, yielding an approximate 96.8% CI.

Results

GWAS for CM and SCS were conducted on two independent datasets as well as a combined dataset. Significant trait-marker associations in both of the two independent datasets for CM was found for 10 SNPs positioned on chromosomes 2, 4, 6, 9, 17 and 20, and for SCS for 4 SNPs positioned on chromosomes 12, 19 and 26. Significant trait-marker associations in the combined dataset for CM were found for 26 SNPs positioned on 10 chromosomes and for SCS for 11 SNPs positioned on 6 chromosomes. A summary of GWAS results are given in Table 2a for CM and in Table 2b for SCS. The analysis revealed no SNPs showing consistently significant trait-marker associations for both CM and lactation average SCS. Correlations between SNP effects on SCS and occurrence of CM in the seven lactational time periods based on all 17,347 SNPs are presented in Table 3. Higher correlations were found between SNP effects on SCS and CM in late lactation than between SNP effects on SCS and CM in the periparturient period. Further, stronger correlations were found between SNP effects on CM in the same phase of lactation than between SNP effects on CM in different phases of lactation. Since the trait of interest in this study was susceptibility to mastitis only putative QTLs for CM were investigated further.

LDLA for CM was performed for all chromosome-trait combinations giving significant associations by GWAS in both the two independent datasets, and for the ten chromosome-trait combinations giving strongest trait-marker associations by GWAS in the combined dataset. LDLA analysis was conducted on BTAs 2, 4, 6 and 20 for CM in time period 1, BTA14 for CM in time period 2, BTA2 for CM in time period 3, BTA6 for CM in time period 4, BTAs 7, 9 and 17 for CM in time period 5 and BTA6 for CM in time period 6. All LDLA analyses were performed on the total dataset of 2,589 paternal half-sib sires from 109 families. The GridQTL software was not able to esitimate haplotype effects for all positions, assumedly due to convergence issues for the remaining positions. For three of the chromosome-trait combinations (BTA6 for time period 4 and 6, and BTA9 for time period 5) the software was only able to estimate haplotype effects for a few positions. Putative QTLs identified by GWAS on the two independent datasets and on the combined dataset on chromosomes 2, 6 and 20 for CM in the periparturient period of first lactation (CM1) were confirmed by LDLA. A putative QTL only identified by GWAS on the combined dataset on BTA14 for CM in late first lactation (CM2) was also confirmed by LDLA. Results of LDLA and GWAS on the combined dataset for these QTLs are presented in Figure 1. Highest LRT scores from LDLA were found for positions on BTA2 at 104Mb, on BTA6 at 95Mb, on BTA14 at 42Mb and on BTA20 at 43Mb. Approximate 96.8% CIs for the four QTL regions based on the LOD drop-off method (Lander & Botstein 1989) included the regions 103.4-104.3Mb on BTA2, 94.2-95.3Mb on BTA6, 41.3-42.3Mb on BTA14 and 41.8-43.2Mb on BTA20. Highest LRT scores from GWAS on the combined dataset for each of the four QTLs were found for SNP BTA-120624 at 103.9Mb on BTA2, for SNP BTA-119376 at 90.7Mb on BTA6, for SNP BTA-34923 at 45.2Mb on BTA14 and for SNP BTA-19985 at 43.3 on BTA20.

Table 2a

Significant trait-marker associations from GWAS on two independent datasets and on the combined dataset for CM on the 29 bovine autosomal chromosomes. BTA, SNP, position, LRT score for datset 1 (D1), dataset 2 (D2) and the combined dataset (CD), and time period (TP) are given.

BTA	SNP	Position (bp)	D1 LRT	D2 LRT	CD LRT	TP
2	BTA-47902	68,074,185	4.92	3.26	15.5	CM3
2	BTA-120621	103,854,622	3.29	3.05	12.72	CM1
2	BTA-120624	103,892,096	5.03	3.26	15.52	CM1
2	rs29025784	112,396,633	4.06	4.1	14.62	CM1
4	rs29020694	90,418,076	5.65	3.23	20.14	CM1
6	BTA-119376	90,670,190	3.28	12.74	32.36	CM1
6	BTA-119376	90,670,190	2.76	3.16	15.46	CM6
6	BTA-77136	94,544,954			18.8	CM4
6	BTA-109071	95,256,811			14.18	CM6
6	BTA-77356	96,189,520	3.08	5.18	15.6	CM1
7	BTA-78563	22,841,729			16.92	CM5
7	BTA-99486	26,347,331			16.8	CM5
9	BTA-84619	88,416,414	2.95	2.71		CM5
10	BTA-79349	92,455,424			14.2	CM6
13	rs29022774	62,928,524			13.9	CM4
13	rs29022775	62,928,533			15.9	CM4
14	rs29012803	17,289,087			14.9	CM5
14	BTA-34796	40,769,096			15.36	CM2
14	BTA-34923	45,153,040			18.34	CM2
14	BTA-111421	47,425,522			15.96	CM2
16	BTA-38543	33,335,777			14.64	CM1
17	BTA-103789	34,861,876	3.02	2.76	12.84	CM5
20	BTA-25160	31,659,731			15.16	CM2
20	BTA-50239	35,530,051	2.96	3.84	16.02	CM1
20	BTA-50236	35,861,339			17.1	CM1
20	rs29021255	38,261,470			14.88	CM3
20	BTA-19985	43,267,496			20.76	CM1
20	BTA-22852	50,780,080			14.36	CM5
29	rs29027496	45,602,144			14.36	CM1

Table 2b

Significant trait-marker associations from GWAS on two independent datasets and on the combined dataset for SCS on the 29 bovine autosomal chromosomes. BTA, SNP, position and LRT score for datset 1 (D1), dataset 2 (D2) and the combined dataset (CD) are given.

BTA	SNP	Position (bp)	D1 LRT	D2 LRT	CD LRT
8	BTA-120681	27,391,619			14
8	BTA-102648	30,020,795			14.34
12	rs29021760	63,796,167			14.46
12	BTA-28030	66,114,084	2.75	3.16	13.76
17	rs29019473	52,225,317			13.98
17	rs29019471	52,229,956			14.26
19	ss46526232	42,098,180			16.2
19	BTA-45702	50,276,021	4.28	3.73	14.86
19	BTA-45709	50,359,720	5.44	3.64	15.66
20	BTA-95391	11,740,236			14.08
21	rs29016404	33,141,160			25.12
21	BTA-52383	46,299,936			14.06
26	BTA-61841	46,096,255	2.88	2.98	

Table 3

Correlations between SNP effects on SCS and CM in the seven lactational time periods CM1, CM2, CM3, CM4, CM5, CM6 and CM7 (the seven lactational time periods are described in Table 1).

	CM1	CM2	CM3	CM4	CM5	CM6	CM7
SCS	0.04	0.11	0.11	0.08	0.15	0.10	0.12
CM1		0.14	0.08	0.34	0.15	0.22	0.07
CM2			0.30	0.16	0.33	0.11	0.21
CM3				0.10	0.31	0.07	0.17
CM4					0.13	0.33	0.09
CM5						0.10	0.31
CM6							0.09

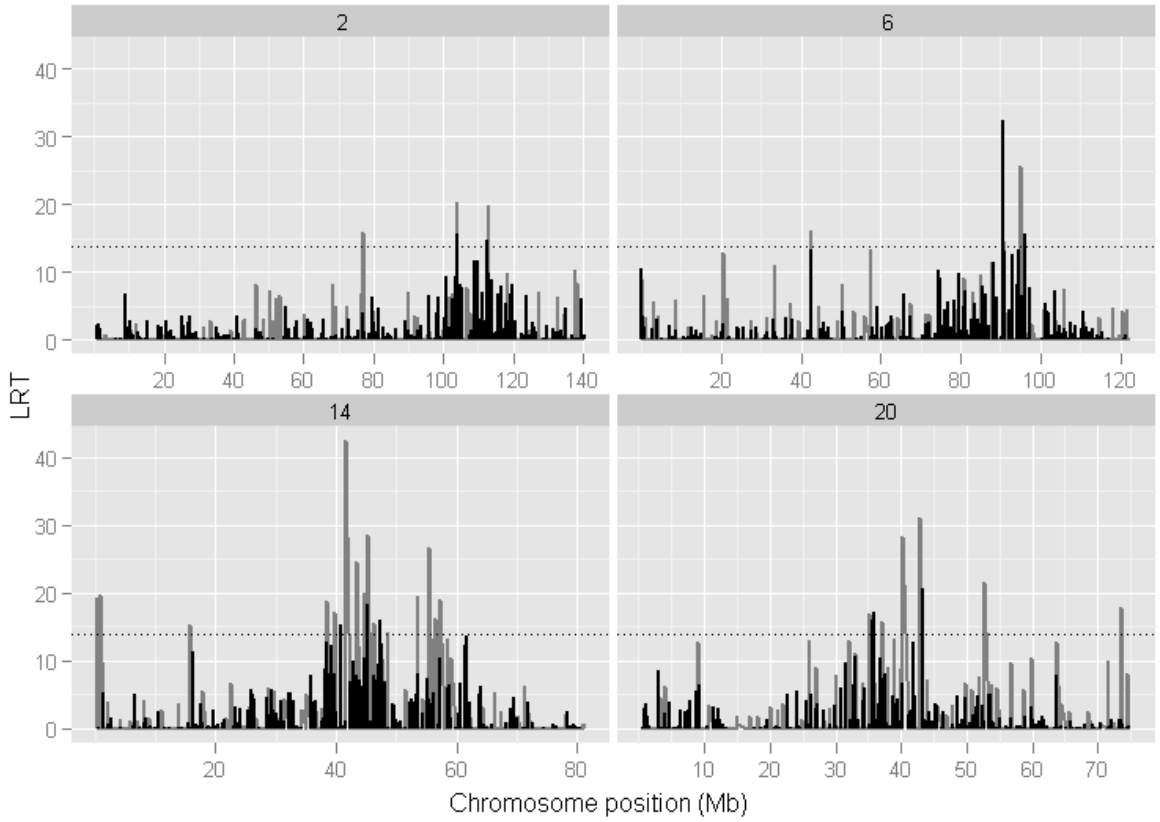


Figure 1
Results from GWAS and LDLA for BTA2 for CM1, BTA6 for CM1, BTA14 for CM2 and BTA20 for CM1. Black bars show LRT score from GWAS, grey lines show LRT scores from LDLA. Dotted horizontal line indicate LRT threshold for GWAS (LRT > 13.81).

For each of these four QTLs effects of alleles and genotypes for the most significant SNP by GWAS for CM are shown in Table 4. Results are in DYD standard deviations from the mean DYD for CM. Difference in effect on DYD between alleles was largest for SNP BTA-119376 on BTA6 for CM in the periparturient period of first lactation. This SNP also gave the highest test-score from GWAS on the combined dataset for CM among all 17,347 SNPs. For SNP BTA-34923 on BTA14 and SNP BTA-19985 on BTA20 low minor-allele frequencies contributed to large standard deviation relative to effect sizes, resulting in more uncertain effect estimates. All the four QTL regions contained several SNPs significantly associated with occurrence of CM (see Table 2a). Therefore, tests for multiple QTLs were performed on the combined dataset to find out if any of these regions contained more than one QTL. For all four QTL regions LRT scores of the remaining SNPs were lowered to non-significant levels (<13.81) when the SNP with the highest LRT score in each of the regions was included in the model as a fixed effect. Highest test scores found in each region by the multiple QTL test were for SNP BTA-122697 on BTA2 at 112.8Mb with LRT score 5.16, SNP BTA-86975 on BTA6 at 88.4Mb with LRT score 4.82, SNP rs29018717 on BTA14 at 61.3Mb with LRT score 6.66 and SNP BTA-50239 on BTA20 at 35.5Mb with LRT score 9.04.

Table 4

Effects of SNPs on DYD for CM for the most significant SNP in each of the four QTL regions. BTA, SNP, alleles and genotypes, frequencies of alleles and genotypes, effect of alleles and genotypes, standard deviation (SD) of effect and time period (TP) is given. Effects are expressed as DYD standard deviations from the mean value. Standard deviation for DYD for clinical mastitis in the seven time periods are given in Table 1.

