

GWAS and Fine-Mapping of Livability and Six Disease Traits in Holstein Cattle

Ellen Freebern^{1,†}, Daniel JA Santos^{1,†}, Lingzhao Fang^{1,2}, Jicai Jiang¹, Kristen L. Parker Gaddis³,
George E. Liu², Paul M. Vanraden², Christian Maltecca⁴, John B. Cole², Li Ma^{1,*}

¹Department of Animal and Avian Sciences, University of Maryland, College Park,
Maryland 20742, USA;

²Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural
Research Service, USDA, Beltsville, Maryland 20705, USA;

³Council on Dairy Cattle Breeding, Bowie, MD 20716, USA;

⁴Department of Animal Science, North Carolina State University, Raleigh, 27695, USA;

[†] Equal contribution

* Corresponding Author:

Li Ma, Department of Animal and Avian Sciences, University of Maryland, College Park,
Maryland 20742, USA.

E-mail: lima@umd.edu, Phone: 301-405-1389

Emails: efreeber@terpmail.umd.edu, djordand@umd.edu, lingzhaofang@gmail.com,
jicai.jiang@gmail.com, kristen.gaddis@uscddb.com, George.Liu@usda.gov,
Paul.Vanraden@usda.gov, cmaltec@ncsu.edu, john.cole@usda.gov, lima@umd.edu.

Abstract

Background: Health traits are of significant economic importance to the dairy industry due to their effects on milk production and associated treatment costs. Genome-wide association studies (GWAS) provide a means to identify associated genomic variants and thus reveal insights into the genetic architecture of complex traits and diseases. The objective of this study is to investigate the genetic basis of seven health traits in dairy cattle and to identify potential candidate genes associated with cattle health using GWAS, fine mapping, and analyses of multi-tissue transcriptome data.

Results: We studied cow livability and six direct disease traits, mastitis, ketosis, hypocalcemia, displaced abomasum, metritis, and retained placenta, using de-regressed breeding values and more than three million imputed DNA sequence variants. After data edits and filtering on reliability, phenotypes for 11,880 to 24,699 Holstein bulls were included in the analyses of the seven traits. GWAS was performed using a mixed-model association test, and a Bayesian fine-mapping procedure was conducted to calculate a posterior probability of causality to each variant and gene in the candidate regions. The GWAS results detected a total of eight genome-wide significant associations for three traits, cow livability, ketosis, and hypocalcemia, including the bovine MHC region associated with livability. Our fine-mapping of associated regions reported 20 candidate genes with the highest posterior probabilities of causality for cattle health. Combined with transcriptome data across multiple tissues in cattle, we further exploited these candidate genes to identify specific expression patterns in disease-related tissues and relevant biological explanations such as the expression of *GC* in the liver and association with mastitis as well as the *CCDC88C* expression in CD8 cells and association with cow livability.

Conclusions: Collectively, our analyses report six significant associations and 20 candidate genes of cattle health. With the integration of multi-tissue transcriptome data, our results provide useful information for future functional studies and better understanding of the biological relationship between genetics and disease susceptibility in cattle.

Keywords: GWAS, Fine-Mapping, Health Trait, Gene Expression, Dairy Cattle

Background

One of the fundamental goals of animal production is to profitably produce nutritious food for humans from healthy animals. Profitability of the dairy industry is influenced by many factors, including production, reproduction, and animal health [1]. Cattle diseases can cause substantial financial losses to producers as the result of decreased productivity, including milk that must be dumped, and increased costs for labor and veterinary care. Indirect costs associated with reduced fertility, reduced production after recovery, and increased risk of culling also can be substantial. For example, ketosis is a metabolic disease that occurs in cows during early lactation and hinders the cow's energy intake, thus subsequently reduces milk yield and increases the risk of displaced abomasum, which is very costly [2]. Mastitis is a major endemic disease of dairy cattle that can lead to losses to dairy farmers due to contamination, veterinary care, and decreased milk production [3]. In addition, cows may develop milk fever, a metabolic disease that is related to a low blood calcium level known as hypocalcemia [4]. Another common disease in cattle is metritis, which is inflammation of the uterus and commonly seen following calving when cows have a suppressed immune system and are vulnerable to bacterial infection [5]. Complications during delivery can also result in a retained placenta [6]. Despite the complex

nature of these cattle diseases, a better understanding of the underlying genetic components can help the management and genetic improvements of cattle health.

