Genomics 95 (2010) 176–182

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

# A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in holstein cattle

Sameer D. Pant [a,*], Flavio S. Schenkel [a], Chris P. Verschoor [a], Qiumei You [a], David F. Kelton [b], Stephen S. Moore [c], Niel A. Karrow [a]

[a] Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario, Canada N1G 2W1
[b] Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada N1G 2W1
[c] Alberta Veterinary Research Institute, University of Alberta, Edmonton, Canada T6G 2P5

ARTICLE INFO

*Article history:*
Received 16 October 2009
Accepted 4 January 2010
Available online 7 January 2010

*Keywords:*
Johne's disease
Whole genome association analysis
Principal component regression
Bovine
QTL

ABSTRACT

Bovine Johne's disease (JD), caused by *Mycobacterium avium* spp. *paratuberculosis* (MAP), causes significant losses to the dairy and beef cattle industries. Effective vaccination or therapeutic strategies against this disease are currently unavailable and infected animals either get culled or die due to clinical disease. An alternative strategy to manage the disease is to selectively breed animals with enhanced resistance to MAP infection. Therefore, the objective of this study was to identify genetic loci putatively associated with MAP infection in a resource population consisting of Holstein cattle using a genome-wide association approach. The BovineSNP50 BeadChip, containing 54,001 single nucleotide polymorphisms (SNPs), was used to genotype 232 animals with known MAP infection status. Since, traditional case-control analytical techniques are based on single-marker analysis and do not account for the existence of linkage disequilibrium (LD) between markers, we used a novel principal component regression approach, where each SNP was fit in a logistic regression model, along with principal components of other SNPs on the same chromosome showing association with the trait, as covariates. Such an approach allowed us to account for the LD that exists between multiple markers showing an association on the same chromosome. Our analysis revealed the presence of at least 12 genomic regions on BTA1, 5, 6, 7, 10, 11 and 14 that were associated with the MAP infection status of our resource population. A brief description of these genomic regions, and a discussion of the analysis used in this study, have been presented.



## Introduction

*Mycobacterium avium* spp. *paratuberculosis* (MAP) causes Johne's disease (JD), also known as paratuberculosis, in ruminants and some wild-type species. MAP is an intracellular, slow-growing, Gram-positive, acid-fast bacterium that causes granulomatous inflammation of the small intestines. There are two principal subtypes of MAP: type 'C' that primarily infects cattle, and type 'S' that primarily infects sheep [1–4]. However, both subtypes can infect multiple species. MAP can also be isolated from some human subjects suffering from a similar chronic intestinal inflammatory disease (IBD) known as Crohn's disease (CD) [5]. This has prompted numerous investigations into the potential relationship between MAP infection and CD [6]. The precise role of MAP in the pathogenesis of CD is controversial, and there are both supporting [7,8], as well as opposing arguments [9–11].

Bovine JD is contagious in nature and spreads mostly through the ingestion of milk, colostrum or feces contaminated with MAP [12]. Vertical transmission *in utero* of the disease has also been reported [12,13], and although mycobacteria are able to persist in the reproductive organs of bulls, and in fresh [14] and cryopreserved semen used for artificial insemination [15], there is currently no evidence to support transmission from sire to calves. Calves are considered most susceptible to the disease, likely due to their low level of immunocompetence [16]. The disease has a prolonged incubation period (between 2 to 10 years) that depends on individual resistance to infection and the level of exposure to MAP [17,18]. During this incubation period, the infected animals remain asymptomatic, but perform poorly and actively shed MAP, exposing herd mates and thereby increasing their risk of infection [19–21]. Progressive weight loss and diarrhea are the main clinical signs, apparent only in the advanced stages of disease that follow the prolonged incubation period [22].

Bovine JD is spread worldwide and prevalence estimates range from 7% in Austria to 60% in New Zealand [19]. In the USA, Animal and Plant Health Inspection Service (APHIS) has reported MAP to be present in 68.1% of dairy operations (http://nahms.aphis.usda.gov/

* Corresponding author. Room 016B, Department of Animal and Poultry Science, 50 Stone Road East, Building # 70, University of Guelph, Guelph, Ontario, Canada N1G 2W1. Fax: +1 519 836 9873.
*E-mail address:* spant@uoguelph.ca (S.D. Pant).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.01.001

dairy/dairy07/Dairy2007_Johnes.pdf). Diagnosis of JD is difficult, and commercially available diagnostic tests are either expensive, and/or low in sensitivity. Common diagnostic tests are milk and serum enzyme-linked immunosorbent assays (ELISA), direct fecal polymerase chain reactions (PCR), and fecal culture [23–25]. The disease has a profound economic impact and annual losses are in excess of US $200 million in the US dairy industry alone [26]. These losses are attributed to reduced production and reproductive efficiency, increased treatment and management costs and premature culling or death from clinical disease [20,26,27].

There is no available cure for this disease to date, and although commercial vaccines targeting MAP are available; they only delay the onset of clinical signs instead of eliminating infection [28]. A widely accepted alternative strategy is to tailor livestock breeding strategies toward improving host genetics [29]. Individual susceptibility to MAP infection in cattle is heritable and estimates of heritability range from 0.06 to 0.183 [22,30,31]. This indicates the presence of genetic determinants of susceptibility to MAP infection in cattle populations. If identified, such genetic determinants could be used to select cattle for increased resistance to MAP infection. An understanding of such genetic determinants could also provide important insight into the potential relationship between CD and MAP infection in human subjects and target genes for drug therapy.

Few studies have attempted to identify genetic loci and variants conferring susceptibility/resistance to MAP infection in cattle. Studies based on a candidate gene approach have investigated associations between microsatellites and single nucleotide polymorphisms (SNPs) within or close to Toll like receptor 1 (*TLR1*),*TLR2*, *TLR4* [32,33], interferon-gamma (*IFNG*), solute carrier family 11, member 1 (*SLC11A1*) [33,34], Interleukin 4 (*IL4*), *IL10*, *IL12A*, *IL12B*, *IL18*, tumor necrosis factor alpha (*TNF*) [34] and caspase recruitment domain family, member 15 (CARD15) [35]. These studies reveal varying degrees of association between MAP infection in cattle and *TLR2*, *TLR4*, *SLC11A1* and *IFNG*, but these results have neither been replicated nor verified. There are also two published studies based on a whole genome association (WGA) analysis approach [36,37]. These studies used microsatellites and SNPs distributed throughout the genome to map quantitative trait loci (QTL) associated with the disease phenotype with limited success.

Bovine JD is a complex disease and is likely governed by a large number of genes, each having a small effect on the disease phenotype. This limits the application of WGA to identify genetic determinants and the focus is generally on identifying putative QTLs that harbor novel genes of interest for further statistical or biological investigation. Even then, there are two major challenges associated with the application of WGA in the study of complex diseases. First, WGA studies assume the existence of linkage disequilibrium (LD) between the marker loci that are genotyped and QTLs in the vicinity of these loci [38,39]. However, multiple markers that are genotyped within a genomic region may also be in strong LD. This causes multiple markers to show an association with the disease phenotype and it is difficult to discern between markers associated with QTLs that have a likely causal relationship with the disease phenotype, and markers that are in LD with other markers showing association [40]. The second challenge is that the population substructure within the diseased and healthy cohorts may result in spurious associations between genotyped markers and disease traits [41].