BTA	SNP	Alleles	Frequency	Effect	SD (Effect)	TP
2	BTA-120624	A	0.50	-0.09	0.04	CM1
		G	0.50	0.06	0.04	
		AA	0.25	-0.17	0.05	
		AG	0.51	-0.02	0.04	
		GG	0.24	0.15	0.05	
6	BTA-119376	C	0.26	0.20	0.05	CM1
		T	0.74	-0.01	0.03	
		CC	0.07	0.36	0.09	
		CT	0.38	0.09	0.04	
		TT	0.55	-0.11	0.04	
14	BTA-34923	C	0.19	-0.03	0.03	CM2
		T	0.81	-0.04	0.03	
		CC	0.03	-0.05	0.17	
		CT	0.31	0.01	0.04	
		TT	0.66	-0.06	0.03	
20	BTA-19985	C	0.19	-0.01	0.03	CM1
		G	0.81	-0.01	0.03	
		CC	0.04	-0.05	0.12	
		CG	0.31	-0.04	0.05	
		GG	0.66	0.00	0.04	

Discussion

GWAS performed on two independent datasets identified 10 SNPs positioned on chromosomes 2, 4, 6, 9, 17 and 20 that were significantly associated with occurrence of CM (LRT > 2.7 in both datasets). GWAS on the combined dataset identified 26 SNPs positioned on 10 chromosomes significantly associated with occurrence of CM (LRT > 13.81). Three QTLs on BTAs 2, 6 and 20 for CM in the periparturient period and a QTL on BTA14 for CM in late lactation were confirmed by LDLA. The QTL on BTA14 was detected by GWAS on the combined dataset but not by GWAS on the two independent datasets. This could be an indication that the requirement for associations to be significant in two independent datasets (LRT>2.7) is a more conservative test than the higher significance threshold value (LRT>13.81) used for the combined dataset. Inability to detect a QTL in a split-dataset analysis could also be caused by lowered power due to fewer observations in each dataset. Highest test scores from GWAS on the combined dataset for each of the four QTLs were found for SNP BTA-120624 at 103.9Mb on BTA2, for SNP BTA-119376 at 90.7Mb on BTA6, for SNP BTA-34923 at 45.2Mb on BTA14 and for SNP BTA-19985 at 43.3 on BTA20. Approximate

96.8% CIs were found by the LOD drop-off method for all four QTLs based on LDLA test scores. The CI on BTA2 included the SNP with the highest test score from GWAS in this region. The CIs on BTA6, BTA14 and BTA20 did not include the SNPs with the highest test score from GWAS from each of these regions. For LDLA LRT scores were only attempted estimated for every 1Mb across each chromosome and unfortunately the GridQTL software was not able to estimate haplotype effects for all positions. Construction of CIs based on LDLA analysis might therefore be an inaccurate approach for this study.
Multiple QTL analysis did not provide evidence of more than one QTL in any of the four regions. However, a relatively high LRT score of 9.04 was found for a SNP at 35.5Mb on BTA20 when including the SNP at 43.27Mb as a fixed effect. This LRT score is below our threshold for GWAS on the combined dataset, but presence of an additional QTL in this region cannot be completely ruled out.
While the QTL on BTA20 has not previously been reported, the QTLs on BTAs 2, 6 and 14 are supported by other studies. A QTL for SCS in German Holsteins has been reported on BTA2 at 100cM (Bennewitz *et al.* 2003) close to our most significant SNP for CM on BTA2 at 103.9Mb. This QTL was not found to affect SCS in our study. For NRF records for SCS are retrieved as lactation averages and are thus not directly comparable with the CM recordings. This may reduce power for QTL detection for SCS. Our analyses showed no consistencies between QTLs for CM and for SCS. Similar results have been reported for the NRF population in a previous study (Klungland *et al.* 2001).
Correlation in SNP effects based on all 17,347 SNPs were higher between SCS and CM in late lactation than between SCS and CM in the periparturient period, supporting a previous report on genetic correlations between SCS and CM in different stages of lactation (Svendsen & Heringstad 2006b).
QTLs affecting both milk production traits and CM around 90Mb on BTA6 have previously been reported for NRF (Nilsen *et al.* 2009) in a study including the same population and many of the same individuals as were included in the GWAS described here. This is in the same region as the QTL detected at approximately 90Mb on BTA6 in this study. QTLs affecting susceptibility to CM on BTA14, close to our most significant SNP on BTA14 at 45.2Mb, have been reported by both Lund *et al.* (2007) and Schulman *et al.* (2004). More milk production trait QTLs have been reported for chromosomes 6, 14 and 20 than for the other bovine chromosomes (Khatkar *et al.* 2004).
Genetic susceptibility to CM could have different biological causes at different stages of lactation (Waller 2000; Sordillo 2005; Østeras 2006), which could reduce power to detect QTLs affecting susceptibility to CM only in a particular phase of lactation. In this study phenotypic records on occurrence of CM were divided into seven time periods (Table 1), consequently treating occurrence of CM at different stages of lactation as different traits. There are more records for CM incidents in the periparturient period as approximately two thirds of incidents occur in the two first months after lactation. Consequently this study had more power to detect QTLs for CM in the periparturient period than for CM in later lactation. The study also had more power to detect QTLs for CM in first lactation than in later lactations. Three of the four QTLs detected here were for CM in the periparturient period of first lactation. Correlations based on all 17,347 SNPs were higher between SNP effects on CM in the same phase of lactation than between SNP effects on CM in different phases of lactation. This is in accordance with a report on genetic correlation between CM in different phases of lactation (Svendsen & Heringstad 2006a), and supports division of records on CM

into lactational time periods.
The strongest trait-marker association by GWAS on the combined dataset was between SNP BTA-119376 at 90.7Mb on BTA6 and CM in the periparturient period of first lactation (CM1). Genes coding for most of the C-X-C motif chemokines are clustered near this SNP on BTA6. In particular, the gene coding for interleukin 8 (IL8) is positioned on BTA6 at 91.78Mb. IL8 is a C-X-C motif chemokine important for initial recruitment of circulating neutrophils to the site of infection. In NRF *Staphylococcus aureus* is the most common cause of CM (Østeras 2006). Gram positive bacteria such as *S. aureus* are assumed to initiate expression of IL8 by toll-like receptor 2 activation by cell wall components lipoteichoic acid or peptidoglycan (Bannerman *et al.* 2004). There has been inconsistent reports on the ability of different cattle breeds to express IL8 upon exposure to *S. aureus* (Bannerman *et al.* 2004; Strandberg *et al.* 2005; Lahouassa *et al.* 2007; Griesbeck-Zilch *et al.* 2008; Yang *et al.* 2008). These inconsistencies might be due to genetic differences between cattle or due to different strains of *S. aureus* being investigated in the different studies.
Candidate genes associated with immunological defence are also found close to other QTLs. On BTA2 genes coding for IL8 receptors chemokine (C-X-C motif) receptor 2 (CXCR2) and chemokine (C-X-C motif) receptor 1 (CXCR1) are both positioned around 110.6Mb, between SNPs BTA-120624 and rs29025784 at 103.9Mb and 112.4Mb. Both these SNP showed significant association with CM in the periparturient period of first lactation (CM1), with LRT values 15.52 and 14.62. Polymorphisms in *CXCR2* have been associated with occurrence of subclinical mastitis and ability to recruit neutrophils to the site of infection (Youngerman *et al.* 2004; Rambeaud & Pighetti 2005). Polymorphisms in *CXCR1* have been associated with SCS level and expression level of CXCR1 (Leyva-Baca *et al.* 2008a; Leyva-Baca *et al.* 2008b). The IL8 receptors have other C-X-C motif chemokines as ligands in addition to IL8. As mentioned above, the QTL on BTA2 was not found to affect lactation average SCS level in this study.
Three genes coding for complement components 6, 7 and 9 (C6, C7 and C9) are positioned at 34,4Mb, 35,7Mb and 37,3Mb on BTA20. By GWAS four SNPs in the interval 35.5 to 45.6Mb on this chromosome showed significant association with CM in the periparturient period. Complement components are an important part of immunological defence. They are involved in inflammation, phagocytosis, attack on bacterial membranes and promotion of antibody production. Although the mentioned genes could potentially have an effect on susceptibility to CM in NRF, the number of other potential candidate genes in these regions is quite high. It may be of particular interest from GWAS for CM that our two most significant SNPs on BTA2 lie near the genes coding for CXCR1 and CXCR2, while our most significant SNP on BTA6 lie near the gene coding for IL8.

Acknowledgements
Thanks to H. Henriksen and T. Nome for technical assistance, and to GENO Breeding and AI association for providing relationship information and DYD values. This project has been funded by The Research Council of Norway, GENO Breeding and AI association and BoviBank Ltd.

References

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. *Clin Diagn Lab Immunol* **11**, 463-72.

Bennewitz J., Reinsch N., Grohs C., Leveziel H., Malafosse A., Thomsen H., Xu N., Looft C., Kuhn C., Brockmann G.A., Schwerin M., Weimann C., Hiendleder S., Erhardt G., Medjugorac I., Russ I., Forster M., Brenig B., Reinhardt F., Reents R., Averdunk G., Blumel J., Boichard D. & Kalm E. (2003) Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genet Sel Evol* **35**, 319-38.

Gibbs R.A., Taylor J.F., Van Tassell C.P., Barendse W., Eversole K.A., Gill C.A., Green R.D., Hamernik D.L., Kappes S.M., Lien S., Matukumalli L.K., McEwan J.C., Nazareth L.V., Schnabel R.D., Weinstock G.M., Wheeler D.A., Ajmone-Marsan P., Boettcher P.J., Caetano A.R., Garcia J.F., Hanotte O., Mariani P., Skow L.C., Sonstegard T.S., Williams J.L., Diallo B., Hailemariam L., Martinez M.L., Morris C.A., Silva L.O., Spelman R.J., Mulatu W., Zhao K., Abbey C.A., Agaba M., Araujo F.R., Bunch R.J., Burton J., Gorni C., Olivier H., Harrison B.E., Luff B., Machado M.A., Mwakaya J., Plastow G., Sim W., Smith T., Thomas M.B., Valentini A., Williams P., Womack J., Woolliams J.A., Liu Y., Qin X., Worley K.C., Gao C., Jiang H., Moore S.S., Ren Y., Song X.Z., Bustamante C.D., Hernandez R.D., Muzny D.M., Patil S., San Lucas A., Fu Q., Kent M.P., Vega R., Matukumalli A., McWilliam S., Sclep G., Bryc K., Choi J., Gao H., Grefenstette J.J., Murdoch B., Stella A., Villa-Angulo R., Wright M., Aerts J., Jann O., Negrini R., Goddard M.E., Hayes B.J., Bradley D.G., Barbosa da Silva M., Lau L.P., Liu G.E., Lynn D.J., Panzitta F. & Dodds K.G. (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528-32.

Gilmour A.R., B. R Cullis, S.J. Welham, and R. Thompson. (2000) ASREML reference manual.

Green P., K. Falls and S. Crooks. (1990) Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St. Louis, Mo., USA.

Griesbeck-Zilch B., Meyer H.H., Kuhn C.H., Schwerin M. & Wellnitz O. (2008) Staphylococcus aureus and Escherichia coli cause deviating expression profiles of cytokines and lactoferrin messenger ribonucleic acid in mammary epithelial cells. *J Dairy Sci* **91**, 2215-24.

Heringstad B., Chang Y.M., Gianola D. & Klemetsdal G. (2005) Genetic association between susceptibility to clinical mastitis and protein yield in norwegian dairy cattle. *J Dairy Sci* **88**, 1509-14.

Heringstad B., Klemetsdal G. & Ruane J. (2000) Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries *Livestock Production Science* **64**, 95-106.

Hernandez-Sanchez J., Grunchec J.A. & Knott S. (2009) A web application to perform linkage disequilibrium and linkage analyses on a computational grid. *Bioinformatics* **25**, 1377-83.

Khatkar M.S., Thomson P.C., Tammen I. & Raadsma H.W. (2004) Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet Sel Evol* **36**, 163-90.

Klungland H., Sabry A., Heringstad B., Olsen H.G., Gomez-Raya L., Vage D.I., Olsaker I., Odegard J., Klemetsdal G., Schulman N., Vilkki J., Ruane J., Aasland M., Ronningen K. & Lien S. (2001) Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mamm Genome* **12**, 837-42.

Kuhn C., Bennewitz J., Reinsch N., Xu N., Thomsen H., Looft C., Brockmann G.A., Schwerin M., Weimann C., Hiendleder S., Erhardt G., Medjugorac I., Forster M., Brenig B., Reinhardt F., Reents R., Russ I., Averdunk G., Blumel J. & Kalm E. (2003) Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J Dairy Sci* **86**, 360-8.

Lahouassa H., Moussay E., Rainard P. & Riollet C. (2007) Differential cytokine and chemokine responses of bovine mammary epithelial cells to Staphylococcus aureus and Escherichia coli. *Cytokine* **38**, 12-21.

Lander E.S. & Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-99.

Leyva-Baca I., Pighetti G. & Karrow N.A. (2008a) Genotype-specific IL8RA gene expression in bovine neutrophils in response to Escherichia coli lipopolysaccharide challenge. *Anim Genet* **39**, 298-300.

Leyva-Baca I., Schenkel F., Martin J. & Karrow N.A. (2008b) Polymorphisms in the 5' upstream region of the CXCR1 chemokine receptor gene, and their association with somatic cell score in Holstein cattle in Canada. *J Dairy Sci* **91**, 407-17.

Leyva-Baca I., Schenkel F., Sharma B.S., Jansen G.B. & Karrow N.A. (2007) Identification of single nucleotide polymorphisms in the bovine CCL2, IL8, CCR2 and IL8RA genes and their association with health and production in Canadian Holsteins. *Anim Genet* **38**, 198-202.

Liu Y., Qin X., Song X.Z., Jiang H., Shen Y., Durbin K.J., Lien S., Kent M.P., Sodeland M., Ren Y., Zhang L., Sodergren E., Havlak P., Worley K.C., Weinstock G.M. & Gibbs R.A. (2009) Bos taurus genome assembly. *BMC Genomics* **10**, 180.

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

MacLeod I.M., Hayes B.J., Savin K.W., Chamberlain A.J., McPartlan H.C. & Goddard M.E. (2010) Power of a genome scan to detect and locate quantitative trait loci in cattle using dense single nucleotide polymorphisms. *J Anim Breed Genet* **127**, 133-42.

Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F., Heaton M.P., O'Connell J., Moore S.S., Smith T.P., Sonstegard T.S. & Van Tassell C.P. (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* **4**, e5350.

Meuwissen T.H. & Goddard M.E. (2001) Prediction of identity by descent probabilities from marker-haplotypes. *Genet Sel Evol* **33**, 605-34.

Mrode R.A. & 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. *Animal Breeding Abstracts* **64**, 847–57.

Nilsen H., Olsen H.G., Hayes B., Nome T., Sehested E., Svendsen M., Meuwissen T.H. & Lien S. (2009) Characterization of a QTL region affecting clinical mastitis and protein yield on BTA6. *Anim Genet* **40**, 701-12.

Rambeaud M. & Pighetti G.M. (2005) Impaired neutrophil migration associated with specific bovine CXCR2 genotypes. *Infect Immun* **73**, 4955-9.

Sahana G., Lund M.S., Andersson-Eklund L., Hastings N., Fernandez A., Iso-Touru T., Thomsen B., Viitala S., Sorensen P., Williams J.L. & Vilkki J. (2008) Fine-mapping QTL for mastitis resistance on BTA9 in three Nordic red cattle breeds. *Anim Genet* **39**, 354-62.