Genome-wide association studies (GWAS) have been successful at interrogating the genetic basis of complex traits and diseases in cattle [7-9]. Due to the high level of linkage disequilibrium (LD) between genomic variants, pinpointing causal variants of complex traits has been challenging [9]. Fine-mapping is a common post-GWAS analysis, where posterior probabilities of causality are assigned to candidate variants and genes [9]. In humans, fine-mapping of complex traits are currently on-going along or following GWAS studies [10]. The utility of fine-mapping in cattle studies, however, has been limited by data availability and the high levels of LD present in cattle populations. To circumvent this challenge, a recent study developed a fast Bayesian Fine-MAPping method (BFMAP), which performs fine-mapping by integrating various functional annotation data [9]. Additionally, this method can be exploited to identify biologically meaningful information from candidate genes to enhance the understanding of complex traits [11].

The U.S. dairy industry has been collecting and evaluating economically important traits in dairy cattle since the late 1800s, when the first dairy improvement programs were formed. Since then, a series of dairy traits have been evaluated, including production, body conformation, reproduction, and health traits. Cow livability was included in the national genomic evaluation system by the Council on Dairy Cattle Breeding (CDCB) in 2016 [12]. This trait reflects a cow's overall ability to stay alive in a milking herd by measuring the percentage of on-farm deaths per lactation. Cow livability is partially attributable to health and can be selected to provide more milk revenue and less replacement of cows. In 2018, six direct health traits were introduced into the U.S. genomic evaluation, including ketosis, mastitis, hypocalcemia or milk fever, metritis,

retained placenta, and displaced abomasum [13]. These phenotypic records along with genotype data collected from the U.S. dairy industry provide a unique opportunity to investigate the genetic basis of cattle health. The aim of our study is, therefore, to provide a powerful genetic investigation of seven health traits in cattle, to pinpoint the candidate disease genes and variants with relevant tissue-specific expression, and to provide insights into the biological relationship between candidate genes and the disease risk they may present on a broad scale.

Results

Genome-wide association study of livability and six direct health traits

We conducted genome-wide association analyses of seven health related traits in 27,214 Holstein bulls that have many daughter records and thus accurate phenotypes using imputed sequence data and de-regressed breeding values. After editing and filtering on reliability, we included 11,880 to 24,699 Holstein bulls across the seven traits (Table 1). Compared to the analysis using predicted transmitting ability (PTA) as phenotype (Additional File 1), GWAS on de-regressed PTA values produced more consistent and reliable results [14]. While different results between analyses of raw and de-regressed PTAs were obtained for the six health traits, little difference was observed for cow livability, which have more records and higher reliabilities (Table 1 and Additional File 2). Therefore, we only considered association results obtained with de-regressed PTAs in all subsequent analyses.

Out of the seven health traits, we detected significantly associated genomic regions only for three traits after Bonferroni correction, hypocalcemia, ketosis, and livability (Figure 1). In total, we had one associated region on BTA 6 for hypocalcemia, one region on BTA 14 for ketosis, and six regions for cow livability on BTA 5, 6, 14, 18, 21, and 23, respectively (Table 2). Notably, the bovine Major Histocompatibility Complex (MHC) region on BTA 23 [15] is

associated with cow livability. Additionally, association signals on BTA 16 for ketosis (P -value = 1.9×10^{-8}) and BTA 6 for mastitis (P -value = 4.2×10^{-8}) almost reached the Bonferroni significance level. Other traits had prominent signals, but their top associations were below the Bonferroni threshold. Since sequence data have the highest coverage of functional variants in our study, we included all these regions to query the Cattle QTLdb for a comparative analysis.