Linkage disequilibrium between genotyped markers could potentially be accounted for using a multiple regression approach and simultaneously fitting markers as covariates in a regression model. However, this approach fails if a large number of correlated markers are fitted in the model, as would be expected on a genome-wide scale. Under such circumstances, regression coefficients of SNPs that are calculated, while controlling for many other correlated SNPs, tend to be unreliable; a problem generally known as multicollinearity [42]. An alternative approach capable of overcoming multicollinearity, while still being able to account for LD between genotyped markers, has been recently described [43]. This approach makes use of a principal component regression (PCReg), first computing principal components (PCs) from the SNP genotype covariance matrix, and then fitting PCs as regressors in a multiple regression model. This allows the model covariates to remain orthogonal, thus overcoming multicollinearity, while still capturing the majority of the variation in the SNP genotypes. While this approach has been described in multilocus genetic studies of quantitative traits, we have not yet seen its implementation in the study of binomial disease phenotypes. Thus, the objective of our study was to use a PCReg approach to map QTLs conferring resistance/susceptibility to MAP infection in dairy cattle. We use a two stage logistic regression approach to analyze our data. In the first step, individual SNPs were tested for association with the disease phenotype using logistic regression. In the second step, all SNPs showing association at $p<0.05$ were analyzed on a chromosome-wise basis, using a combination of PCReg and multiple logistic regression to account for LD between markers showing association on a single chromosome. Hence, in this study, we describe the use of a principal component logistic regression approach to identify putative QTLs for MAP resistance in a population of Canadian dairy Holstein cattle. We have also discussed some biologically relevant genes in close proximity to the identified QTLs.

## Results

In the first step of the analysis, a total of 34,759 SNPs were individually tested for association with the binomial disease trait using a logistic regression model. We found 2679 SNPs, including 59 SNPs on unassigned contigs, to be associated with the disease trait at $p<0.05$. The 59 SNPs belonging to unassigned contigs were not carried into the second step of the analysis. The remaining 2620 SNPs were analyzed chromosome-wise in the second step using a PC multiple logistic regression model. A total of 550 SNPs were found to retain significance at $p<0.05$ after the second step. The genome-wide threshold for multiple testing ($p=1.96E-05$) was computed based only on the 2620 SNPs that were analyzed in the second step. Twenty-two SNPs on 7 different chromosomes were significantly associated with the disease trait using this genome-wide threshold. Among these SNPs, there were four pairs of SNPs that were in complete linkage with each other. These completely linked SNP pairs were situated on BTA5 (ss61491182, ss86274666), BTA6 (ss86341209, ss86284128), BTA7 (ss61558503, rs43505295) and BTA10 (ss61553578, ss86341021). We assumed SNPs having overlapping chromosomal regions within 1 Mbp of their map positions to represent a single QTL. Based on this assumption, these 22 SNPs were grouped to represent 12 QTLs. A brief description of these SNPs, including their accession numbers, chromosomal positions, major and minor alleles, minor allele frequency, odds ratio for the minor allele, *p*-values obtained after the second step of analysis, along with symbols of genes located within one mega basepairs (Mbp) of the QTLs is described in Table 1.

## Discussion

Our analysis revealed the presence of at least 12 putative QTLs for MAP resistance in the resource Holstein population and several interesting genes were identified within, or close to these QTLs. The most interesting genomic region associated with MAP resistance in our study was found on BTA7; it contained four SNPs (ss61491930, ss61558503, rs43505295 and ss86310793) associated with the disease phenotype after applying a genome-wide correction for multiple testing. More than 90 genes exist either inside this region or within 1 Mbps of the immediate surrounding genomic region. The most relevant genes among these are *IRF1* (interferon regulatory factor 1), *IL4*, *IL5*, *IL13*, *SLC39A3* (solute carrier family 39, member 3), *TNFAIP8L1* (tumor necrosis factor, alpha-induced protein 8-like 1) and

**Table 1**
A description of the SNPs within different genomic regions association with MAP resistance/susceptibility.

| Accession No. | Bta | Position (bp) | Maj./Min. all. | Min. all. freq. | Odds ratio (Min. all.) | *p*-value | Gene symbols (within 1 Mbp) |
|---|---|---|---|---|---|---|---|
| ss61563380 | 1 | 40,758,982 | G/T | 0.22 | 7.71 | 1.9E−06 | *TUBA3D* |
| ss61491182 | 5 | 14,416,892 | A/T | 0.26 | 0.05 | 6.6E−06 | *CCDC59, TMTC2* |
| ss86274666 | 5 | 14,500,783 | C/T | 0.27 | 0.04 | 1.1E−05 | |
| rs42852055 | 5 | 37,186,379 | C/T | 0.42 | 6.47 | 4.7E−06 | *FAM113B, AMIGO2, SLC38A4, SLC38A2, SLC38A1, SFRS2IP, ARID2* |
| ss61555725 | 5 | 41,512,973 | C/T | 0.25 | 0.03 | 3.8E−06 | *PRICKLE1, PPHLN1, ZCRB1, YAF2, GXYLT1* |
| rs29023629 | 5 | 87,137,388 | G/T | 0.25 | 7.45 | 5.6E−07 | *TMTC1, OVCH1, ERGIC2, FAR2, TM7SF3,* |
| ss61498341 | 5 | 88,577,519 | A/G | 0.41 | 8.87 | 2.2E−07 | *CCDC91, PTHLH, KLHDC5, MRPS35, FGFR10P2,* |
| ss61477621 | 5 | 88,838,035 | T/G | 0.36 | 11.85 | 1.7E−07 | *PPFIBP1, ARNTL2, STK38L, MED21, ITPR2* |
| ss86341209 | 6 | 108,202,550 | T/C | 0.38 | 0.04 | 4.1E−06 | *CRMP1, EVC, EVC2, STX18, STK32B,* |
| ss86284128 | 6 | 108,228,418 | C/T | 0.38 | 0.04 | 4.1E−06 | *MSX1, CYTL1* |
| ss61491930 | 7 | 17,929,919 | G/A | 0.22 | 15.40 | 5.5E−06 | *IRF1, IL5, IL13, IL4* |
| ss61558503 | 7 | 17,957,376 | T/C | 0.33 | 0.20 | 4.0E−08 | |
| rs43505295 | 7 | 17,983,797 | T/G | 0.33 | 0.20 | 4.0E−08 | |
| ss86310793 | 7 | 19,746,463 | T/C | 0.34 | 0.19 | 6.7E−08 | |
| rs42445244 | 7 | 83,397,046 | T/G | 0.27 | 5.27 | 1.0E−05 | *SSBP2, ATG10, XRCC4* |
| rs42556851 | 10 | 51,097,099 | T/C | 0.41 | 5.24 | 2.6E−07 | *NARG2, ANXA2, FOXB1, ADAM10, GRINL1A,* |
| ss61553578 | 10 | 52,944,489 | C/T | 0.32 | 4.79 | 4.3E−07 | *BNIP2, GTF2A2, LIPC, TCF12, MYO1E, CCNB2,* |
| ss86341021 | 10 | 52,966,013 | G/A | 0.32 | 4.79 | 4.3E−07 | *RNF111, AQP9* |
| ss86315269 | 10 | 60,612,527 | G/A | 0.19 | 11.94 | 4.9E−08 | *GLDN, SCG3, LYSMD2, TMOD2, TMOD3, LEO1, GNB5, MYO5C, AP4E1, TRPM7, USP50, USP8, GABPB2, HDC, SLC27A2* |
| ss86328445 | 11 | 30,601,961 | T/C | 0.43 | 0.08 | 1.7E−08 | *PRKCE, EPAS1, ATP6V1E2, PIGF, CRIPT, SOCS5, MCFD2, TTC7A, EPCAM, MSH2, KCNK12, MSH6, FBXO11* |
| ss61521480 | 14 | 53,982,484 | A/G | 0.36 | 0.12 | 7.8E−09 | *ANGPT1, RABL4, RSPO2, EIF3E, TTC35,* |
| rs42413954 | 14 | 54,033,703 | A/C | 0.41 | 0.16 | 6.8E−08 | *TMEM74, TRHR* |