Schulman N.F., Viitala S.M., de Koning D.J., Virta J., Maki-Tanila A. & Vilkki J.H. (2004) Quantitative trait Loci for health traits in Finnish Ayrshire cattle. *J Dairy Sci* **87**, 443-9.

Sordillo L.M. (2005) Factors affecting mammary gland immunity and mastitis susceptibility. *Livestock Production Science* **98**, 89-99.

Stephens M., Smith N.J. & Donnelly P. (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89.

Strandberg Y., Gray C., Vuocolo T., Donaldson L., Broadway M. & Tellam R. (2005) Lipopolysaccharide and lipoteichoic acid induce different innate immune responses in bovine mammary epithelial cells. *Cytokine* **31**, 72-86.

Svendsen M. & Heringstad B. (2006a) New Genetic Evaluation for Clinical Mastitis in Multiparous Norwegian Red Cows. *Interbull Bulletine* **35**, 8-11.

Svendsen M. & Heringstad B. (2006b) Somatic Cell Count as an Indicator of Subclinical Mastitis. Genetic Parameters and Correlations with Clinical Mastitis. *Interbull Bulletine* **35**, 12-6.

Syvajarvi J., Saloniemi H. & Grohn Y. (1986) An epidemiological and genetic study on registered diseases in Finnish Ayrshire cattle. IV. Clinical mastitis. *Acta Vet Scand* **27**, 223-34.

Viguier C., Arora S., Gilmartin N., Welbeck K. & O'Kennedy R. (2009) Mastitis detection: current trends and future perspectives. *Trends Biotechnol* **27**, 486-93.

Waller K.P. (2000) Mammary gland immunology around parturition. Influence of stress, nutrition and genetics. *Adv Exp Med Biol* **480**, 231-45.

Weller J.I., Kashi Y. & Soller M. (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci* **73**, 2525-37.

Yang W., Zerbe H., Petzl W., Brunner R.M., Gunther J., Draing C., von Aulock S., Schuberth H.J. & Seyfert H.M. (2008) Bovine TLR2 and TLR4 properly transduce signals from Staphylococcus aureus and E. coli, but S. aureus fails to both activate NF-kappaB in mammary epithelial cells and to quickly induce TNFalpha and interleukin-8 (CXCL8) expression in the udder. *Mol Immunol* **45**, 1385-97.

Youngerman S.M., Saxton A.M., Oliver S.P. & Pighetti G.M. (2004) Association of CXCR2 polymorphisms with subclinical and clinical mastitis in dairy cattle. *J Dairy Sci* **87**, 2442-8.

Østeras O., Kruse, H, Sølverød, L, Gjestvang, J, Mørk, T (2006) Nordic View Concerning Mastitis Pathogen Resistance. *Proceedings NMC 45th Annual Meeting. Tampa, Florida.*

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. *J Dairy Sci* **90**, 4483-97.

Paper III

Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle

Marte Sodeland[1], Harald Grove[1, 2], Matthew Kent[1, 2], Simon Taylor[1], Morten Svendsen[3], Ben J. Hayes[2, 4], and Sigbjørn Lien[1, 2]

[1]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway.[2] Centre for Integrative Genetics, Norwegian University of Life Sciences, N-1432 Aas, Norway. [3]Geno Breeding and AI organization, Norwegian University of Life Sciences, Box 5003, N-1432 Aas, Norway.[4] Biosciences Research Division, Department of Primary Industries Victoria, Melbourne, Australia, 3083.

Abstract

Background

Previous fine mapping studies in Norwegian Red cattle (NRF) in the region 86-90.4Mb on *Bos Taurus* chromosome 6 (BTA6) has revealed one quantitative trait locus (QTL) for protein yield (PY) around 88Mb and another for clinical mastitis (CM) around 90Mb. The close proximity of these QTLs may partly explain the unfavorable genetic correlation between these two traits in NRF. A long range haplotype covering this region was introduced into the NRF population through the importation of a Holstein-Friesian bull (1606 Frasse) from Sweden in the 1970s. It has been suggested that this haplotype has a favorable effect on milk protein content but an unfavorable effect on mastitis susceptibility. Selective breeding for milk production traits is likely to have increased the frequency of this haplotype in the NRF population.

Results

Association mapping for PY and CM in NRF was performed using genotypes from 556 SNPs throughout the region 86-97Mb on BTA6 and daughter-yield-deviations from 2601 bulls made available from the Norwegian dairy herd recording system. Highest test scores for PY were found in and around the genes *CSN2* and *CSN1S2*, coding for the β-casein and α_{S2}-casein proteins. High coverage re-sequencing by high throughput sequencing technology enabled molecular characterization of a long range haplotype from 1606 Frasse encompassing these two genes. Haplotype analysis of a large number of descendants from this bull indicated that the haplotype was not markedly disrupted by recombination in this region. The haplotype was associated with both increased milk protein content and increased susceptibility to mastitis, which might explain parts of the observed genetic correlation between PY and CM in NRF. Plausible causal polymorphisms affecting PY were detected in the promoter region and in the 5'-flanking UTR of *CSN1S2*. These polymorphisms could affect transcription or translation of *CSN1S2* and thereby affect the amount of α_{S2}-casein in milk.

Conclusion

Molecular characterization of the long range haplotype from the Holstein-Friesian bull 1606 Frasse, imported into NRF in the 1970s, revealed polymorphisms that could affect transcription or translation of the casein gene *CSN1S2*. Sires with this haplotype had daughters with significantly elevated milk protein content and selection for milk production traits is likely to have increased the frequency of this haplotype in the NRF population. The haplotype was also associated with increased mastitis susceptibility, which might explain parts of the genetic correlation between PY and CM in NRF.

Background

It has been suggested by Lien *et al.* [1] that a haplotype encompassing the casein gene cluster around 88Mb on BTA6, which confers a favorable effect on milk production traits, was introduced into the NRF population through the importation of a Swedish Holstein-Friesian bull (1606 Frasse) in the 1970s. Association mapping has revealed that a QTL for PY coincides with the casein gene cluster [2-6]. Casein proteins constitute approximately 80% of dairy cattle milk protein and polymorphisms in these genes have been shown to contain variation associated with milk protein composition and protein content in other populations [1, 3, 7-12]. An unfavourable genetic correlation between PY and CM has been reported for NRF, with estimates ranging from 0.21 to 0.55 [13]. This genetic correlation could be both due to pleiotropic effects and due to QTLs affecting the two traits being closely positioned on bovine chromosomes. The heritability for PY is estimated to be 0.19 in NRF [13], and is

higher than the heritability for CM for which estimates range between 0.02 and 0.12 in Nordic cattle populations [14-16]. Selective breeding for PY will therefore be more efficient than for CM, and could have the side-effect of increasing the frequency of variants with undesirable effects on CM. It has further been reported that the haplotype from 1606 Frasse encompassing the casein gene cluster is associated with increased mastitis susceptibility, and a QTL for CM in the periparturient period has been found around 90Mb, close to the QTL affecting PY around 88Mb [2]. Taken together these results suggest that causal polymorphisms residing within this genomic region are influencing these two important traits.
Opportunities for fine mapping and molecular characterization of QTL regions have been improved by recent developments in high throughput sequencing and genotyping technologies [17-21], and it is well established that genotyping of related animals with well documented pedigree increases the accuracy of haplotyping and imputation methods [22, 23]. Accuracy of imputation and power of association mapping using imputed genotypes is increased in populations with extensive linkage disequilibrium (LD) [24-27] , meaning that a combination of these approaches is a feasible strategy in cattle populations [28-32].
The aim of this study was to fine map the region on BTA6 containing QTLs for PY and CM, and perform molecular characterization of the 1606 Frasse haplotype by re-sequencing in order to identify plausible causal polymorphisms underlying the two QTLs.

Results and discussion

Re-sequencing in the genomic region between 86 and 97Mb on BTA6 was done by first capturing sequence from seven genomic DNA samples using a Nimblegen sequence capture array, and then sequencing the product on a Roche 454 GS-FLX sequencer [19]. A total of 269 new single-nucleotide polymorphisms (SNPs) were revealed in this region and were genotyped in 768 NRF sires. The resulting dataset was joined with datasets containing previously genotyped SNPs in NRF by haplotyping and imputation of untyped genotypes. Imputation was facilitated both by the elevated LD in NRF and by extensive pedigree records being available [28-32]. Pedigree records improve haplotyping accuracy and thereby improve accuracy of association mapping [22, 23]. The final imputed dataset contained genotypes for 556 SNPs in 2601 NRF sires, with typically only 1.2% of SNPs missing for each individual sire. Average distance between adjacent markers for the 556 SNPs included in this study was approximately 20kb, with some variation in SNP density across the 11Mb genomic region. The 556 SNPs are presented in Additional file 1.

Association mapping

Association mapping was performed to map single SNPs and haplotypes associated with PY or CM in the genomic interval between 86 and 97Mb on BTA6. Three mastitis traits were included in the analyses; incidences of CM in the periparturient period of first (CM1), second (CM2), and third lactation (CM3). Haplotype blocks were defined by the algorithm developed by Gabriel *et al.* [33] (GAB) and by the four-gamete rule algorithm (GAM) described by Wang *et al.* [34], and all haplotype blocks defined by the GAB or the GAM algorithm were included in haplotype association mapping. Results for single-marker association mapping for CM1 and PY for the highest scoring region 88 to 94Mb are shown in Figure 1, whereas single SNPs and haplotype blocks giving highest test scores for each of the four traits are presented in Table 1.
Highest likelihood-ratio test (LRT) scores for PY were found for the SNPs ss86217862 and ss86217864 located at positions 88.410Mb and 88.414Mb. Both SNPs were positioned within the gene *CSN1S2* and in complete LD with each other (r^2=1). The most significant haplotype results for PY were detected for a GAM block in the interval 88.33 to 88.43Mb and a smaller

GAB block lying within this interval (88.33 to 88.42Mb). Both blocks encompassed the genes *CSN2*, *HSTN*, *STATH* and *CSN1S2*. Haplotype analyses for these two blocks did not reveal genome-wide significant test scores for any of the three mastitis traits.

Single-marker association mapping for CM1 and CM3 gave highest test scores for SNP rs42766480 at 90.07Mb, whereas highest test score from haplotype association mapping were found for a three-marker GAM block in the interval 90.64 to 90.67Mb. In contrast to CM1 and CM3, single-marker association mapping for CM2 gave highest test scores for SNP S1_1625793 at 89.63Mb and the most significant haplotype association for CM2 was for a four-marker GAM block in the interval 89.62 to 89.67Mb. For all three mastitis traits highest test scores from both single-marker and haplotype association mapping were found within the interval 89 to 91Mb.

The SNPs that gave highest test scores for the three mastitis traits (rs42766480 and S1_1625793) also gave high test scores for PY (Figure 1). However, SNPs rs42766480 and S1_1625793 were not in LD with the SNPs that gave the highest test scores for PY in this study (ss86217862 and ss86217864).

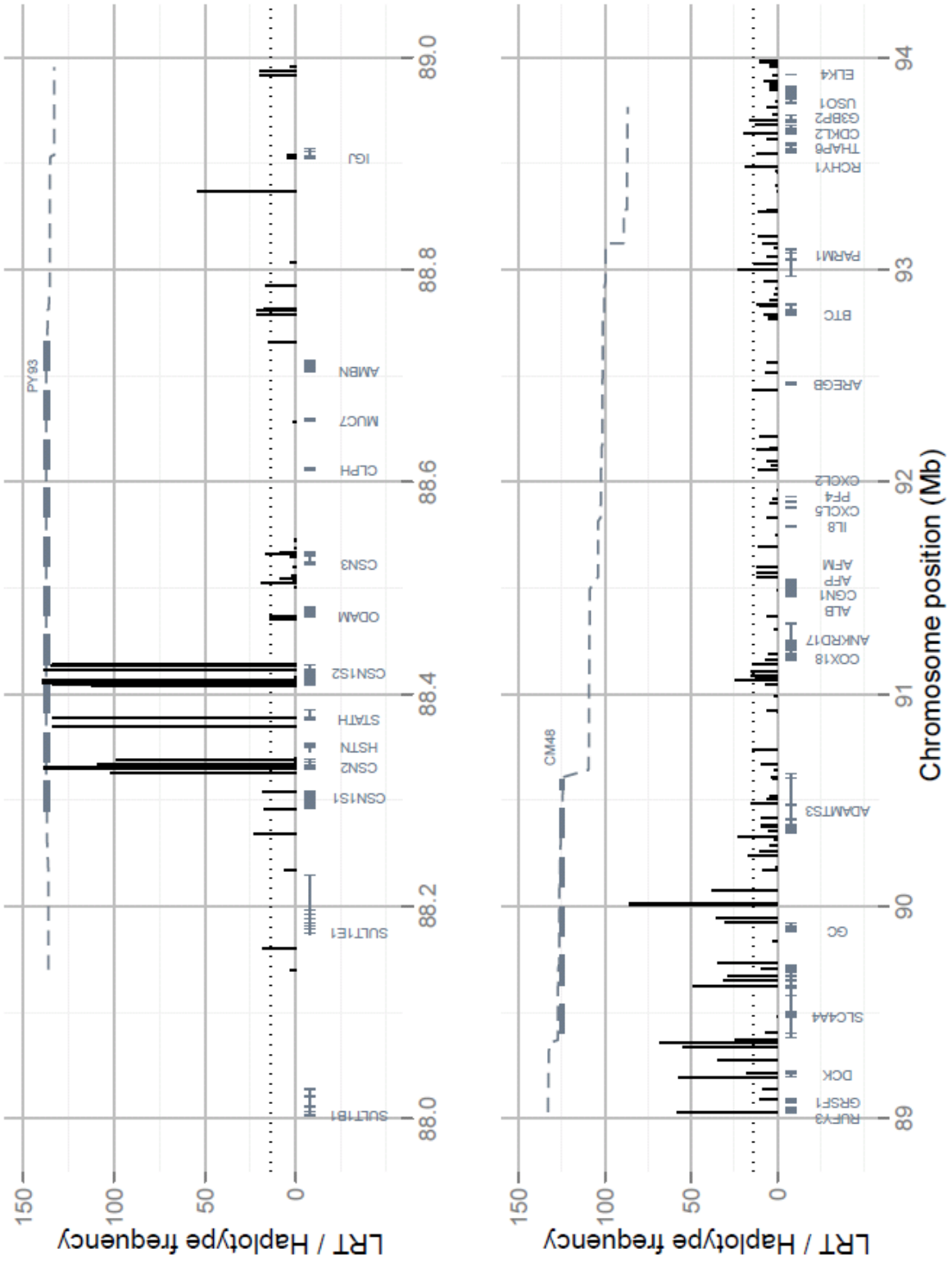


Figure 1a – Single-marker association mapping for protein yield
LRT scores from single-marker association mapping for protein yield (black) in the interval 88 to 89Mb (left) and 89 to 94Mb (right) on BTA6. Genome-wide significance threshold for LRT (>13.81) is indicated (black dotted line). Frequencies (frequency ·1000) of the 1606 Frasse haplotype (grey thin dotted line) extending from the PY93 haplotype window, as well as haplotype windows PY93 and CM48 (grey thick dotted lines), are shown.