When compared to existing studies, many of these health related regions have been previously associated with milk production or disease related traits in cattle (Table 2) [16]. The top associated region for hypocalcemia is around 10,521,824 bp on BTA 6, where QTLs were reported for body/carcass weight and reproduction traits with nearby genes being *TRAMIL1* and *NDST4*. The region around 2,762,595 bp on BTA 14 for ketosis is involved with milk and fat metabolism and the well-known *DGATI* gene. The region around 7,048,452 bp on BTA 16 for ketosis was also previously associated with fat metabolism. The region around 88,868,886 bp on BTA 6 associated with mastitis is close to the *GC* gene with many reported QTLs associated with mastitis [17-20]. This region was also associated with cow livability in this study with QTLs involved with the length of productive life [21]. For the six regions associated with cow livability (Table 2), we found reported QTLs related to productive life, somatic cell count, immune response, reproduction, and body conformation traits [21]. The top associated regions for displaced abomasum (DSAB) on BTA 4 and BTA 8 have been previously associated with cattle reproduction and body conformation traits [22-24]. For metritis, the top associated variant, 3,662,486 bp on BTA4, is close to *RF00322*, and around ± 1 Mb upstream and downstream were QTLs associated with production, reproduction, and dystocia [25]. Genes *RUNDC3B* (BTA 4), *QDPR* (BTA 6), *TMEM182* (BTA 11), *ZFP28* (BTA 18) are the closest genes to the retained

placenta (RETP) signals with previous associations related to milk production, productive life, health and reproduction traits, including calving ease and stillbirth [8].

Association of livability QTL with other disease traits

Cow livability is a health-related trait that measures the overall robustness of a cow. As the GWAS of cow livability was the most powerful among the seven traits and detected six QTL regions, we evaluated whether these livability QTLs were also associated with other disease traits. Out of the six livability QTLs, four of them were related to at least one disease trait at the nominal significance level (Table 3). All these overlapped associations exhibited consistent directions of effect: alleles related to longer productive life were more resistant to diseases. The most significant QTL of livability on BTA 18 is associated with displaced abomasum and metritis, both of which can occur after abnormal birth. This QTL has been associated with gestation length, calving traits, and other gestation and birth related traits [11]. The QTL on BTA 6 is associated with hypocalcemia, ketosis, and mastitis. The BTA 21 QTL is associated with hypocalcemia and mastitis. The BTA 5 QTL is related to displaced abomasum and ketosis. Interestingly, the bovine MHC region on BTA 23 is not associated with any direct disease traits, which suggests that those genes do not explain substantial variation for the presence or absence of a disease during a lactation.

Fine-mapping analyses and validation from tissue-specific expression

Focusing on the candidate QTL regions in Table 2, the fine-mapping analysis calculated posterior probabilities of causalities (PPC) for individual variants and genes to identify candidates (Table 4), which were largely consistent with the GWAS results. A total of eight genes detected in GWAS signals were also successfully fine-mapped, including *PLXNA4* (DSAB), *FANCC* (DSAB), *NTM* (DSAB), *GC* (mastitis and livability), *ABCC9* (livability),

QDPR (RETP), *ZFAT* (livability), and *CCDC88C* (livability). In addition, fine-mapping identified new candidate genes, including *COBL* on BTA 4 for metritis, *LOC783947* on BTA 16 for ketosis, *LOC783493* on BTA 18 for RETP, and *LOC618463* on BTA 18 and *LOC101908667* on BTA 23 for livability. The genes *LOC107133096* on BTA 14 and *LOC100296627* on BTA 4 detected respectively for ketosis and RETP by fine mapping were close to two genes (*DGATI* and *ABCB1*) that have known biological association with milk production and other traits. In addition to the detected genes in these two cases, we further investigated genes with a potential biological link with disease, and genes with the highest PPC (*PARP10* and *MALSUI*) that were located between these two references (Table 4). No genes were detected by fine-mapping in the signal on BTA 6 for hypocalcemia (Figure 1), given that the nearest genes were beyond a 1 Mb window boundary.