*TICAM1* (Toll-interleukin 1 receptor domain containing adaptor molecule).

*IRF1* is an important transcription factor involved in the Type 1 ($T_h1$) cell-mediated immune response and is known to regulate the expression of many genes playing a role in the pathogenesis of human IBD such as *IL6, IL12B*, inducible nitric oxide synthase (*NOS2*) and major histocompatibility complex class II molecules [44–46]. Cell-mediated immunity is an important host defense mechanism against intracellular pathogens including MAP [47].

Interleukin-4, *IL5* and *IL13* are all type 2 cytokines that promote the $T_h2$ antibody-mediated immune response. Antibody-mediated immunity is relatively ineffective against intracellular pathogens, and a shift from a $T_h1$ to a $T_h2$ immune response can render the host incapable of combating MAP infection [48]. The clinical phase of both bovine JD and CD in humans is characterized by a gradual shift in the immune responses from cell-mediated immune response to antibody-mediated immune response [49–52]. Therefore, these $T_h2$ cytokines might play an important role in the pathogenesis of the disease.

TNFAIP8 is a cytosolic, antiapoptotic protein that can be induced by nuclear factor kappa beta (NF-κβ) as well as tumor necrosis factor alpha [53,54]. Both NF-κβ and TNF are critical for mediating inflammatory and immune responses against MAP [55,56]. TICAM1 is a toll-like receptor adaptor molecule that can induce type-I interferons [57] that play a role in the immune responses directed against Mycobacterial spp. [58].

We also found *SLC27A2*, *SLC38A1*, *SLC38A2*, *SLC38A4* and *SLC39A3* members of the solute carrier (SLC) superfamily within QTLs on BTA5, BTA7 and BTA10 that are associated with the disease phenotype. The SLC superfamily is a diverse group of more than 300 membrane transport proteins, many of which are also expressed in the intestines, where they play an important role in the uptake of macronutrients [59]. Candidate gene association studies have previously reported SLC11A1 to be associated with MAP infection in cattle [33,34]. Other members such as *SLC22A5*, *SLC22A23* and *SLC26A3* have also been implicated in CD [60]. None of the SLC members found within QTL regions in our study have been previously associated with bovine JD or CD in humans. In the case of cattle, this could very well be due to relatively few studies that have been published. Although little is known about the specific role that SLC family members could play in the pathogenesis associated with MAP infection in cattle or humans, the fact that many members of this family are involved in macronutrient uptake, and are expressed in the small intestines, which also serves as the portal of entry for MAP, makes them interesting candidate genes for future investigation.

Other important genes that exist either inside a putative QTL region or within 1 Mbps of the immediate surrounding genomic region are: annexin A2 (*ANXA2*), suppressor of cytokine signaling 5 (*SOCS5*), cytokine like-1 (*CYTL1*), and autophagy-related 10 homolog (*ATG10*). The protein encoded by *ANXA2* has multiple functions [61] that include mediating a plasmin-induced pro-inflammatory response in human peripheral blood monocytes [62] and mediating the growth factor effects of progastrin and gastrin peptides in intestinal epithelial cells [63]. *SOCS5* encodes a member of the SOCS family of proteins that are induced by cytokines and act in a negative feedback loop to regulate cytokine signaling and inflammation [64]. SOCS5 in particular, is expressed in lymphoid organs and is involved in regulating IL-4 signaling [65]. *CYTL1* encodes a protein that appears to be a novel cytokine based on its structural similarities with different cytokines [66–69]. *ATG10* encodes an enzyme that catalyzes the conjugation of ATG15-ATG5; a complex that is essential for autophagy [70], a process by which cells digest and recycle self-organelles. This is an important host defense mechanism and many autophagy-related genes play a role in the innate immune response against different pathogens [71]. Variants in three different autophagy-related genes have been implicated to play a role in the pathogenesis of CD [72].

The chromosomal region found to be associated with MAP resistance on BTA1 only contained the tubulin alpha 3d (*TUBA3D*) gene. *TUBA3D* encodes a member of the tubulin family of globular proteins that are involved in microtubule formation. The precise role that microtubules may play in pathogenesis of MAP infection is unclear. However, disruption of microtubules does have an effect on neutrophil motility [73].

Some QTL regions described in this study are located near previously described QTL regions for disease resistance traits in cattle. Putative QTLs for general disease resistance on BTA11 [74]; for clinical mastitis on BTA5, BTA11 and BTA14 [74–76]; for somatic cell count on

BTA10 and BTA11 [74,77]; and for somatic cell score on BTA1, BTA5, BTA6, and BTA14 [78,79] are situated in the vicinity of some of the putative QTL regions described in this study. This indicates that some of these putative QTLs could also influence other important disease traits in cattle populations.