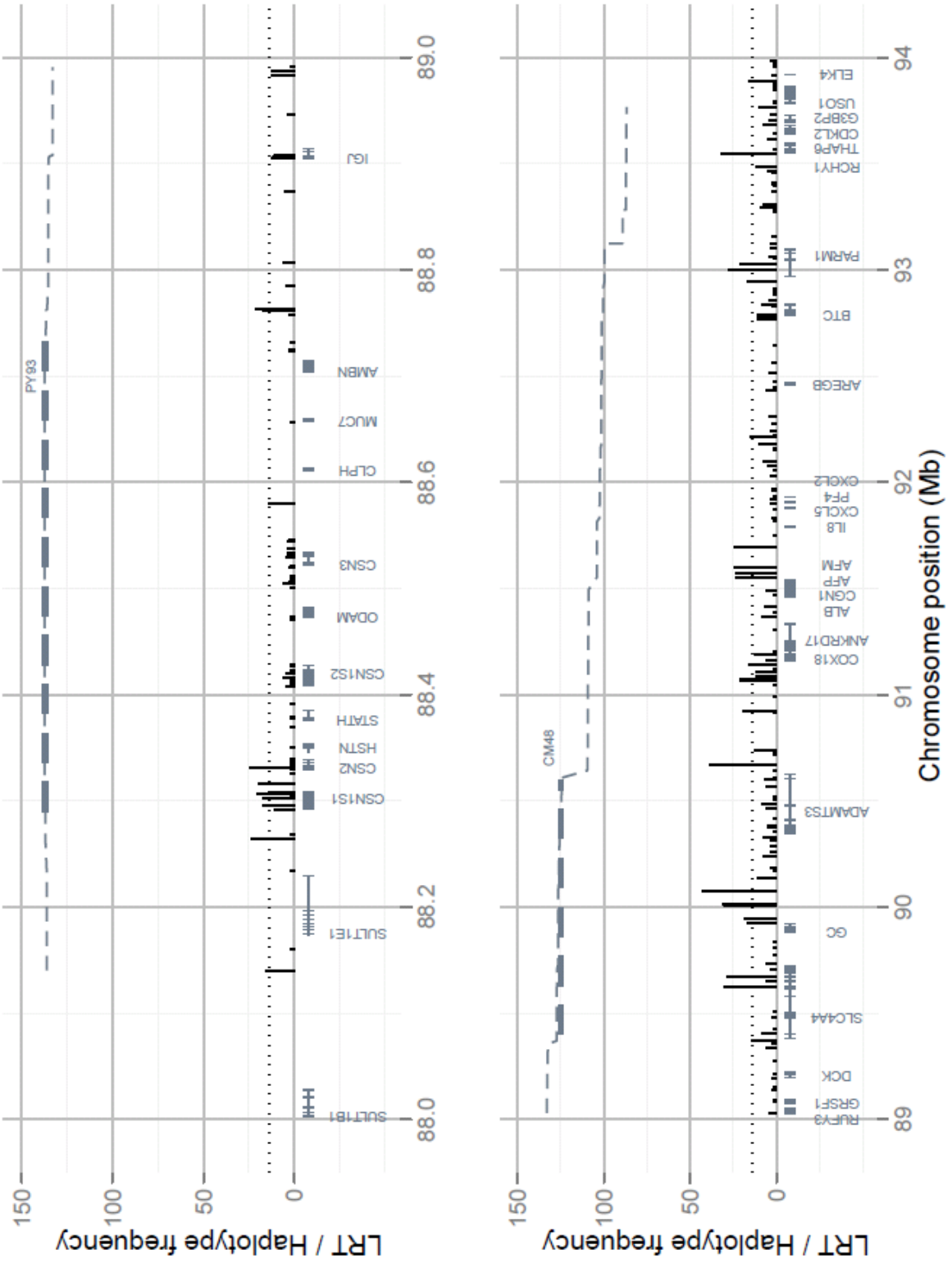


Figure 1b – Single-marker association mapping for clinical mastitis
LRT scores from single-marker association mapping for clinical mastitis in the periparturient period of first lactation (black) in the interval 88 to 89Mb (left) and 89 to 94Mb (right) on BTA6. Genome-wide significance threshold for LRT (>13.81) is indicated (black dotted line). Frequencies (frequency ·1000) of the 1606 Frasse haplotype (grey thin dotted line) extending from the PY93 haplotype window, as well as haplotype windows PY93 and CM48 (grey thick dotted lines), are shown.

Table 1 – Association mapping results
Highest test scores from single-marker (SM) and haplotype (GAB and GAM) association mapping for protein yield (PY) and clinical mastitis in the periparturient period of first (CM1), second (CM2) and third lactation (CM3). Trait, analysis, chromosome position, LRT score and SNPs are presented for the highest scoring SNP or haplotype block from each analysis.

Trait	Analysis	Position (bp)	LRT score	SNPs
CM1	SM	90,075,263	42	rs42766480
CM1	GAB	90,670,190 - 90,725,368	37	ss61522200 to S1_2725368
CM1	GAM	90,642,598 - 90,670,190	42	rs29024027 to ss61522200
CM2	SM	89,625,793	26	S1_1625793
CM2	GAB	89,623,896 - 89,625,793	23	ss61524338 to S1_1625793
CM2	GAM	89,623,896 - 89,668,440	28	ss61524338 to ss86278591
CM3	SM	90,075,263	26	rs42766480
CM3	GAB	90,670,190 - 90,725,368	18	ss61522200 to S1_2725368
CM3	GAM	90,642,598 - 90,670,190	19	rs29024027 to ss61522200
PY	SM	88,410,501	139	ss86217862
PY	SM	88,413,712	139	ss86217864
PY	GAB	88,333,706 - 88,422,590	78	ss86217849 to ss86217869
PY	GAM	88,333,706 - 88,427,761	78	ss86217849 to ss117968525

Table 2 - Haplotype classification for the PY93 haplotype window
Haplotype 1 and haplotype 2 are given for NRF animals 1606 Frasse, 1893 Rud, 2005 Smidesang, 2636 Vik and 3454 J. Steinsvik for the PY93 haplotype window. Population frequencies (PF) were found based on phased chromosomes of 2601 NRF animals. Elevated (E) or reduced (R) protein yield (PY) is indicated when found significantly differing from the population mean (p-value).

Animal	Haplotype 1 (PF)	PY (p-value)	Haplotype 2 (PF)	Effect (p-value)
1606 Frasse	PY93_C (0.137)	E (4.189^{-10})	PY93_C (0.137)	E (4.189^{-10})
1893 Rud	PY93_C (0.137)	E (4.189^{-10})	PY93_D (0.070)	-
2005 Smidesang	PY93_C (0.137)	E (4.189^{-10})	PY93_B (0.146)	R (2.200^{-16})
2636 Vik	PY93_B (0.146)	R (2.200^{-16})	PY93_E (0.025)	-
3454 J. Steinsvik	PY93_A (0.197)	E (2.200^{-16})	PY93_E (0.025)	-

Molecular characterization of a long range haplotype affecting protein yield

Very high LRT scores for PY were found for SNPs in and around the casein genes *CSN2* and *CSN1S2* (Figure 1a). A previous study using the same population also identified this region and postulated that a influential haplotype associated with elevated protein yield was introduced into NRF through importation of the bull 1606 Frasse in the 1970s [3]. To identify possible causal polymorphisms underlying this QTL whole genome re-sequencing was conducted of five elite sires in the NRF population including 1606 Frasse (10x coverage) and two of his sons; 1893 Rud and 2005 Smidesang (both at 4x coverage).
A comparison of the sequence data covering the genes *CSN2* and *CSN1S2* and their 2000bp

5'-flanking promoters with 28 previously genotyped SNPs within these regions showed that only one SNP out of the 28 SNPs was undetected by the re-sequencing. Altogether 93 polymorphisms were detected in the re-sequencing of these two regions, corresponding to one SNP approximately every 360bp.

To be able to group genetically similar sires for the QTL for PY a 93 marker haplotype window (PY93) covering the region from 88.29 to 88.75Mb was defined (Figure 1). Haplotype classification within PY93 was performed based on phased chromosomes for the five re-sequenced sires (Table 2) and other NRF sires for which genotypes were available. Both re-sequencing and haplotype classification indicated that 1606 Frasse was homozygous for the QTL for PY while his sons 1893 Rud and 2005 Smidesang both were heterozygous. Sires with at least one copy of the haplotype PY93_C, likely descendents from 1606 Frasse, had daughters with significantly elevated PY (p-value of 4.189^{-10}).

The re-sequencing allowed for molecular characterization of the long range 1606 Frasse haplotype associated with increased PY. Comparison with other haplotypes revealed 6 polymorphisms in coding regions of casein genes *CSN2* and *CSN1S2* or in their 2000bp 5'-flanking promoter regions (Table 3). Positive and negative alleles were assigned for these polymorphisms, with a positive allele defined as one found in PY93_C and therefore associated with increased PY (Table 3). The only non-synonymous substitution detected within the genes *CSN2* and *CSN1S2* was in amino acid 82 in *CSN2* (Ref NM_181008.2), previously reported in NRF by Nilsen *et al.* [3]. This substitution has been reported to have differing effects on PY in various cattle breeds and is therefore not likely to be a significant causal polymorphism [4, 36-38]. The re-sequencing also detected a silent substitution (C>T) in amino acid 125 of *CSN2* (Ref NM_181008.2), previously reported by Lien *et al.* [1].

Of greater interest, a SNP (A>C) was detected in the promoter region of *CSN1S2* at -7bp relative to the transcription initiation site. The SNP, which has previously been reported by Schild and Geldermann [10], was positioned three base-pairs downstream of a CCAAT motif stretching from -14 to -10bp. In the mammary gland the transcription factor C/EBP β functions as an enhancer of transcription by binding to CCAAT motifs and is crucial for transcription of casein genes [39-41]. It is possible that the polymorphism described here affects the binding affinity of enhancers to the CCAAT motif, and thereby affect transcription efficiency of *CSN1S2*.

A second SNP (T>C), that has not previously been reported, was detected in the 5'- UTR of the gene *CSN1S2* at position -5 bp distant from the initiation codon in the sequence GYAAACatgG (Figure 2), and could directly influence translation. Bevilacqua *et al.* [42] found that while transcripts from all four casein genes are found at similar concentrations in mammary tissue, translated α_{S2}-casein and κ-casein are found in much lower concentrations in cow milk than α_{S1}-casein and β-casein. The 5'- UTR sequence for the four caseins were strictly conserved between cattle, sheep and goat, and they suggested that variation in the Kozak consensus sequence (GCCRCCatgG [43]) might be the cause of the observed variation in translational efficiency between casein genes (Figure 2). Matching well with the higher protein levels associated with the PY93_C haplotype from 1606 Frasse; the C allele found in PY93_C was in better accordance with the Kozak consensus sequence than the alternative T allele, and therefore expected to produce a more efficient translation initiation site within the *CSN1S2* transcript. Work is in progress to deduce functionality of the detected polymorphisms on the transcription and translation of *CSN1S2* by expression profiling and quantitative determination of α_{S2}-casein in milk.

Table 3 – Polymorphisms associated with protein yield
Btau_4.0 position [44], gene, description of region (promoter, exon or UTR),translated protein position (TPP), polymorphic nucleotide (NT) alleles denoted as positive or negative [+/-], polymorphic translated protein (TP) alleles denoted as positive or negative [+/-] and promoter position (PP) relative to transcription initiation site. * Translated protein position in Ref NM_181008.2 for CSN2 and Ref NM_ 174528.2 for CSN1S2. † Polymorphism previously described in Lien *et al.* [1]. ‡Polymorphism previously described in Nilsen *et al.* [3].