In addition, we investigated the expression levels of fine-mapped candidate genes across cattle tissues using existing RNA-Seq data from public databases. While many genes are ubiquitously expressed in multiple tissues, several fine-mapped genes were specifically expressed in a few tissues relevant to cattle health (Table 4). Interesting examples of tissue-specific expression and candidate genes included liver with mastitis and livability (*GC*), and CD8 cells with livability (*CCDC88C*). Although this analysis is preliminary, these results provide additional support for these candidate genes of cattle health and help the understanding of how and where their expression is related with dairy disease resistance.

Discussion

In this study, we performed powerful GWAS analyses of seven health and related traits in Holstein bulls. The resulting GWAS signals were further investigated by a Bayesian fine-

mapping approach to identify candidate genes and variants. Additionally, we included tissue-specific expression data of candidate genes to reveal a potential biological relationship between genes, tissues and cattle diseases. Finally, we provide a list of candidate genes of cattle health with associated tissue-specific expression that can be readily tested in future functional validation studies.

In our GWAS analysis, we used de-regressed PTA as phenotype and incorporated the reliabilities of the de-regressed PTAs of livability and six disease traits. Three traits were found to have significant association signals, hypocalcemia, ketosis, and livability, which demonstrated the power of our GWAS study. For instance, we observed tissue-specific expression with hypocalcemia for body/carcass weight and reproduction traits that are supported by a previous study where hypocalcemia compromised immunosuppression in cows at calving [26]. Interestingly, this finding coincides with the incidence of hypocalcemia that occurs in human infants born to mothers who had complications during pregnancy or labor [27]. This suggests that the association signals related to hypocalcemia in these regions may be a common phenomenon across species. In addition to hypocalcemia, we also observed regions associated with livability, in particular, with the region around 58,194,319 on BTA 18 to possess a large effect on dairy and body traits. Our finding was corroborated by a BLAST analysis that identified a related molecule, Siglec-6, which is expressed in tissues such as the human placenta [28]. Further analyses can be performed to characterize the functional implications of these association regions for the seven health and related traits in cattle.

When using PTA values as phenotype in GWAS, we observed different regions to be associated, compared to the GWAS with de-regressed PTA (Figure 1 and Additional File 2). For example, a genomic region larger than 4 Mb on BTA 12 was associated with most of the health

traits (Additional File 2). Although these generally appeared as clear association signals, we observed only a few HD SNP markers to be associated, which may be due to poor imputation. Additionally, this region was reported by VanRaden et al., (2017) as having low imputation accuracy [29]. The lower imputation accuracy on BTA 12 was determined to be caused by a gap between the 72.4 and 75.2 Mb region where no SNPs were present on the HD SNP array [29]. Additional studies are needed to address this imputation issue in order to improve the accuracy and power of future analysis on this region. In sum, this comparison of GWAS using PTA and de-regressed PTA supports the use of de-regressed PTA values with reliabilities accounted for in future GWAS studies in cattle.