The application of WGA analysis to identify QTLs for complex disease traits presents many difficulties pertaining to the sample size, selection of animals in the healthy and diseased cohorts, choice of markers for genotyping, and the type of model used to analyze the data. Specifically, in the case of mapping QTLs for MAP resistance in dairy cattle, the basis of classifying animals into diseased and healthy cohorts is a major concern. Available diagnostic tests are specific but lack desired sensitivity. This means that while positive results are indicative of MAP infection, negative results do not necessarily indicate that the animals are free of infection. This is especially true for younger infected animals that do not manifest any noticeable symptoms during the sub-clinical stages of infection. Some of these animals could potentially escape detection by the common ELISA diagnostic kits, and could thus be misclassified as being healthy (false negatives). To mitigate the inclusion of such false negatives, we avoided the inclusion of younger animals (<5.8 years) in our healthy cohort (mean age = 7.3 years). In addition, more than two-thirds of the healthy cohort had tested negative for MAP infection in previous years. Another potential concern regarding the healthy cohort is that healthy animals do not necessarily reflect enhanced genetic resistance to MAP. Animals coming from farms or herds with no prevalence of MAP will also test negative for MAP infection. In such cases, it is the absence of exposure to MAP, rather than enhanced genetic resistance of the animals, that translates into a negative test result. Therefore, in our study, animals that were classified as being healthy were only picked from farms with a high prevalence of MAP infection in the herd, ensuring the exposure of healthy animals.

Two previously published studies have attempted to map QTLs for MAP resistance in cattle using microsatellite and SNP markers [36,37]. It is known that multiallelic markers like microsatellites are more informative than bi-allelic markers like SNPs. We used SNPs in our study as they are relatively abundant in the genome and because an automated, high-throughput and cost-effective genotyping technology is available. The Illumina BovineSNP50 BeadChip used in this study contained 54,001 SNPs. It has been estimated that 10,000 SNPs are sufficient to find associations within cattle breeds [39]. Since our study focused on a single breed (Holsteins), the number of markers on the Illumina BovineSNP50 BeadChip provided sufficient power to our study.

Strict quality control (QC) measures were implemented prior to data analyses. All SNPs with a minor allele frequency (MAF) less than 10%, and all SNPs and individual animals with a genotype call rate of less than 95% were removed from the dataset prior to analysis. However, none of the animals were removed due to QC, as all animals had a genotype call rate of >95%. It is important to exclude SNPs with low MAF, especially in studies with a limited sample size in order to control type-I error. Insufficient observations to compute the estimates for rare alleles could potentially produce inflated estimates and spurious associations. This seems to be evident in the study published by Settles et al., who used the Illumina BovineSNP50 BeadChip to map QTLs for MAP resistance in Holstein cattle, but used a threshold of 1% for MAF during QC [37]. Consequently, 5 out of the 16 SNPs that were reported to be associated with MAP infection status, have a MAF of 1% with unrealistically high odds ratios, and are most likely spurious associations. A potential pitfall of using a high threshold for the MAF, is the possibility of overlooking rare disease alleles that might be in LD with the some of the markers that are excluded from the analyses due to low MAF [80]. However, within the limitations of the available sample size of our study, it would not have been possible for us to distinguish between rare alleles showing spurious associations and rare disease alleles having a true association with the disease phenotype.

We performed statistical analysis using a two-step logistic regression approach. In the first step, SNPs were tested for association with the disease phenotype, one at a time. In the second step, all SNPs found to be associated at $p<0.05$ in the first step, were re-analyzed, fitting PCs of SNPs that were also associated at $p<0.05$, and were on the same chromosome, as covariates. Our choice of logistic regression over other traditional case-control analytical techniques was based on two main reasons. First, logistic regression allows for joint analysis of multiple loci and second, it mitigates the effect of population substructure without a significant loss in power [81]. Accounting for population substructure was necessary since the pedigree of the resource population was unavailable due to client anonymity, and hence population substructure could potentially confound the analyses resulting in a higher type-I error rate.

Since we were limited by the number of animals in our resource population ($n = 232$), a preliminary step was required in our analyses to remove the majority of SNPs that were not associated with the disease trait. Computation of PCs of thousands of SNPs (34,759 in our study), based on the genotypes of a few hundred animals (232 in our study), results in PCs that individually explain very minute proportions of the total genotypic variance. Since inclusion of such PCs in the regression model leads to a perfect prediction the disease phenotype, resulting in almost infinite likelihoods and inaccurate estimates [82], the first step in our analyses was necessary to overcome this problem. The second step of the analyses was similar to the PCReg approach proposed by Wang and Abbot for quantitative traits [43]. Although PCs are able to summarize the total variance of the original genotype scores, and remain orthogonal, the PCReg approach only assesses the association between the disease phenotype and all SNPs used to compute the PCs, as a whole. Therefore, interpretations about the importance of individual SNPs cannot be made, based on the associations shown by PCs that are computed from them [83]. Joint analysis of an individual SNP along with PCs computed from the genotype covariance matrix of the rest of the SNPs on a chromosome, allows the multiple regression model to account for correlation between SNPs (which is mostly due to LD), without suffering from severe multicollinearity. Also, by fitting individual SNPs along with the PCs derived from the rest of the SNPs on the same chromosome, we avoided making any interpretation about the PCs and were able to test the association between the disease phenotype and individual SNPs while accounting for the association of the rest of the SNPs on the same chromosome.

## Conclusions

In conclusion, this GWA study, like the earlier two studies that were published based on a GWA approach, is an early attempt to reveal putative candidate genes that could play a role in conferring resistance to MAP infection in cattle populations. The statistical approach used in this study can accommodate for LD and thus, could be extended to other case-control association studies where accounting for LD among genetic markers is a problem. Our objectives were limited to identifying putative candidate genes that could be investigated further in future candidate gene studies. Once the role of such candidates is firmly established and characterized, variation within these genes could be exploited in order to make cattle populations more resistant to MAP infections for the benefit of the livestock industry.

## Materials and methods

### *Resource population*

Six commercial operations in Southwestern and Eastern Ontario were selected for sample collection based on a previous history of high prevalence of MAP infection. Blood was collected between the

months of July and September 2007 via the coccygeal (tail) vein from more than 400 cows based on age, breed, stage of lactation, infection status, and history of MAP screening. Current infection status was confirmed in blood plasma using the commercially available Herd-Chek M. pt. Antibody ELISA Test Kit (IDEXX Laboratories, Westbrook, ME, USA) according to manufacturer's instructions. Infection-free Holsteins that were older than 5.8 years of age were chosen for the healthy (MAP negative) cohort ($n=142$). There were 107 animals, within the healthy cohort, that had also tested negative for MAP infection in previous years. The mean age of this cohort was 7.3 years (range, 5.8 to 12.7 years). The infected (MAP positive) cohort ($n=90$) consisted of those Holsteins considered to be infected according to blood plasma MAP screening ($n=34$) and a second group considered to be infected according to milk ELISA MAP screening ($n=56$). Milk samples from these animals were generously provided by Canwest DHI (Guelph, ON, CAN) between July 2006 and November 2007, and due to client anonymity information such as age, pedigree, and location was not available.

*Genotyping and quality control*

Genomic DNA was extracted from the blood buffy coat using the DNeasy blood and tissue kit (Qiagen, Santa Clara, CA, USA), and from milk according to the methods described in Murphy et al. [84]. Sample DNA was quantified and subsequently genotyped using the Illumina BovineSNP50 BeadChip. The Illumina BovineSNP50 BeadChip assay contained 54,001 SNPs, of which, about 1672 SNPs were on unassigned contigs. The mean spacing between SNPs is 51.5 kb (median spacing of 37 kb; maximum spacing of 1.45 Mb) based on the BTAU 4.0 assembly (ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/). Samples were genotyped at the University of Alberta using BEADSTUDIO (Illumina) software.