Btau_4.0 (bp)	Gene	Region	TPP *	NT [+/-]	TP [+/-]	PP (bp)
88331023	*CSN2*	Exon	125	[T/C] †	P	
88331153	*CSN2*	Exon	82	[C/A] ‡	[P/H]	
88338754	*CSN2*	Promoter		[G/A]		-367
88339252	*CSN2*	Promoter		[T/A]		-865
88409015	*CSN1S2*	Promoter		[C/A]		-7
88410860	*CSN1S2*	UTR		[C/T]		

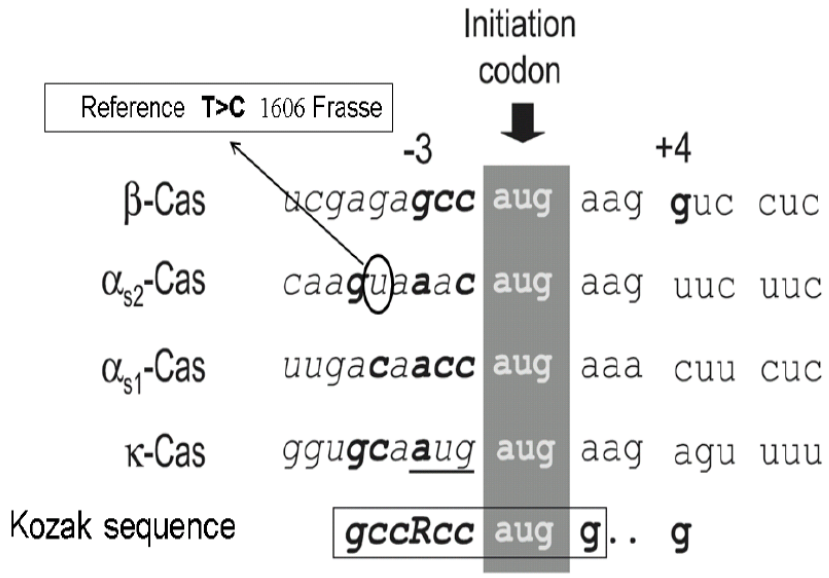


Figure 2 – A novel polymorphism in the Kozak sequence of α_{S2}-casein
A novel single-nucleotide polymorphism was discovered in a motif known to regulate efficiency of translation (the Kozak sequence) [43] in the 5'-flanking UTR of CSN1S2 coding for α_{S2}-casein. The polymorphic nucleotide is indicated and alleles found in the reference sequence and in the Holstein-Fresian Bull 1606 Frasse are presented. The figure is a modification of the figure presented in Bevilacqua *et al.* [42].

Long range haplotypes affecting clinical mastitis

Highest test scores for CM1, CM2 and CM3 from both single-marker and haplotype association mapping were found within the genomic interval 89 to 91Mb on BTA6 (Table 1). To be able to group genetically similar sires for the QTL for CM a 48 marker haplotype window (CM48) covering the genomic interval 89.4 to 90.6Mb was defined (Figure 1). This interval contained the SNPs producing the highest test scores from association mapping for CM. Haplotype classification within CM48 was performed based on phased chromosomes for NRF sires for which genotypes were available. The data indicated that 1606 Frasse was homozygous for the QTL for CM, and sires with the haplotype from Frasse 1606 had daughters with significantly elevated mastitis susceptibility (p-value of 1.469^{-07}). As can be seen from Figure 1, the 1606 Frasse haplotype was not markedly disrupted by recombination

between the haplotype windows PY93 and CM48. As previously noted the haplotype from Frasse 1606 was also significantly associated with higher levels of protein content in milk [1, 3], which might partly explain the observed genetic correlation between PY and CM in NRF [13].
In contrast to the QTL for PY which gave very high test scores for potent candidate genes (Figure 1a), results for CM were much more dispersed (Figure 1b). SNPs strongly associated with CM were not concentrated to a few specific genes, meaning that a number of genes in the QTL region for CM could harbor polymorphisms potentially affecting mastitis susceptibility. The highest test scores for CM were found within and around the three genes *SLC4A4*, *GC* and *ADAMTS3*. The first of these (*SLC4A4*) codes for a sodium bicarbonate co-transporter involved in maintaining normal blood pH [45, 46], the second (*GC*) encodes the main carrier protein of vitamin D in plasma, and finally the third gene (*ADAMTS3*) shows high similarity with *ADAMTS2*, which codes for a pro-collagen N-proteinase [47]. A cluster of genes coding for the CXC chemokines *IL8*, *CXCL5*, *PF4* and *CXCL2* are positioned around 92Mb, quite close to the highest scoring region for CM. CXC chemokines are important pro-inflammatory mediators and might therefore contain variation affecting mastitis susceptibility. Previously the genes *MUC7* and *IGJ* have been proposed as candidate genes for mastitis susceptibility in this region [2], but elevated test scores for CM were not found in or around these two genes in the current study (Figure 1b). Fine mapping with higher SNP density will be necessary in order to identify the most plausible candidate genes in the QTL region for CM.

Conclusion

Highest test scores from association mapping for PY were found in and around the casein genes *CSN2* and *CSN1S2*. Haplotype classification and high-coverage re-sequencing data indicated that the Holstein-Friesian bull 1606 Frasse, imported into the NRF cattle population in the 1970s, was homozygous for a haplotype encompassing these two genes. As previously suggested the haplotype from 1606 Frasse was significantly associated with elevated PY and selection for milk production traits is likely to have increased the frequency of this haplotype in the NRF population [1]. Data available from high throughput re-sequencing allowed for molecular characterization of the haplotype from 1606 Frasse, and plausible causal polymorphisms were detected in a regulatory element in the promoter region of the gene *CSN1S2* as well as in a motif that regulates translation efficiency of *CSN1S2* [42, 43].
It was further shown that the long range haplotype from 1606 Frasse is highly conserved in the NRF population for the region spanning the two QTLs affecting PY and CM on BTA6. The positive effect on milk protein content and the negative effect on mastitis susceptibility of this haplotype might partly explain the observed genetic correlation between these two traits in NRF [13].

Methods

Animals and phenotypes

NRF is an admixed breed formed from Norwegian breeds and imported animals from other Nordic countries. Norway has a dairy herd recording system which has included veterinary reported clinical mastitis (VRCM) since 1975 [48]. Records of VRCM in the periparturient period (-15 to 30 days post partum) of first (CM1), second (CM2) and third (CM3) lactation were available from GENO Breeding and AI Association [49] as daughter-yield-deviations (DYDs) for NRF sires. A sire with a high DYD value for CM has daughters with increased susceptibility to CM. Here records of VRCM were retrieved as a binary trait for daughters of sires from paternal half-sib families, yielding a large number of records per sire and a

reduction in variance due to environmental effects compared with other designs [50]. The mastitis traits CM1, CM2 and CM3 are described in Table 4 together with heritabilities and genetic correlations reported by Svendsen and Heringstad [49].
Records of PY were also available from GENO Breeding and AI Association as DYDs for NRF sires and were retrieved as 3,481,538 daughter records of 2,596 sires from paternal half-sib families. A sire with a high DYD value for PY has daughters with increased PY. For both CM and PY number of daughters per sire is highly variable and influenced by a small number of elite sires with a large number of daughters.

Table 4 – The mastitis traits
Trait and number of sire and daughter records for the mastitis traits CM1, CM2 and CM3 are presented. The final columns give heritabilities (on the diagonal) and genetic correlations reported by Svendsen and Heringstad [49].

	Number of records		Heritability and genetic correlation		
Trait	Daughters	Sires	CM1	CM2	CM3
CM1	1,755,649	2,596	0.03	0.74	0.68
CM2	1,256,887	2,532	-	0.02	0.85
CM3	805,376	2,440	-	-	0.02

Re-sequencing of candidate region and SNP detection
Sequence capture using a Roche-Nimblegen product was performed to isolate the region of interest on BTA6. Roche NimbleGen designed and manufactured a 5Mb sequence capture array targeting BTA6 coordinates 88-97Mb, standard repeat masking was applied in the design with 80% of the targeted bases being within a 100bp window of the final probe set. Sequence capture library construction was performed on seven samples, four NRF sires (2005 Smidesang, 10243 Rishaugen, 10263 Frestad and 10553 Nordbø) and three pools of old Norwegian breeds. Samples were sequenced using a 454 GS-FLX platform with the number of reads generated from each sample ranging from 97-460k. Sequence data was aligned to the BTA_4.0 reference genome [44] using the MOSAIK software package and standard alignment parameters [51]. SNP detection was performed with GigaBayes [52]. Criteria for filtering SNPs included minimum number of reads of each variant in non-coding regions ($\geq$2), minimum number of reads of the variant differing from the reference sequence for coding region ($\geq$2), GigaBayes score ($\geq$0.95) and minimum distance to closest SNP (>5bp). SNPs positioned in homopolymer regions (>5bp) were also rejected.

Genotype dataset
After initial filtering 269 SNPs remained from re-sequencing of candidate region and SNP detection by sequence capture and 454 sequencing. These SNPs were genotyped in 768 sires from paternal half-sib families using the Sequenom MassARRAY system. Genotypes were also retrieved for 84 SNPs for 2164 NRF sires from paternal half-sib families genotyped with the Affymetrix 25k MIP array and for 198 SNPs for 2596 NRF sires from paternal half-sib families genotyped with the Illumina Bovine SNP50 BeadChip. In addition, genotypes for 102 SNPs in the genomic region 86-90.4Mb on BTA6 were retrieved for 1143 sires [2]. Some SNPs were present in more than one dataset.
The data were checked for mendelian errors and based on the observed results, a cut-off of 4% was set to identify samples not fitting the pedigree. Pedigree errors were resolved by either identifying a new sire or setting parental information to unknown for the affected animal. New sires were assigned when the number of mendelian errors was equal to or lower than the background (0.4%) and there was only one candidate. After correcting for pedigree

errors, any remaining mendelian errors were corrected in addition to imputing those untyped genotypes that were possible based on inheritance. Linkage analysis, haplotyping and imputation were conducted with CRIMAP [53], PHASE [54] and locally developed software to combine the four datasets and fill in untyped genotypes. The local software was developed to handle genotyping errors and replace incorrect genotypes with correct ones where possible. The phasing procedure was implemented based on the six rules algorithm presented by Qian *et al.* [55], with modifications to fit half-sib families with missing data. The basic strategy was to first decide on the parental phases, starting with the youngest generation. Next step was to impute any remaining ambiguous or untyped positions in parents and offspring. Imputation in paternal haplotypes was performed by assuming no recombination between informative markers of the same phase. Imputation in maternal haplotypes, where no genotype information from parent were available, was performed by searching the rest of the dataset for equal haplotypes at surrounding informative positions and imputing when the untyped base could be decided uniquely. The final dataset contained genotypes for 556 SNPs in the BTA6 86-97Mb genomic region for 2601 sires. In addition to the pedigree checks performed on the filtered genotypes, possible pedigree problems were also tested by considering initial phasing results. Connections between offspring and sire showing a consistently high number of recombinations (above 20) for all chromosomes were removed. After removing animals due to initial phasing, genotypes were corrected to remove all double recombinants caused by single markers. Usually, the genotypes for the animal were deleted, but if several offspring of the same sire had problems for a single marker, the sire was corrected instead. Depending on the number of corrections needed to remove the double recombinants, the genotypes for the whole family might be deleted for the marker in question.

Single-marker association mapping

Single-marker association mapping for CM1, CM2, CM3 and PY were performed for all SNPs. The mixed model was:

$$P_i = Xg_j + Ya_i + Zm_k + e_{ijk}$$

Here phenotypic value P is DYD of sire i weighted by number of daughters, g is fixed effect of grandsire j, a is random effect of sire i where co-variance structure between sires is determined from pedigree relationships, m is random effect of genetic marker k and e is an error term. Estimation was conducted with the ASREML software [56]. MacLeod *et al.* [57] demonstrated that including effect of sire based on pedigree relationships reduces the number of false positives in association studies.

Haplotype association mapping

Pair-wise LD measure r^2 was found for all SNP pairs with the Haploview 4.1 software [58] and haplotype blocks were defined by the method described in Gabriel *et al.* [33] (GAB) and by the four gamete rule [34](GAM). A perl script was written to classify sires according to haplotypes for each of the defined haplotype blocks. The classification into GAB and GAM blocks were implemented in haplotype association mapping for CM1, CM2, CM3 and PY. The mixed model was:

$$P= Xg_i + Ya_j + Zh_k + e_{ijk}$$

Here h is random effect of haplotype k. Remaining terms are as described above for the single-marker association mapping. Estimation was conducted with the ASREML software [56].

Test score

LRT scores were calculated as two times the log-likelihood (LogL) ratio. LogL ratios were obtained with the ASREML software [56] for each SNP or haplotype as the difference between the LogL of a model containing the effect of the SNP or haplotype and the LogL of a model not containing this effect. LRT scores were expected to be distributed as a mixture of two χ^2 distributions with 0 and 1 degree of freedom. A logarithm of odds (LOD) score larger than 3, corresponding to a LRT score larger than 13.81 (p-value ≤ 0.0001), is an indication of genome-wide significance [59].

Predicted phenotypic value

Predicted phenotypic values were found as standard deviations from mean DYD for sires with at least one copy of a haplotype by applying a model with DYD as response variable and haplotype and grandsire as fixed effects. A two sample t-test was used to test difference in means of predicted phenotypic values between sires with at least one copy of the haplotype and remaining sires.

Genome re-sequencing and detection of polymorphisms

Genome re-sequencing of five NRF sires (1606 Frasse, 2636 Vik , 3454 J. Steinsvik, 2005 Smidesang and 1893 Rud) was performes on an Illumina GAIIx platform. Reads were generated as 2x108 paired-ends, coverage was 10Gb for 1606 Frasse and 2636 Vik, and 4Gb for 3454 J. Steinsvik, 2005 Smidesang and 1893 Rud. The FASTQ/A Clipper program from the FASTX-Toolkit [60] was used to remove adapter sequence and to discard reads based on average quality score (<10) or untyped bases (>7 Ns). Sequence data was assembled by mapping reads to the BTA_4.0 reference genome [44] using the BWA software package [61] and standard alignment parameters. Polymorphism detection was performed with SAMtools [62]. Criteria for filtering included minimum number of reads (≥2), maximum number of reads (≤100), minimum number of reads of the variant differing from the reference sequence (≥2) and minimum RMS mapping value (≥25). Polymorphisms positioned in homopolymer or repeat regions were discharged.

Author's contributions

MS bioinformatics for assembly and SNP detection for 454 sequence data, SNP filtering, association mapping, prediction of phenotypic values, haplotype classification, identification of relevant polymorphism and writing the manuscript. HG joining datasets by haplotyping and imputation and filtering of SNPs. MK conducted and coordinated molecular genetics work. ST bioinformatics for assembly and SNP detection for Illumina sequence data. MS provided DYDs and pedigree information. BJH assisted in finalizing the manuscript. SL conceived of the study, coordination and assistance in drafting the manuscript. All authors helped finalize the manuscript and read and approved of the final version.