Application of BFMAP for fine-mapping allowed us to identify 20 promising candidate genes (Table 4) and a list of candidate variants (Additional File 3) for health traits in dairy cattle. We found that most of the genes possess tissue-specific expression, notably the detected gene *LOC107133096* on BTA 14 for ketosis. This gene is located close to the *DGATI* gene that affects milk fat composition. A previous candidate gene association study by Tetens et al. (2013) proposed *DGATI* to be an indicator of ketosis [30]. In that study, the *DGATI* gene was determined to be involved in cholesterol metabolism, which is known to be an indicator of a ketogenic diet in humans [30]. This result highlights a potential pathway in the pathogenesis of ketosis that may be an area for future research. Additionally, ketosis is a multifactorial disease that is likely influenced by multiple loci. Therefore, implementation of a functional genomics approach would allow identification of more genetic markers, and in doing so, improve resistance to this disease. For DSAB, the gene *PLXNA4* was observed to have an association with the variant 97,101,981 bp on BTA 4 (Table 4 and Additional File 3). Our analysis also detected tissue-specific expression for *PLXNA4* in the aorta. A previous study on atherosclerosis

found that Plexin-A4 knockout mice exhibited incomplete aortic septation [31]. In addition, it is worth noting that one of the known clinical signs of dairy cows with displaced abomasum is an increased heart rate. Taken together, these findings provide support for the association of *PLXNA4* with DSAB as we found in this study.

Six signals were observed as clear association peaks for livability (Figure 1). The associated variant at 8,144,774 – 8,305,775 bp on BTA 14 was close to the gene *ZFAT*, which is known to be expressed in the human placenta [32]. In particular, the expression of this gene is downregulated in placentas from complicated pregnancies. Additionally, a GWAS study performed in three French dairy cattle populations found the *ZFAT* gene to be the top variant associated with fertility [33]. These results lend support to our findings of this associated variant with the livability health trait. On BTA18, the associated variant at 57,587,990 – 57,594,549 bp was near the gene *LOC618463*, which has been previously identified as a candidate gene associated with calving difficulty in three different dairy populations [34]. For the associated variant at 56,645,629 – 56,773,438 bp on BTA21, it is located close to the *CCDC88C* gene (Table 4). In addition to our detection of tissue-specific expression with the CD8 cell, this gene has been associated with traits such as dairy form and days to first breeding in cattle [19]. Interestingly, in humans, mutation of the *CCDC88C* gene caused a recessive form of congenital hydrocephalus, which presents symptoms of an elevated CD4/CD8 ratio [35, 36].

It is notable that our GWAS signal for livability at 25,904,084 – 25,909,461 bp on BTA 23 is located in the bovine MHC region (Table 4). The gene we detected was *LOC101908667*, which is one of the immune genes of MHC. This is of considerable interest because MHC genes have a role in immune regulation. The MHC complex of cattle located on BTA 23 is called the bovine leukocyte antigen (*BoLA*) region. This complex of genes has been extensively studied,

such as in research investigating the polymorphism of genes in *BoLA* and their association with disease resistance [37]. Therefore, our research highlights a gene of considerable interest that should be further explored to understand its importance in breeding programs and its potential role in resistance to infectious diseases.

Additionally, we identified an associated variant for livability at 88,687,845 - 88,739,292 bp on BTA6 was close to the gene *GC*, which was specifically expressed in tissues such as the liver (Table 4). This gene has been previously studied in an association analysis that investigated the role of *GC* on milk production [17]. It found that the gene expression of *GC* in cattle is predominantly expressed in the liver. Moreover, affected animals displayed decreased levels of the vitamin D-binding protein (DBP) encoded by *GC*, highlighting the importance of *GC* for a cow's production. Additionally, liver-specific *GC* expression has been identified in humans, specifically regulated through binding sites for the liver-specific factor HNF1 [38]. Collectively, these results offer evidence for *GC* expression in the liver, which may be an important factor for determining cow livability.