Quality control measures were applied to the genotype data by excluding SNPs with low call rates and minor allele frequencies. In total, 15,194 SNPs with a minor allele frequency (maf) <10%, 153 markers with a low call rate <95% and all SNPs on the allosomes were excluded. Quality control measures to exclude individuals with low call rates <95% were also applied but none of the individuals was excluded. Genome-wide, 34,759 SNPs and 232 animals passed these quality control measures and were used for statistical analysis.

*Statistical analyses*

A two stage logistic regression approach was used to analyze SNP association with the binomial disease phenotype. In the first step, individual SNPs were tested for association with the phenotype using the following model:

$$\text{Log it}(Y_i) = \mu + \beta\alpha + e_i$$

where: $Y_i$ = binomial response phenotype of the $i$th animal; $\mu$ = overall mean; $\beta$ = regression coefficient for the additive effect of the SNP, $e_i$ = random error. The binomial response phenotype (dependent variable) was coded based on the presence (coded as '1') or absence (coded as '0') of MAP infection as defined earlier. The coded coefficients for the additive ($\alpha$) effect (independent variable) were

$\alpha_i = -1$ for a homozygote (MM)
0 for a heterozygote (Mm)
1 for the other homozygote (mm).

In the second step, all SNPs significant at $p<0.05$ were selected and analyzed chromosome-wise. Consider n SNPs significant on chromosome 'x' at $p<0.05$ in the preliminary analysis. Let the SNP being tested be denoted as $SNP_t$, '$g_{ij}$' be the genotype code for $i$th individual at the $j$th SNP and 'G' = $(g_{ij})$ be an $m \times n-1$ matrix of genotypes of all SNPs except $SNP_t$ on chromosome 'x'. PCs of the matrix 'G' were computed by singular value decomposition of the matrix 'G'. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ denote the eigenvalues (variances) of $k$ PCs that explain 90% of the total variance of the original 'G' matrix and $a_k$ denote the eigenvector associated with the eigenvalue $\lambda_k$. Then, the value of the $k$th PC term for the $i$th individual ($P_{ik}$) was computed as $(g_{i1}, g_{i2}, \ldots, g_{i(n-1)})a_k$. A stepwise logistic regression approach was implemented for further analyses, and initially all PC terms ($P$) explaining 90% of the total variance in the 'G' matrix, were included as covariates along with the $SNP_t$ in the model. A backward model selection based on Akaike's information criterion (AIC) was applied to the PC terms in the full model. Thus, PC terms were dropped from the model unless they improved the fit of the model based on AIC. The model selection procedure terminated when no further PC terms were dropped from the model and the odds ratio, confidence intervals and $p$-value for $SNP_t$ were obtained at this stage. The full model including $SNP_t$ and all PC terms explaining 90% of the total variation of the remaining SNPs on the same chromosome is as follows:

$$\text{Log it}(Y_i) = \mu + \beta_t\alpha_t + \sum_{j=1}^{k} \beta_j P_j + e_i$$

where: $Y_i$ = binomial response phenotype of the $i$th animal; $\mu$ = overall mean; $\beta_t$ = regression coefficient for the additive effect of the $SNP_t$, $\beta_j$ = multiple regression coefficients for the PC terms, $e_i$ = random error. The binomial response phenotype (dependent variable) was coded as in the preliminary analysis. The coded coefficients for the additive ($\alpha_t$) effect (independent variable) of $SNP_t$ were as in the single SNP regression analysis. This procedure was repeated for each SNP on each chromosome and $p$-values, odds ratios and confidence intervals were obtained for all SNPs. Multiple testing correction was applied using Sidak correction [85] after the second stage of the analysis, and was only based on the number of markers included in the second stage of our analysis.

## Acknowledgments

We are grateful to Dr. Mehdi Sargolzaei for his valuable suggestions and to Jackson Mah for his assistance with genotyping. We are also grateful to Jeremy Mount and Graham Biggar for their assistance in sample collection. This research was financed by Natural Sciences and Egnineering Research Council of Canada and DairyGen.

## References

[1] D.M. Collins, D.M. Gabric, G.W. de Lisle, Identification of two groups of *Mycobacterium paratuberculosis* strains by restriction endonuclease analysis and DNA hybridization, J. Clin. Microbiol. 28 (1990/7) 1591–1596.
[2] M. Florou, L. Leontides, P. Kostoulas, C. Billinis, M. Sofia, I. Kyriazakis, F. Lykotrafitis, Isolation of *Mycobacterium avium* subspecies *paratuberculosis* from non-ruminant wildlife living in the sheds and on the pastures of Greek sheep and goats, Epidemiol. Infect. 136 (2008/5) 644–652.
[3] B.J. Moloney, R.J. Whittington, Cross species transmission of ovine Johne's disease from sheep to cattle: an estimate of prevalence in exposed susceptible cattle, Aust. Vet. J. 86 (2008/4) 117–123.
[4] D.R. Ris, K.L. Hamel, J.M. Ayling, Can sheep become infected by grazing pasture contaminated by cattle with Johne's disease? N. Z. Vet. J. 35 (1987/8) 137.
[5] W.M. Chamberlin, S.A. Naser, Integrating theories of the etiology of Crohn's disease. On the etiology of Crohn's disease: questioning the hypotheses, Med. Sci. Monit. 12 (2006/2) RA27–RA33.
[6] L.A. Waddell, A. Rajic, J. Sargeant, J. Harris, R. Amezcua, L. Downey, S. Read, S.A. McEwen, The zoonotic potential of *Mycobacterium avium* spp. *paratuberculosis*: a systematic review, Can. J. Public Health 99 (2008/3) 145–155.
[7] W. Chamberlin, G. Ghobrial, M. Chehtane, S.A. Naser, Successful treatment of a Crohn's disease patient infected with bacteremic *Mycobacterium paratuberculosis*, Am. J. Gastroenterol. 102 (2007/3) 689–691.
[8] M. Feller, K. Huwiler, R. Stephan, E. Altpeter, A. Shang, H. Furrer, G.E. Pfyffer, T. Jemmi, A. Baumgartner, M. Egger, *Mycobacterium avium* subspecies *paratuberculosis* and Crohn's disease: a systematic review and meta-analysis, Lancet Infect. Dis. 7 (2007/9) 607–613.
[9] C.N. Bernstein, J.F. Blanchard, P. Rawsthorne, M.T. Collins, Population-based case control study of seroprevalence of *Mycobacterium paratuberculosis* in patients