Acknowledgements

Thanks to Paul R. Berg, Marianne H. S. Hansen, Hanne Hamland, Arne Roseth and Kristil K. Sundsaasen for sample processing, to Tina Graceline for bioinformatics assistance and to GENO Breeding and AI association for providing relationship information and DYD values. This project has been funded by The Research Council of Norway, GENO Breeding and AI association and BoviBank Ltd.

References

1. Lien S, Gomez-Raya L, Steine T, Fimland E, Rogne S: **Associations between casein haplotypes and milk yield traits**. *J Dairy Sci* 1995, **78**(9):2047-2056.
2. Nilsen H, Olsen HG, Hayes B, Nome T, Sehested E, Svendsen M, Meuwissen TH, Lien S: **Characterization of a QTL region affecting clinical mastitis and protein yield on BTA6**. *Anim Genet* 2009, **40**(5):701-712.
3. Nilsen H, Olsen HG, Hayes B, Sehested E, Svendsen M, Nome T, Meuwissen T, Lien S: **Casein haplotypes and their association with milk production traits in Norwegian Red cattle**. *Genet Sel Evol* 2009, **41**:24.
4. Bovenhuis H, Weller JI: **Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers**. *Genetics* 1994, **137**(1):267-280.
5. Velmala RJ, Vilkki HJ, Elo KT, de Koning DJ, Maki-Tanila AV: **A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle**. *Anim Genet* 1999, **30**(2):136-143.
6. Schopen GC, Koks PD, van Arendonk JA, Bovenhuis H, Visker MH: **Whole genome scan to detect quantitative trait loci for bovine milk protein composition**. *Anim Genet* 2009, **40**(4):524-537.
7. Farrell HM, Jr., Jimenez-Flores R, Bleck GT, Brown EM, Butler JE, Creamer LK, Hicks CL, Hollar CM, Ng-Kwai-Hang KF, Swaisgood HE: **Nomenclature of the proteins of cows' milk--sixth revision**. *J Dairy Sci* 2004, **87**(6):1641-1674.
8. Martin P, Szymanowska M, Zwierzchowski L, Leroux C: **The impact of genetic polymorphisms on the protein composition of ruminant milks**. *Reprod Nutr Dev* 2002, **42**(5):433-459.
9. Caroli AM, Chessa S, Erhardt GJ: **Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition**. *J Dairy Sci* 2009, **92**(11):5335-5352.
10. Schild TA, Geldermann H: **Variants within the 5'-flanking regions of bovine milk-protein-encoding genes. III. Genes encoding the Ca-sensitive caseins αs1, αs2 and β** *Theoretical and Applied Genetics* 1996, **93**:887-893.
11. Hallen E, Wedholm A, Andren A, Lunden A: **Effect of beta-casein, kappa-casein and beta-lactoglobulin genotypes on concentration of milk protein variants**. *J Anim Breed Genet* 2008, **125**(2):119-129.
12. Szymanowska M, Siadkowska E, Lukaszewicz M, Zwierzchowski L: **Association of nucleotide-sequence polymorphism in the 5'-flanking regions of bovine casein genes with casein content in cow's milk**. *Le Lait* 2004, **84**:579-590.
13. Heringstad B, Chang YM, Gianola D, Klemetsdal G: **Genetic association between susceptibility to clinical mastitis and protein yield in norwegian dairy cattle**. *J Dairy Sci* 2005, **88**(4):1509-1514.
14. Heringstad B, Klemetsdal G, Ruane J: **Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries** *Livestock Production Science* 2000, **64**(2-3):95-106.
15. Heringstad B, Chang YM, Gianola D, Klemetsdal G: **Genetic analysis of clinical mastitis, milk fever, ketosis, and retained placenta in three lactations of Norwegian red cows**. *J Dairy Sci* 2005, **88**(9):3273-3281.
16. Lund MS, Jensen J, Petersen PH: **Estimation of genetic and phenotypic parameters for clinical mastitis, somatic cell production deviance, and protein yield in dairy cattle using Gibbs sampling**. *J Dairy Sci* 1999, **82**(5):1045-1051.
17. Mardis ER: **The impact of next-generation sequencing technology on genetics**. *Trends Genet* 2008, **24**(3):133-141.
18. Stratton M: **Genome resequencing and genetic variation**. *Nat Biotechnol* 2008, **26**(1):65-66.

19. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ *et al*: **Direct selection of human genomic loci by microarray hybridization**. *Nat Methods* 2007, **4**(11):903-905.
20. Weaver TA: **High-throughput SNP discovery and typing for genome-wide genetic analysis**. *Trends in Genetics* 2000, **December 2000**:36-42.
21. Vignal A, Milan D, SanCristobal M, Eggen A: **A review on SNP and other types of molecular markers and their use in animal genetics**. *Genet Sel Evol* 2002, **34**(3):275-305.
22. Lindholm E, Zhang J, Hodge SE, Greenberg DA: **The reliability of haplotyping inference in nuclear families: misassignment rates for SNPs and microsatellites**. *Hum Hered* 2004, **57**(3):117-127.
23. Gao G, Allison DB, Hoeschele I: **Haplotyping methods for pedigrees**. *Hum Hered* 2009, **67**(4):248-266.
24. Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A: **A comprehensive evaluation of SNP genotype imputation**. *Hum Genet* 2009, **125**(2):163-171.
25. Hao K, Chudin E, McElwee J, Schadt EE: **Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies**. *BMC Genet* 2009, **10**:27.
26. Pei YF, Zhang L, Li J, Deng HW: **Analyses and comparison of imputation-based association methods**. *PLoS One* 2010, **5**(5):e10827.
27. Browning SR: **Missing data imputation and haplotype phase inference for genome-wide association studies**. *Hum Genet* 2008, **124**(5):439-450.
28. Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M *et al*: **Extensive genome-wide linkage disequilibrium in cattle**. *Genome Res* 2000, **10**(2):220-227.
29. Vallejo RL, Li YL, Rogers GW, Ashwell MS: **Genetic diversity and background linkage disequilibrium in the North American Holstein cattle population**. *J Dairy Sci* 2003, **86**(12):4137-4147.
30. Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM: **Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes**. *J Anim Sci* 2003, **81**(3):617-623.
31. Odani M, Narita A, Watanabe T, Yokouchi K, Sugimoto Y, Fujita T, Oguni T, Matsumoto M, Sasaki Y: **Genome-wide linkage disequilibrium in two Japanese beef cattle breeds**. *Anim Genet* 2006, **37**(2):139-144.
32. Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Cavanagh JA, Barris W, Schnabel RD, Taylor JF, Raadsma HW: **Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel**. *BMC Genomics* 2008, **9**:187.
33. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M *et al*: **The structure of haplotype blocks in the human genome**. *Science* 2002, **296**(5576):2225-2229.
34. Wang N, Akey JM, Zhang K, Chakraborty R, Jin L: **Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation**. *Am J Hum Genet* 2002, **71**(5):1227-1234.
35. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins**. *Nucleic Acids Res* 2007, **35**(Database issue):D61-65.
36. Boettcher PJ, Caroli A, Stella A, Chessa S, Budelli E, Canavesi F, Ghiroldi S, Pagnacco G: **Effects of casein haplotypes on milk production traits in Italian Holstein and Brown Swiss cattle**. *J Dairy Sci* 2004, **87**(12):4311-4317.
37. Ikonen T, Bovenhuis H, Ojala M, Ruottinen O, Georges M: **Associations between casein haplotypes and first lactation milk production traits in Finnish Ayrshire cows**. *J Dairy Sci* 2001, **84**(2):507-514.

38. Velmala R, Vilkki J, Elo K, Maki-Tanila A: **Casein haplotypes and their association with milk production traits in the Finnish Ayrshire cattle**. *Anim Genet* 1995, **26**(6):419-425.
39. Wyszomierski SL, Rosen JM: **Cooperative effects of STAT5 (signal transducer and activator of transcription 5) and C/EBPbeta (CCAAT/enhancer-binding protein-beta) on beta-casein gene transcription are mediated by the glucocorticoid receptor**. *Mol Endocrinol* 2001, **15**(2):228-240.
40. Robinson GW, Johnson PF, Hennighausen L, Sterneck E: **The C/EBPbeta transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland**. *Genes Dev* 1998, **12**(12):1907-1916.
41. Rosen JM, Wyszomierski SL, Hadsell D: **Regulation of milk protein gene expression**. *Annu Rev Nutr* 1999, **19**:407-436.
42. Bevilacqua C, Helbling JC, Miranda G, Martin P: **Translational efficiency of casein transcripts in the mammary tissue of lactating ruminants**. *Reprod Nutr Dev* 2006, **46**(5):567-578.
43. Kozak M: **Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6**. *EMBO J* 1997, **16**(9):2482-2492.
44. Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y *et al*: **Bos taurus genome assembly**. *BMC Genomics* 2009, **10**:180.
45. Yu H, Riederer B, Stieger N, Boron WF, Shull GE, Manns MP, Seidler UE, Bachmann O: **Secretagogue stimulation enhances NBCe1 (electrogenic Na(+)/HCO(3)(-) cotransporter) surface expression in murine colonic crypts**. *Am J Physiol Gastrointest Liver Physiol* 2009, **297**(6):G1223-1231.
46. Igarashi T, Sekine T, Watanabe H: **Molecular basis of proximal renal tubular acidosis**. *J Nephrol* 2002, **15 Suppl 5**:S135-141.
47. Tang BL: **ADAMTS: a novel family of extracellular matrix proteases**. *Int J Biochem Cell Biol* 2001, **33**(1):33-44.
48. Østeras O, Solbu H, Refsdal AO, Roalkvam T, Filseth O, Minsaas A: **Results and evaluation of thirty years of health recordings in the Norwegian dairy cattle population**. *J Dairy Sci* 2007, **90**(9):4483-4497.
49. Svendsen M, Heringstad B: **New Genetic Evaluation for Clinical Mastitis in Multiparous Norwegian Red Cows**. *Interbull Bulletine* 2006, **35**:8-11.
50. Weller JI, Kashi Y, Soller M: **Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle**. *J Dairy Sci* 1990, **73**(9):2525-2537.
51. Strömberg M: **Mosaik 1.0 Documentation**. In.; 2009.
52. Marth GT: **GigaBayes: SNP and Short-INDEL Polymorphism Discovery Tool**. 2009.
53. Green P, K. Falls and S. Crooks.: **Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St. Louis, Mo., USA**. 1990.
54. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data**. *Am J Hum Genet* 2001, **68**(4):978-989.
55. Qian D, Beckmann L: **Minimum-recombinant haplotyping in pedigrees**. *Am J Hum Genet* 2002, **70**(6):1434-1445.
56. Gilmour AR, B. R Cullis, S.J. Welham, and R. Thompson. : **ASREML reference manual.** 2000.
57. MacLeod IM, Hayes BJ, Savin KW, Chamberlain AJ, McPartlan HC, Goddard ME: **Power of a genome scan to detect and locate quantitative trait loci in cattle using dense single nucleotide polymorphisms**. *J Anim Breed Genet* 2010, **127**(2):133-142.
58. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps**. *Bioinformatics* 2005, **21**(2):263-265.
59. Lander ES, Botstein D: **Mapping mendelian factors underlying quantitative traits using RFLP linkage maps**. *Genetics* 1989, **121**(1):185-199.

60. **FASTX-Toolkit** [http://hannonlab.cshl.edu/fastx_toolkit/index.html]
61. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2009, **25**(14):1754-1760.
62. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.

Additional file 1

Table A1 – Genotyped single-nucleotide polymorphisms

The 556 SNPs genotyped in this study are presented by position, alleles (A1 and A2) and fraction of missing genotypes.

SNP	Position (bp)	A1	A2	Missing	SNP	Position (bp)	A1	A2	Missing
BTA-76946	86,091,437	A	G	0.0069	BTA-77101	88,268,695	A	G	0.0062
ss86309338	86,091,957	A	C	0.0062	ss86217839	88,291,433	A	G	0.1207
rs43703008	86,104,461	C	G	0.0065	ss86217840	88,291,473	C	G	0.0042
rs41654417	86,123,130	C	T	0.0042	ss86217841	88,295,268	C	T	0.0081
rs41654416	86,127,539	G	T	0.0065	ss117968347	88,302,639	A	G	0.0081
rs29001782	86,128,027	A	G	0.0069	S1_305616	88,305,616	A	G	0.005
ss61506487	86,365,126	A	C	0.0065	ss86217842	88,306,150	C	G	0.0073
ss61557722	86,434,938	A	G	0.0088	rs43703010	88,307,280	A	G	0.12
ss61491570	86,467,724	A	G	0.0115	ss86217843	88,307,439	A	G	0.0042
rs41618641	86,613,445	A	G	0.0058	BTA-115153	88,315,660	A	G	0.0054
ss86291015	86,671,685	A	G	0.0108	ss117968764	88,326,006	A	G	0.0119
ss86297176	86,712,348	A	G	0.0119	ss86217844	88,330,008	C	T	0.0119
rs29010229	86,810,566	A	T	0.0119	ss86217845	88,330,265	G	T	0.0115
rs41570706	86,908,337	A	T	0.01	rs43703013	88,330,987	C	G	0.0027
BTA-113299	86,965,533	C	G	0.0062	ss86217846	88,331,026	C	T	0.0054
ss117967957	87,103,072	A	C	0.0138	rs43703011	88,331,153	A	C	0.0111
rs29011726	87,242,360	C	T	0.0012	ss86217847	88,332,840	G	T	0.0111
rs29011727	87,242,379	A	G	0.0012	ss86217848	88,333,146	A	G	0.0119
rs29011728	87,242,429	A	G	0.0012	ss86217849	88,333,706	C	T	0.1027
ss86307579	87,255,540	A	G	0.0012	ss86217850	88,335,937	A	G	0.0119
ss86291546	87,370,506	A	G	0.0104	ss86217851	88,337,212	A	G	0.0023
ss61466227	87,480,010	A	C	0.015	ss86217852	88,337,966	A	G	0.0023
ss86324844	87,663,732	A	C	0.0096	ss86217853	88,338,919	A	T	0.0119
ss86317213	87,879,379	A	G	0.0058	ss86217854	88,339,983	A	T	0.0023
rs41610994	87,903,788	C	T	0.0223	ss86217855	88,340,058	C	T	0.0058
rs41610993	87,903,902	C	T	0.005	ss117968472	88,350,096	A	G	0.0058
ss61516066	87,904,282	A	G	0.0038	ss117968030	88,370,146	A	C	0.0054
rs29010267	87,989,290	A	G	0.0038	ss86217856	88,377,887	A	G	0.0104
BTA-77094	87,989,565	A	G	0.0038	ss86217857	88,378,201	C	T	0.0104
S1_140290	88,140,290	A	G	0.01	ss86217858	88,378,904	C	T	0.0104
rs29015040	88,160,178	A	T	0.0012	ss117968093	88,391,613	A	C	0.005
rs29015039	88,160,213	C	G	0.0012	ss86217859	88,407,310	A	G	0.005
S1_233640	88,233,640	A	C	0.0185	ss86217860	88,407,939	C	T	0.0119
ss117968170	88,263,656	A	G	0.01	ss86217861	88,408,758	A	G	0.0042