with Crohn's disease and ulcerative colitis, J. Clin. Microbiol. 42 (2004/3) 1129–1135.
[10] J.L. Ellingson, J.C. Cheville, D. Brees, J.M. Miller, N.F. Cheville, Absence of *Mycobacterium avium* subspecies *paratuberculosis* components from Crohn's disease intestinal biopsy tissues, Clin. Med. Res. 1 (2003/7) 217–226.
[11] D.S. Rowbotham, N.P. Mapstone, L.K. Trejdosiewicz, P.D. Howdle, P. Quirke, *Mycobacterium paratuberculosis* DNA not detected in Crohn's disease tissue by fluorescent polymerase chain reaction, Gut 37 (1995/11) 660–667.
[12] S.S. Nielsen, H. Bjerre, N. Toft, Colostrum and milk as risk factors for infection with *Mycobacterium avium* subspecies *paratuberculosis* in dairy cattle, J. Dairy Sci. 91 (2008/12) 4610–4615.
[13] R.J. Whittington, P.A. Windsor, In utero infection of cattle with *Mycobacterium avium* subsp. *paratuberculosis*: a critical review and meta-analysis, Vet. J. 179 (2009/1) 60–69.
[14] A.B. Larsen, O.H. Stalheim, D.E. Hughes, L.H. Appell, W.D. Richards, E.M. Himes, *Mycobacterium paratuberculosis* in the semen and genital organs of a semen-donor bull, J. Am. Vet. Med. Assoc. 179 (1981/7/15) 169–171.
[15] M.C. Jorge, M. Catena, J. Cabodevila, P. Soto, Micobacterias en semen bovino criopreservado (in Spanish), Veterinaria Argentina 145 (1998) 337–340.
[16] C.C. Chase, D.J. Hurley, A.J. Reber, Neonatal immune development in the calf and its impact on vaccine response, Vet. Clin. North Am. Food Anim. Pract. 24 (2008/3) 87–104.
[17] P.M. Coussens, Model for immune responses to *Mycobacterium avium* subspecies *paratuberculosis* in cattle, Infect. Immun. 72 (2004/6) 3089–3096.
[18] M. Pence, C. Baldwin, C.C. Black III, The seroprevalence of Johne's disease in Georgia beef and dairy cull cattle, J. Vet. Diagn. Invest. 15 (2003/9) 475–477.
[19] I.R. Grant, Zoonotic potential of *Mycobacterium avium* ssp. *paratuberculosis*: the current position, J. Appl. Microbiol. 98 (2005) 1282–1293.
[20] J.E. Lombard, F.B. Garry, B.J. McCluskey, B.A. Wagner, Risk of removal and effects on milk production associated with paratuberculosis status in dairy cows, J. Am. Vet. Med. Assoc. 227 (2005/12/15) 1975–1981.
[21] S.S. Nielsen, M.A. Krogh, C. Enevoldsen, Time to the occurrence of a decline in milk production in cows with various paratuberculosis antibody profiles, J. Dairy Sci. 92 (2009/1) 149–155.
[22] M.G. Gonda, Y.M. Chang, G.E. Shook, M.T. Collins, B.W. Kirkpatrick, Genetic variation of *Mycobacterium avium* ssp. *paratuberculosis* infection in US Holsteins, J. Dairy Sci. 89 (2006).
[23] Z. Lu, R.M. Mitchell, R.L. Smith, J.S. Van Kessel, P.P. Chapagain, Y.H. Schukken, Y.T. Grohn, The importance of culling in Johne's disease control, J. Theor. Biol. 254 (2008/9/7) 135–146.
[24] H. Mikkelsen, G. Jungersen, S.S. Nielsen, Association between milk antibody and interferon-gamma responses in cattle from *Mycobacterium avium* subsp. *paratuberculosis* infected herds, Vet. Immunol. Immunopathol. 127 (2009/2/15) 235–241.
[25] S.S. Nielsen, N. Toft, Assessment of management-related risk factors for paratuberculosis in Danish dairy herds using Bayesian mixture models, Prev. Vet. Med. 81 (2007/10/16) 306–317.
[26] S.L. Ott, S.J. Wells, B.A. Wagner, Herd-level economic losses associated with Johne's disease on US dairy operations, Prev. Vet. Med. 40 (1999/6/11) 179–192.
[27] M.G. Gonda, Y.M. Chang, G.E. Shook, M.T. Collins, B.W. Kirkpatrick, Effect of Mycobacterium paratuberculosis infection on production, reproduction, and health traits in US Holsteins, Prev. Vet. Med. 80 (2007/7/16) 103–119.
[28] D.L. Emery, R.J. Whittington, An evaluation of mycophage therapy, chemotherapy and vaccination for control of *Mycobacterium avium* subsp. *paratuberculosis* infection, Vet. Microbiol. 104 (2004/12/9) 143–155.
[29] B. Lindhe, J. Philipsson, Conventional breeding programmes and genetic resistance to animal diseases, Rev. Sci. Tech. 17 (1998) 291–301.
[30] A.P. Koets, G. Adugna, L.L. Janss, H.J. van Weering, C.H. Kalis, G.H. Wentink, V.P. Rutten, Y.H. Schukken, Genetic variation of susceptibility to *Mycobacterium avium* subsp. *paratuberculosis* infection in dairy cattle, J. Dairy Sci. 83 (2000) 2702–2708.
[31] H. Mortensen, S.S. Nielsen, P. Berg, Genetic variation and heritability of the antibody response to Mycobacterium avium subspecies paratuberculosis in Danish Holstein cows, J. Dairy Sci. 87 (2004) 2108–2113.
[32] R. Mucha, M.R. Bhide, E.B. Chakurkar, M. Novak, I. Mikula Sr., Toll-like receptors TLR1, TLR2 and TLR4 gene mutations and natural resistance to *Mycobacterium avium* subsp. *paratuberculosis* infection in cattle, Vet. Immunol. Immunopathol. 128 (2009) 381–388.
[33] P.J. Pinedo, C.D. Buergelt, G.A. Donovan, P. Melendez, L. Morel, R. Wu, T.Y. Langaee, D.O. Rae, Candidate gene polymorphisms (BoIFNG, TLR4, SLC11A1) as risk factors for paratuberculosis infection in cattle, Prev. Vet. Med. 91 (2–4) (2009) 189–196.
[34] M. Hinger, H. Brandt, S. Horner, G. Erhardt, Short communication: association analysis of microsatellites and *Mycobacterium avium* subspecies *paratuberculosis* antibody response in German Holsteins, J. Dairy Sci. 90 (2007) 1957–1961.
[35] P.J. Pinedo, C.D. Buergelt, G.A. Donovan, P. Melendez, L. Morel, R. Wu, T.Y. Langaee, D.O. Rae, Association between CARD15/NOD2 gene polymorphisms and paratuberculosis infection in cattle, Vet. Microbiol. 134 (2009) 346–352.
[36] M.G. Gonda, B.W. Kirkpatrick, G.E. Shook, M.T. Collins, Identification of a QTL on BTA20 affecting susceptibility to *Mycobacterium avium* ssp. *paratuberculosis* infection in US Holsteins, Anim. Genet. 38 (2007) 389–396.
[37] M. Settles, R. Zanella, S.D. McKay, R.D. Schnabel, J.F. Taylor, R. Whitlock, Y. Schukken, J.S. Van Kessel, J.M. Smith, H. Neibergs, A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. *paratuberculosis* infection status in US holstein cattle, Anim. Genet. 40 (5) (2009) 655–662.
[38] J. Couzin, J. Kaiser, Genome-wide association. Closing the net on common disease genes, Science 316 (2007) 820–822.
[39] M.E. Goddard, B.J. Hayes, Mapping genes for complex traits in domestic animals and their use in breeding programmes, Nat. Rev. Genet. 10 (2009) 381–391.
[40] N. Malo, O. Libiger, N.J. Schork, Accommodating linkage disequilibrium in genetic-association analyses via ridge regression, Am. J. Hum. Genet. 82 (2008) 375–385.
[41] M.L. Freedman, D. Reich, K.L. Penney, G.J. McDonald, A.A. Mignault, N. Patterson, S.B. Gabriel, E.J. Topol, J.W. Smoller, C.N. Pato, M.T. Pato, T.L. Petryshen, L.N. Kolonel, E.S. Lander, P. Sklar, B. Henderson, J.N. Hirschhorn, D. Altshuler, Assessing the impact of population stratification on genetic association studies, Nat. Genet. 36 (2004) 388–393.
[42] S. Weisberg, Applied Linear Regression, John Wiley and Sons, Inc., New York, NY, 1985.
[43] K. Wang, D. Abbott, A principal components regression approach to multilocus genetic association studies, Genet. Epidemiol. 32 (2008) 108–118.
[44] M. Lohoff, D. Ferrick, H.W. Mittrucker, G.S. Duncan, S. Bischof, M. Rollinghoff, T.W. Mak, Interferon regulatory factor-1 is required for a T helper 1 immune response in vivo, Immunity 6 (1997) 681–689.
[45] D.L. McElligott, J.A. Phillips, C.A. Stillman, R.J. Koch, D.E. Mosier, M.V. Hobbs, CD4+ T cells from IRF-1-deficient mice exhibit altered patterns of cytokine expression and cell subset homeostasis, J. Immunol. 159 (1997) 4180–4186.
[46] S. Taki, T. Sato, K. Ogasawara, T. Fukuda, M. Sato, S. Hida, G. Suzuki, M. Mitsuyama, E.H. Shin, S. Kojima, T. Taniguchi, Y. Asano, Multistage regulation of Th1-type immune responses by the transcription factor IRF-1, Immunity 6 (1997) 673–679.
[47] A. Koets, V. Rutten, A. Hoek, M.F. van, K. Muller, D. Bakker, E. Gruys, E.W. van, Progressive bovine paratuberculosis is associated with local loss of CD4(+) T cells, increased frequency of gamma delta T cells, and related changes in T-cell function, Infect. Immun. 70 (2002/7) 3856–3864.
[48] P.M. Coussens, *Mycobacterium paratuberculosis* and the bovine immune system, Anim. Health Res. Rev. 2 (2001/12) 141–161.
[49] P.M. Coussens, C.B. Pudrith, K. Skovgaard, X. Ren, S.P. Suchyta, J.R. Stabel, P.M. Heegaard, Johne's disease in cattle is associated with enhanced expression of genes encoding IL-5, GATA-3, tissue inhibitors of matrix metalloproteinases 1 and 2, and factors promoting apoptosis in peripheral blood mononuclear cells, Vet. Immunol. Immunopathol. 105 (2005) 221–234.
[50] Z. Ren, J. Turton, T. Borody, G. Pang, R. Clancy, Selective Th2 pattern of cytokine secretion in *Mycobacterium avium* subsp. *paratuberculosis* infected Crohn's disease, J. Gastroenterol. Hepatol. 23 (2008) 310–314.
[51] J.S. Sohal, S.V. Singh, P. Tyagi, S. Subhodh, P.K. Singh, A.V. Singh, K. Narayanasamy, N. Sheoran, S.K. Singh, Immunology of mycobacterial infections: with special reference to Mycobacterium avium subspecies paratuberculosis, Immunobiology 213 (2008) 585–598.
[52] J.R. Stabel, Transitions in immune responses to *Mycobacterium paratuberculosis*, Vet. Microbiol. 77 (2000) 465–473.
[53] D. Kumar, P. Gokhale, C. Broustas, D. Chakravarty, I. Ahmad, U. Kasid, Expression of SCC-S2, an antiapoptotic molecule, correlates with enhanced proliferation and tumorigenicity of MDA-MB 435 cells, Oncogene 23 (2004) 612–616.
[54] C. Zhang, D. Chakravarty, I. Sakabe, R.R. Mewani, H.E. Boudreau, D. Kumar, I. Ahmad, U.N. Kasid, Role of SCC-S2 in experimental metastasis and modulation of VEGFR-2, MMP-1, and MMP-9 expression, Mol. Ther. 13 (2006) 947–955.
[55] L.G. Bekker, S. Freeman, P.J. Murray, B. Ryffel, G. Kaplan, TNF-alpha controls intracellular mycobacterial growth by both inducible nitric oxide synthase-dependent and inducible nitric oxide synthase-independent pathways, J. Immunol. 166 (2001) 6728–6734.
[56] S.E. Girardin, I.G. Boneca, J. Viala, M. Chamaillard, A. Labigne, G. Thomas, D.J. Philpott, P.J. Sansonetti, Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection, J. Biol. Chem. 278 (2003) 8869–8872.
[57] T. Seya, H. Oshiumi, M. Sasai, T. Akazawa, M. Matsumoto, TICAM-1 and TICAM-2: toll-like receptor adapters that participate in induction of type 1 interferons, Int. J. Biochem. Cell. Biol. 37 (2005) 524–529.
[58] R. Lande, E. Giacomini, T. Grassi, M.E. Remoli, E. Iona, M. Miettinen, I. Julkunen, E.M. Coccia, IFN-alpha beta released by *Mycobacterium tuberculosis*-infected human dendritic cells induces the expression of CXCL10: selective recruitment of NK and activated T cells, J. Immunol. 170 (2003) 1174–1182.
[59] M.A. Hediger, M.F. Romero, J.B. Peng, A. Rolfs, H. Takanaga, E.A. Bruford, The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction, Pflugers Arch. 447 (2004) 465–468.
[60] J.D. Rioux, R.J. Xavier, K.D. Taylor, M.S. Silverberg, P. Goyette, A. Huett, T. Green, P. Kuballa, M.M. Barmada, L.W. Datta, Y.Y. Shugart, A.M. Griffiths, S.R. Targan, A.F. Ippoliti, E.J. Bernard, L. Mei, D.L. Nicolae, M. Regueiro, L.P. Schumm, A.H. Steinhart, J.I. Rotter, R.H. Duerr, J.H. Cho, M.J. Daly, S.R. Brant, Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis, Nat. Genet. 39 (2007) 596–604.
[61] P. Raynal, H.B. Pollard, Annexins: the problem of assessing the biological role for a gene family of multifunctional calcium- and phospholipid-binding proteins, Biochim. Biophys. Acta 1197 (1994) 63–93.
[62] Y. Laumonnier, T. Syrovets, L. Burysek, T. Simmet, Identification of the annexin A2 heterotetramer as a receptor for the plasmin-induced signaling in human peripheral monocytes, Blood 107 (2006) 3342–3349.
[63] P. Singh, H. Wu, C. Clark, A. Owlia, Annexin II binds progastrin and gastrin-like peptides, and mediates growth factor effects of autocrine and exogenous gastrins on colon cancer and intestinal epithelial cells, Oncogene 26 (2007) 425–440.
[64] J. Elliott, J.A. Johnston, SOCS: role in inflammation, allergy and homeostasis, Trends Immunol. 25 (2004) 434–440.
[65] A. Ozaki, Y. Seki, A. Fukushima, M. Kubo, The control of allergic conjunctivitis by suppressor of cytokine signaling (SOCS)3 and SOCS5 in a murine model, J. Immunol. 175 (2005) 5489–5497.