SNP	Position (bp)	A1	A2	Missing	SNP	Position (bp)	A1	A2	Missing
ss86217862	88,410,501	A	G	0.0104	ss86217887	88,532,740	A	G	0.0135
ss86217863	88,412,404	C	T	0.0108	ss86217888	88,532,923	C	T	0.0031
ss86217864	88,413,712	C	T	0.0104	ss86217889	88,532,930	A	C	0.0031
ss86217865	88,415,611	A	G	0.0108	ss86217890	88,533,205	G	T	0.0031
ss86217866	88,415,827	A	G	0.0023	ss86217891	88,533,423	A	G	0.0031
ss86217867	88,416,651	A	G	0.0042	ss86217892	88,533,570	A	G	0.0127
ss86217868	88,419,759	A	T	0.0038	ss86217893	88,533,625	A	G	0.0031
ss86217869	88,422,590	C	T	0.0042	ss86217894	88,534,065	C	G	0.0031
ss86217870	88,423,433	C	T	0.0111	rs29024681	88,537,898	A	G	0.0031
ss86217871	88,426,655	A	G	0.0104	rs29024683	88,537,969	A	G	0.0027
ss86217872	88,427,363	C	T	0.0104	rs29024684	88,538,027	A	C	0.0031
ss86217873	88,427,486	A	G	0.0104	rs29024685	88,538,077	A	G	0.0031
ss117968525	88,427,761	A	G	0.0142	S1_544997	88,544,997	G	A	0.0027
rs41588955	88,470,657	A	G	0.0108	S1_545414	88,545,414	A	G	0.0027
rs41588953	88,470,917	C	T	0.0073	S1_545614	88,545,614	A	G	0.015
ss117968780	88,473,588	A	G	0.0046	S1_579785	88,579,785	C	G	0.0027
S2_500334	88,500,334	C	A	0.0046	rs29025858	88,657,039	A	T	0.0111
ss86217874	88,505,291	A	G	0.0181	S1_723114	88,723,114	A	G	0.0085
ss86217875	88,505,604	C	T	0.0327	ss61465597	88,724,564	A	C	0.0119
rs43703014	88,505,736	A	T	0.005	S1_731394	88,731,394	A	G	0.0088
rs41588950	88,508,849	A	G	0.0319	S1_731444	88,731,444	G	A	0.0154
rs41588946	88,508,981	G	T	0.0123	ss86217895	88,757,210	C	T	0.03
rs41588945	88,509,069	C	T	0.0323	ss86217896	88,761,333	C	T	0.015
rs41588944	88,509,123	C	T	0.0323	ss99307233	88,761,588	C	G	0.015
S1_512065	88,512,065	G	A	0.0323	ss86217897	88,761,753	A	C	0.0115
ss86217876	88,519,758	A	T	0.0161	ss86217898	88,761,866	A	G	0.0058
ss86217877	88,520,726	A	G	0.0031	ss86217899	88,761,904	A	T	0.0069
ss86217878	88,520,893	G	T	0.0323	ss86217900	88,762,614	C	T	0.0058
ss86217879	88,520,981	C	T	0.0319	ss99307234	88,762,822	A	T	0.0119
ss86217880	88,521,023	C	T	0.0319	S4_784759	88,784,759	A	G	0.0196
ss86217881	88,528,999	A	C	0.0319	ss86312906	88,806,897	A	C	0.0277
ss86217882	88,530,213	C	T	0.0323	S1_874119	88,874,119	G	A	0.0115
ss86217883	88,531,652	A	G	0.0054	ss86217901	88,905,399	C	T	0.0127
S3_532297	88,532,297	G	A	0.0127	ss86217902	88,905,683	A	G	0.0115
rs43703015	88,532,298	C	T	0.0031	ss86217903	88,905,911	A	G	0.0154
rs43703016	88,532,334	A	C	0.0031	ss86217904	88,907,156	C	G	0.0204
rs43703017	88,532,354	A	G	0.0031	S1_907992	88,907,992	C	G	0.0115
ss86217884	88,532,395	A	G	0.0081	rs29019575	88,946,761	A	G	0.0146
ss86217885	88,532,403	A	T	0.0031	ss117968738	88,983,535	A	G	0.0019
ss86217886	88,532,715	C	T	0.0031	rs41655346	88,987,517	A	T	0.0073

SNP	Position (bp)	A1	A2	Missing	SNP	Position (bp)	A1	A2	Missing
rs41655347	88,990,898	A	C	0.0073	S1_2139654	90,139,654	A	G	0.0115
S1_1030092	89,030,092	G	A	0.0054	S1_2172727	90,172,727	G	A	0.0085
ss86326721	89,030,229	A	G	0.01	ss61489238	90,184,757	A	G	0.0073
rs29010354	89,081,281	A	G	0.01	rs29025895	90,241,567	A	G	0.0081
S3_1090122	89,090,122	C	A	0.0081	ss86302265	90,259,358	A	G	0.0138
S1_1135182	89,135,182	A	G	0.005	rs41629222	90,284,840	A	T	0.01
S1_1135353	89,135,353	G	A	0.0058	BTA-122716	90,284,888	A	G	0.0088
ss61557767	89,150,759	A	G	0.0131	ss61528083	90,288,991	A	C	0.0058
S1_1195672	89,195,672	C	G	0.0027	S1_2312723	90,312,723	G	A	0.0062
ss86341106	89,212,072	A	G	0.0135	ss61562683	90,317,539	A	C	0.0077
rs29010419	89,274,692	A	G	0.0092	S1_2327562	90,327,562	A	G	0.0054
S1_1339428	89,339,428	C	A	0.0104	ss61524397	90,356,013	A	G	0.0108
ss86285294	89,355,142	A	G	0.0111	S1_2374858	90,374,858	A	G	0.0054
rs41588980	89,355,672	A	G	0.0115	S1_2380155	90,380,155	G	A	0.01
rs41655357	89,369,237	A	C	0.0054	ss86297489	90,415,520	A	G	0.0096
rs41655356	89,369,291	C	T	0.0119	S2_2466078	90,466,078	A	G	0.0081
S1_1374004	89,374,004	G	A	0.015	ss86317874	90,485,680	A	G	0.0085
S3_1402237	89,402,237	G	A	0.0131	rs41629221	90,508,812	A	C	0.0085
S1_1403467	89,403,467	A	C	0.0069	S1_2516475	90,516,475	A	T	0.0073
S1_1428900	89,428,900	A	C	0.0165	S1_2518605	90,518,605	G	A	0.0085
S1_1482383	89,482,383	C	G	0.0065	ss61507506	90,564,544	A	G	0.0073
rs43474199	89,510,017	A	C	0.0088	S1_2600152	90,600,152	A	G	0.0038
rs29022799	89,603,520	A	G	0.0073	S1_2600253	90,600,253	G	A	0.0073
ss61524338	89,623,896	A	C	0.0012	S1_2605540	90,605,540	A	G	0.0096
S1_1625793	89,625,793	A	G	0.0108	S1_2608741	90,608,741	C	G	0.0031
S1_1650093	89,650,093	A	T	0.01	rs29024027	90,642,598	C	G	0.0096
S1_1650183	89,650,183	G	A	0.0104	rs29024026	90,644,772	A	C	0.0465
ss86278591	89,668,440	C	G	0.0081	ss61522200	90,670,190	A	G	0.0035
S1_1702619	89,702,619	C	G	0.0081	rs43052931	90,715,456	A	G	0.0058
ss38332444	89,730,160	A	G	0.0123	S1_2724105	90,724,105	A	G	0.0065
ss86337596	89,774,922	A	G	0.0104	S1_2725368	90,725,368	A	G	0.0065
S4_1808406	89,808,406	G	A	0.0088	rs43052940	90,737,717	A	G	0.01
rs43338539	89,838,827	A	G	0.0092	S1_2919075	90,919,075	C	G	0.0127
S1_1921855	89,921,855	A	G	0.0096	S1_2919904	90,919,904	C	G	0.0027
rs43338568	89,926,345	A	G	0.01	rs42932743	90,990,506	A	C	0.01
S1_1942988	89,942,988	G	A	0.0077	ss61557794	91,047,325	A	C	0.0088
ss86296213	90,008,099	A	G	0.01	S1_3064920	91,064,920	G	A	0.0058
S1_2011530	90,011,530	A	G	0.0111	S4_3074321	91,074,321	G	A	0.0085
S1_2012190	90,012,190	A	G	0.0108	S1_3089897	91,089,897	G	A	0.0085
rs42766480	90,075,263	A	G	0.0058	S1_3106225	91,106,225	G	A	0.0104

SNP	Position (bp)	A1	A2	Missing	SNP	Position (bp)	A1	A2	Missing
S4_3121513	91,121,513	A	G	0.0088	ss61557797	92,307,060	A	G	0.0081
S2_3121785	91,121,785	A	G	0.0146	rs42581544	92,434,963	A	G	0.0111
BTA-77209	91,138,928	A	T	0.0104	S2_4445535	92,445,535	C	A	0.0131
ss61523674	91,140,233	A	G	0.0092	ss61496193	92,473,530	A	C	0.0085
S3_3160463	91,160,463	A	G	0.0081	ss86285114	92,517,198	A	G	0.0023
rs41869408	91,190,937	A	G	0.0108	S1_4564474	92,564,474	A	C	0.0104
S1_3199714	91,199,714	G	A	0.0108	S1_4564611	92,564,611	G	A	0.0123
rs41870471	91,303,460	A	C	0.0088	ss86300582	92,646,173	A	G	0.01
S1_3365786	91,365,786	A	G	0.0081	S1_4648026	92,648,026	A	C	0.0111
S1_3387149	91,387,149	A	C	0.01	S1_4767066	92,767,066	A	G	0.0111
rs42149268	91,417,374	A	G	0.0081	S1_4767192	92,767,192	G	A	0.0135
S3_3472826	91,472,826	G	A	0.01	ss86335839	92,767,317	A	G	0.0135
S3_3493203	91,493,203	A	C	0.0085	S1_4769567	92,769,567	A	G	0.0131
ss61550746	91,553,825	A	G	0.0127	S1_4784117	92,784,117	G	A	0.0135
S1_3571393	91,571,393	C	G	0.0135	S1_4784162	92,784,162	A	G	0.0115
S1_3599731	91,599,731	C	A	0.0135	S3_4786144	92,786,144	A	G	0.0119
ss61557678	91,692,660	A	G	0.0135	S3_4786324	92,786,324	A	C	0.0119
rs29012368	91,747,096	A	G	0.0127	S3_4787256	92,787,256	A	C	0.0138
S1_3815688	91,815,688	G	A	0.0031	ss61568524	92,788,188	A	G	0.0246
S2_3831341	91,831,341	A	C	0.0104	S3_4791899	92,791,899	G	A	0.0119
rs43471504	91,874,621	A	G	0.015	S1_4827164	92,827,164	G	A	0.0127
rs43471476	91,897,799	A	C	0.005	ss86310987	92,835,501	A	C	0.0115
ss86334878	91,918,867	A	G	0.0092	rs42773532	92,854,146	A	G	0.0123
S1_3933203	91,933,203	G	A	0.01	ss86333471	92,886,879	A	G	0.0135
ss61557674	91,961,905	A	G	0.0085	S1_4895286	92,895,286	A	G	0.0115
S1_3966668	91,966,668	A	C	0.0062	rs42775634	92,914,806	A	G	0.0096
ss86311372	92,030,874	A	G	0.0062	ss61496362	92,949,328	A	G	0.0115
S1_4060584	92,060,584	A	C	0.01	rs42592169	93,002,336	A	G	0.005
S3_4060620	92,060,620	G	A	0.0119	S1_5030857	93,030,857	A	G	0.0115
S1_4060723	92,060,723	A	G	0.0108	rs42592200	93,031,311	A	G	0.0111
S1_4078268	92,078,268	G	A	0.0108	ss86313553	93,055,656	A	G	0.0135
BTA-76997	92,095,682	A	G	0.01	S1_5059912	93,059,912	A	G	0.0127
ss61557664	92,150,444	A	G	0.0092	S1_5061634	93,061,634	A	G	0.0127
S1_4157186	92,157,186	A	G	0.0042	S1_5102712	93,102,712	G	A	0.0142
ss86339292	92,179,319	C	G	0.0046	BTA-95635	93,124,191	A	G	0.0127
S1_4181027	92,181,027	A	G	0.01	S1_5124587	93,124,587	A	C	0.0092
ss86339125	92,216,838	A	G	0.0077	S1_5124691	93,124,691	G	A	0.0315
S1_4219093	92,219,093	C	G	0.0085	rs42582518	93,155,307	A	G	0.0085
rs29016177	92,240,040	A	G	0.0092	S1_5273006	93,273,006	A	G	0.0088
rs42959067	92,274,348	A	G	0.0046	S1_5283723	93,283,723	A	G	0.0054

SNP	Position (bp)	A1	A2	Missing	SNP	Position (bp)	A1	A2	Missing
S1_5283784	93,283,784	A	G	0.0108	S3_6064479	94,064,479	A	C	0.005
S1_5295667	93,295,667	A	G	0.0519	S2_6065214	94,065,214	A	G	0.0046
S1_5297134	93,297,134	C	G	0.0135	S3_6081399	94,081,399	A	T	0.0108
ss61493860	93,305,750	A	G	0.0077	ss117968835	94,082,193	A	C	0.0108
ss61569057	93,371,538	A	G	0.0138	ss46526537	94,129,168	A	G	0.0058
S1_5396499	93,396,499	C	G	0.0023	rs41256838	94,129,426	A	G	0.0069
S1_5407278	93,407,278	C	A	0.0119	S1_6203179	94,203,179	A	G	0.0069
S1_5409101	93,409,101	A	G	0.0119	ss86289048	94,204,423	A	G	0.0146
S1_5458165	93,458,165	A	G	0.0119	S4_6220618	94,220,618	G	A	0.0158
rs42615162	93,463,204	A	G	0.0119	S3_6223260	94,223,260	A	G	0.0065
S1_5468271	93,468,271	C	A	0.0185	S3_6223635	94,223,635	A	G	0.0154
S1_5468295	93,468,295	G	A	0.0119	S1_6251020	94,251,020	G	A	0.0081
S1_5486836	93,486,836	G	A	0.0123	S3_6259201	94,259,201	A	G	0.0073
rs43479253	93,545,537	A	G	0.0131	ss46526588	94,267,998	A	G	0.0165
S1_5546692	93,546,692	C	G	0.0096	S3_6272032	94,272,032	A	G	0.0096
S2_5546941	93,546,941	G	A	0.0108	S3_6272044	94,272,044	A	G	0.0096
S1_5566788	93,566,788	C	A	0.0108	S3_6286726	94,286,726	A	G	0.0211
S1_5613019	93,613,019	G	A	0.0031	ss86294120	94,293,113	A	C	0.01
ss86316121	93,640,925	A	G	0.0054	ss86336873	94,318,084	A	G	0.0092
S1_5645877	93,645,877	C	G	0.0196	S1_6336752	94,336,752	G	A	0.0088
rs42553777	93,682,966	A	G	0.0096	S1_6336772	94,336,772	A	G	0.0085
rs42553790	93,704,721	A	G	0.02	S1_6336828	94,336,828	G	A	0.0085
ss61495618	93,729,871	A	G	0.0246	S1_6337075	94,337,075	G	A	0.0085
rs42992679	93,767,254	A	G	0.0161	S1_6339238	94,339,238	A	G	0.0085
S1_5789318	93,789,318	A	G	0.0208	S1_6343946	94,343,946	C	G	0.0231
rs42555873	93,850,919	A	C	0.0177	S1_6369447	94,369,447	G	A	0.0158
S3_5864750	93,864,750	A	G	0.0104	rs43477315	94,384,509	A	G	0.0154
S1_5877815	93,877,815	A	G	0.0104	S1_6409488	94,409,488	A	G	0.01
S1_5890378	93,890,378	C	G	0.0104	S1_6413736	94,413,736	A	G	0.0092
rs42553820	93,918,271	A	C	0.0069	S1_6414384	94,414,384	G	A	0.0096
S1_5955044	93,955,044	A	G	0.0288	ss86313930	94,433,956	A	G	0.0054
S4_5955540	93,955,540	A	G	0.0042	S1_6487721	94,487,721	G	A	0.0081
ss86291415	93,962,055	A	G	0.0104	S1_6487844	94,487,844	A	G	0.0046
S1_5977073	93,977,073	A	G	0.03	S1_6526163	94,526,163	G	A	0.0046
S1_5984576	93,984,576	G	A	0.0185	S1_6540023	94,540,023	A	G	0.005
S1_6023891	94,023,891	G	A	0.0085	S1_6541850	94,541,850	A	G	0.0146
S1_6023958	94,023,958	A	G	0.0131	S1_6541909	94,541,909	C	A	0.0054
S3_6027168	94,027,168	C	A	0.0092	ss61557749	94,544,954	A	G	0.0092
S3_6027231	94,027,231	A	G	0.0092	S1_6607514	94,607,514	A	T	0.0085
ss117967932	94,050,759	A	G	0.0092	S1_6610441	94,610,441	C	G	0.0065

SNP	Position (bp)	A1	A2	Missing	SNP	Position (bp)	A1	A2	Missing
S1_6610634	94,610,634	G	A	0.0069	S1_7030075	95,030,075	A	G	0.0127
S1_6622612	94,622,612	A	G	0.0065	S4_7035373	95,035,373	G	A	0.0127
S1_6624188	94,624,188	C	G	0.0054	ss61557762	95,042,615	A	C	0.0473
ss86322902	94,640,252	A	G	0.0131	S1_7054092	95,054,092	G	A	0.0138
S1_6667541	94,667,541	G	A	0.0158	S1_7069325	95,069,325	C	G	0.0115
S1_6674676	94,674,676	A	G	0.0135	rs43480110	95,080,504	A	G	0.0131
S1_6688056	94,688,056	A	G	0.0088	S2_7081622	95,081,622	C	A	0.0092
S1_6688324	94,688,324	G	A	0.0096	S1_7094642	95,094,642	G	A	0.0131
rs43475613	94,709,427	A	G	0.0096	S3_7095881	95,095,881	G	A	0.0131
S4_6716379	94,716,379	A	G	0.0092	S3_7095929	95,095,929	A	C	0.0131
rs29016391	94,720,776	A	G	0.0092	ss86330106	95,103,436	A	G	0.0131
rs29016392	94,723,804	A	C	0.0092	S1_7125409	95,125,409	G	A	0.0135
ss86341394	94,736,057	A	G	0.0092	rs43470932	95,137,925	A	G	0.0127
rs29016270	94,737,420	A	G	0.0088	rs43470953	95,163,007	A	G	0.0058
S1_6738925	94,738,925	G	A	0.0181	S1_7183015	95,183,015	G	A	0.0058
S1_6740736	94,740,736	A	G	0.0085	S1_7217082	95,217,082	G	A	0.0115
S1_6786769	94,786,769	G	A	0.0092	S1_7217309	95,217,309	G	A	0.0108
ss86339206	94,788,665	A	G	0.0242	ss61472560	95,229,149	A	G	0.0108
S4_6790634	94,790,634	A	G	0.0081	S1_7240127	95,240,127	C	A	0.0108
ss86293643	94,827,334	A	G	0.0085	BTA-109071	95,256,811	A	G	0.0138
S1_6833805	94,833,805	G	A	0.0123	S1_7257839	95,257,839	A	T	0.0062
S4_6856096	94,856,096	C	G	0.015	ss86286430	95,269,007	A	C	0.0127
S1_6861220	94,861,220	A	G	0.0154	S1_7272273	95,272,273	G	A	0.0127
S1_6866365	94,866,365	C	A	0.0127	S4_7286837	95,286,837	A	G	0.0123
S1_6871888	94,871,888	A	G	0.0108	S3_7313175	95,313,175	G	A	0.0123
S1_6872070	94,872,070	G	A	0.0108	S1_7315550	95,315,550	G	A	0.0123
S1_6872209	94,872,209	C	G	0.0108	S1_7349924	95,349,924	G	A	0.0123
ss117963883	94,872,476	A	G	0.0108	S1_7356138	95,356,138	C	A	0.0115
S1_6873310	94,873,310	A	G	0.0108	S1_7357823	95,357,823	A	G	0.0131
BTA-77150	94,876,519	A	G	0.0115	S1_7361612	95,361,612	C	G	0.0115
S1_6890932	94,890,932	A	G	0.0042	S1_7381933	95,381,933	A	G	0.0131
S1_6895845	94,895,845	G	A	0.0154	S1_7409611	95,409,611	G	A	0.01
S1_6929113	94,929,113	C	G	0.0138	S1_7410456	95,410,456	C	G	0.0215
S1_6934119	94,934,119	A	G	0.0062	S1_7418018	95,418,018	A	C	0.0219
BTA-77152	94,957,486	A	G	0.0115	S1_7510680	95,510,680	C	G	0.0138
ss86332222	94,977,130	A	G	0.0038	ss86332567	95,528,037	A	C	0.0085
S1_7018465	95,018,465	G	A	0.0135	ss117968857	95,643,454	A	G	0.0115
S1_7025716	95,025,716	G	A	0.0119	S1_7696847	95,696,847	C	G	0.0135
S1_7027742	95,027,742	G	A	0.0127	ss86306932	95,704,460	A	G	0.0081
S1_7027990	95,027,990	A	G	0.0131	S1_7707881	95,707,881	A	G	0.0135

SNP	Position (bp)	A1	A2	Missing	SNP	Position (bp)	A1	A2	Missing
rs29020798	95,739,958	A	C	0.0115	S1_8397663	96,397,663	A	T	0.0108
rs29020799	95,740,171	A	T	0.0123	S1_8400023	96,400,023	G	A	0.0085
rs29020800	95,740,392	A	C	0.0315	S1_8405322	96,405,322	A	G	0.0027
rs43479594	95,770,022	A	G	0.0123	S1_8416402	96,416,402	A	G	0.0131
S1_7779152	95,779,152	G	A	0.0119	S1_8422848	96,422,848	G	A	0.0038
rs43475934	95,800,600	A	G	0.0131	S1_8423037	96,423,037	A	G	0.0365
ss61523677	95,840,994	A	G	0.0115	S1_8426125	96,426,125	G	A	0.01
ss61557804	95,925,105	A	G	0.0035	S1_8439911	96,439,911	C	A	0.01
ss61557805	95,958,447	A	G	0.0088	S1_8445514	96,445,514	C	A	0.0115
rs42800221	95,984,333	A	C	0.0096	rs29014369	96,447,880	C	G	0.0165
rs42801113	96,023,301	A	G	0.0096	S1_8461261	96,461,261	A	G	0.0065
ss86341675	96,063,579	A	G	0.0096	S1_8469079	96,469,079	A	G	0.0077
S1_8068950	96,068,950	A	G	0.0038	S1_8469211	96,469,211	A	G	0.0169
S2_8086729	96,086,729	C	G	0.0092	S1_8470042	96,470,042	C	A	0.0131
S1_8120611	96,120,611	G	A	0.0042	S1_8485952	96,485,952	C	A	0.0131
BTA-77356	96,189,520	A	C	0.0123	S1_8511422	96,511,422	G	A	0.0711
ss86289414	96,193,687	A	G	0.0119	rs43482362	96,513,910	A	G	0.0058
S1_8195493	96,195,493	A	G	0.0092	S1_8534488	96,534,488	C	G	0.0062
S1_8217099	96,217,099	A	G	0.0096	S1_8590026	96,590,026	G	A	0.0088
BTA-77352	96,217,245	A	G	0.0127	BTA-77248	96,597,778	A	G	0.0062
ss86310942	96,220,857	A	C	0.0046	ss86340510	96,601,544	A	G	0.0038
S1_8222411	96,222,411	A	G	0.0127	S4_8640753	96,640,753	G	A	0.0123
S1_8222975	96,222,975	G	A	0.0131	S1_8641446	96,641,446	A	G	0.0046
S1_8240448	96,240,448	C	G	0.0211	rs29022916	96,641,479	A	G	0.0108
S1_8247074	96,247,074	A	T	0.0115	S1_8664883	96,664,883	A	G	0.0042
ss61557846	96,259,022	A	G	0.0108	S1_8678259	96,678,259	G	A	0.0111
S1_8285649	96,285,649	C	A	0.0042	S1_8680977	96,680,977	G	A	0.0208
S1_8287325	96,287,325	A	G	0.0127	rs43479020	96,704,582	A	G	0.0115
ss86324329	96,299,549	A	G	0.0135	BTA-06258	96,723,957	A	G	0.0127
S1_8304280	96,304,280	G	A	0.0131	rs29020620	96,724,593	A	G	0.0058
S1_8306691	96,306,691	A	G	0.0135	S1_8762880	96,762,880	C	A	0.0058
rs29011685	96,314,998	C	G	0.0138	S1_8782146	96,782,146	A	G	0.015
rs43476086	96,322,848	A	G	0.0046	S1_8786997	96,786,997	G	A	0.015
S2_8337250	96,337,250	A	G	0.0111	BTA-27374	96,799,041	A	G	0.0146
S1_8341752	96,341,752	G	A	0.0127	S1_8821525	96,821,525	A	G	0.0065
S1_8342307	96,342,307	A	G	0.0042	S1_8821967	96,821,967	C	A	0.015
S1_8342553	96,342,553	A	G	0.0042	S1_8821975	96,821,975	G	A	0.015
rs43476044	96,344,643	A	G	0.0042	S1_8831027	96,831,027	A	G	0.015
S1_8359318	96,359,318	G	A	0.0027	rs42790107	96,844,636	A	G	0.0154
ss117968766	96,394,191	A	G	0.0104	S1_8861052	96,861,052	A	C	0.0135

SNP	Position (bp)	A1	A2	Missing
S1_8861166	96,861,166	G	A	0.0631
S1_8861667	96,861,667	G	A	0.118
S1_8861779	96,861,779	G	A	0.0631
S1_8876675	96,876,675	C	G	0.0631
S1_8881771	96,881,771	G	A	0.0654
ss61518585	96,884,950	A	G	0.0654
ss117968707	96,907,794	A	G	0.0627
ss61540596	96,979,904	A	C	0.0069