[66] X. Liu, N. Rapp, R. Deans, L. Cheng, Molecular cloning and chromosomal mapping of a candidate cytokine gene selectively expressed in human CD34+ cells, Genomics 65 (2000) 283–292.
[67] M. Ficker, N. Powles, N. Warr, U. Pirvola, M. Maconochie, Analysis of genes from inner ear developmental-stage cDNA subtraction reveals molecular regionalization of the otic capsule, Dev. Biol. 268 (2004) 7–23.
[68] M. Hermansson, Y. Sawaji, M. Bolton, S. Alexander, A. Wallace, S. Begum, R. Wait, J. Saklatvala, Proteomic analysis of articular cartilage shows increased type II collagen synthesis in osteoarthritis and expression of inhibin betaA (activin A), a regulatory molecule for chondrocytes, J. Biol. Chem. 279 (2004) 43514–43521.
[69] J.S. Kim, Z.Y. Ryoo, J.S. Chun, Cytokine-like 1 (Cytl1) regulates the chondrogenesis of mesenchymal cells, J. Biol. Chem. 282 (2007) 29359–29367.
[70] Y. Shao, Z. Gao, T. Feldman, X. Jiang, Stimulation of ATG12-ATG5 conjugation by ribonucleic acid, Autophagy 3 (2007) 10–16.
[71] H.W. Virgin, B. Levine, Autophagy genes in immunity, Nat. Immunol. 10 (2009) 461–470.
[72] J.C. Barrett, S. Hansoul, D.L. Nicolae, J.H. Cho, R.H. Duerr, J.D. Rioux, S.R. Brant, M.S. Silverberg, K.D. Taylor, M.M. Barmada, A. Bitton, T. Dassopoulos, L.W. Datta, T. Green, A.M. Griffiths, E.O. Kistner, M.T. Murtha, M.D. Regueiro, J.I. Rotter, L.P. Schumm, A.H. Steinhart, S.R. Targan, R.J. Xavier, C. Libioulle, C. Sandor, M. Lathrop, J. Belaiche, O. Dewit, I. Gut, S. Heath, D. Laukens, M. Mni, P. Rutgeerts, A. Van Gossum, D. Zelenika, D. Franchimont, J.P. Hugot, M. de Vos, S. Vermeire, E. Louis, L.R. Cardon, C.A. Anderson, H. Drummond, E. Nimmo, T. Ahmad, N.J. Prescott, C.M. Onnie, S.A. Fisher, J. Marchini, J. Ghori, S. Bumpstead, R. Gwilliam, M. Tremelling, P. Deloukas, J. Mansfield, D. Jewell, J. Satsangi, C.G. Mathew, M. Parkes, M. Georges, M.J. Daly, Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease, Nat. Genet. 40 (2008) 955–962.
[73] E. Ben-Chetrit, S. Bergmann, R. Sood, Mechanism of the anti-inflammatory effect of colchicine in rheumatic diseases: a possible new outlook through microarray analysis, Rheumatology (Oxford) 45 (2006) 274–282.
[74] M.S. Lund, G. Sahana, L. Andersson-Eklund, N. Hastings, A. Fernandez, N. Schulman, B. Thomsen, S. Viitala, J.L. Williams, A. Sabry, H. Viinalass, J. Vilkki, 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 (2007) 5282–5290.
[75] M.S. Lund, B. Guldbrandtsen, A.J. Buitenhuis, B. Thomsen, C. Bendixen, Detection of quantitative trait loci in Danish Holstein cattle affecting clinical mastitis, somatic cell score, udder conformation traits, and assessment of associated effects on milk yield, J. Dairy Sci. 91 (2008) 4028–4036.
[76] H. Klungland, A. Sabry, B. Heringstad, H.G. Olsen, L. Gomez-Raya, D.I. Vage, I. Olsaker, J. Odegard, G. Klemetsdal, N. Schulman, J. Vilkki, J. Ruane, M. Aasland, K. Ronningen, S. Lien, Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle, Mamm. Genome 12 (2001) 837–842.
[77] C. Kuhn, J. Bennewitz, N. Reinsch, N. Xu, H. Thomsen, C. Looft, G.A. Brockmann, M. Schwerin, C. Weimann, S. Hiendleder, G. Erhardt, I. Medjugorac, M. Forster, B. Brenig, F. Reinhardt, R. Reents, I. Russ, G. Averdunk, J. Blumel, E. Kalm, Quantitative trait loci mapping of functional traits in the German Holstein cattle population, J. Dairy Sci. 86 (2003) 360–368.
[78] H.D. Daetwyler, F.S. Schenkel, M. Sargolzaei, J.A. Robinson, A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map, J. Dairy Sci. 91 (2008) 3225–3236.
[79] N.F. Schulman, S.M. Viitala, D.J. de Koning, J. Virta, A. Maki-Tanila, J.H. Vilkki, Quantitative trait Loci for health traits in Finnish Ayrshire cattle, J. Dairy Sci. 87 (2004) 443–449.
[80] M. Sargolzaei, F.S. Schenkel, G.B. Jansen, L.R. Schaeffer, Extent of linkage disequilibrium in Holstein cattle in North America, J. Dairy Sci. 91 (2008) 2106–2117.
[81] E. Setakis, H. Stirnadel, D.J. Balding, Logistic regression protects against population structure in genetic association studies, Genome Res. 16 (2006) 290–296.
[82] A. Albert, J.A. Anderson, On the existence of maximum likelihood estimates in logistic regression models, Biometrika 71 (1984) 1–10.
[83] R.J. Harris, A Primer of Multivariate Statistics, Academic Press, New York, 1975.
[84] M.A. Murphy, M.R. Shariflou, C. Moran, High quality genomic DNA extraction from large milk samples, J. Dairy Res. 69 (2002) 645–649.
[85] Z. Sidak, Rectangular confidence regions for the means of multivariate normal distributions, J. Am. Stat. Association 62 (1967) 626–633.