Interestingly, the *GC* gene was also detected to have tissue-specific expression in the liver for mastitis (Table 4). This is corroborated by a study on cattle infected with mastitis to possess limited DBP concentration [17]. Vitamin D plays a key part in maintaining serum levels of calcium when it is secreted into the milk [39]. Since *GC* encodes DBP, it was suggested that the *GC* gene has a role in regulating milk production and the incidence of mastitis infection in dairy cattle. It is important to note that bovine mastitis pathogens, such as *Staphylococcus aureus* and *Escherichia coli*, also commonly occur as pathogens of humans. Therefore, development of molecular methods to contain these pathogens is of considerable interest for use in human medicine to prevent the spread of illness and disease. For instance, the use of enterobacterial

repetitive intergenic consensus typing enables trace back of clinical episodes of *E. coli* mastitis, thus allowing for an evaluation of antimicrobial products for the prevention of mastitis [40].

Continued investigation using molecular methods are needed to understand the pathogenesis of mastitis and its comparative relevance to human medicine. Based on the fine mapping for metritis, the new gene assigned was *COBL* on BTA 6 (Table 4). However, this candidate gene was found to have variants only passing the nominal significance level for causality and for GWAS. Further exploration of this candidate gene is needed to contribute to our understanding of its function and potential tissue-specific expression.

For RETP, the gene *TMEM182* was observed to have an association with a variant between 7,449,519 – 7,492,871 bp on BTA11 (Table 4). Our tissue-specific analysis identified *TMEM182* to have an association in muscle tissues. A study performed in Canchim beef cattle investigated genes for male and female reproductive traits and identified *TMEM182* on BTA 11 as a candidate gene that could act on fertility [41]. Additionally, the gene *TMEM182* has been found to be up-regulated in brown adipose tissue in mice during adipogenesis, which suggests a role in the development of muscle tissue [42]. One important factor that causes retention of fetal membranes in cattle is the impaired muscular tone of organs such as the uterus and abdomen [43]. This suggests the importance of the *TMEM182* gene and the need for future studies to better understand its role in the cattle breeding program.

Conclusions

In this study, we reported eight significant associations for seven health and related traits in dairy cattle. In total, we identified 20 candidate genes of cattle health with the highest

posterior probability, which are readily testable in future functional studies. Several candidate genes exhibited tissue-specific expression related to immune function, muscle growth and development, and neurological pathways. The identification of a novel association for cow livability in the bovine MHC region also represented an insight into the biology of disease resistance. Overall, our study offers a promising resource of candidate genes associated with complex diseases in cattle that can be applied to breeding programs and future studies of disease genes for clinical utility.

Methods

Ethics statement

This study didn't require the approval of the ethics committee, as no biological materials were collected.

Genotype data

Using 444 ancestor Holstein bulls from the 1000 Bull Genomes Project as reference, we previously imputed sequence variants for 27,214 progeny-tested Holstein bulls that have highly reliable phenotypes via FindHap version 3 [44]. The original 777,962 HD SNPs were reduced to 312,614 by removing highly correlated SNP markers with a $|r|$ value higher than 0.95 and by prior editing. Variants with a minor allelic frequency (MAF) lower than 0.01, incorrect map locations (UMD3.1 bovine reference assembly), an excess of heterozygotes, or low correlations ($|r| < 0.95$) between sequence and HD genotypes for the same variant were removed. The final imputed data was composed of 3,148,506 sequence variants for 27,214 Holstein bulls. Details about the genomic data and imputation procedure are described by VanRaden et al., (2017).

After imputation, we only retained autosomal variants with $MAF \geq 0.01$ and P -value of Hardy-Weinberg equilibrium test $>10^{-6}$.

Phenotype data

The data used were part of the 2018 U.S. genomic evaluations from the Council on Dairy Cattle Breeding (CDCB), consisting of 1,922,996 Holstein cattle from the national dairy cattle database. De-regressed predicted transmitting ability (PTA) values according to Garrick et al., (2009) for livability, hypocalcemia, displaced abomasum, ketosis, mastitis, metritis, and retained placenta were analyzed. We restricted the de-regression procedure to those bulls with PTA reliability greater than parent average reliability, thus reducing the total number of animals from 27,214 to 11,880, 13,229, 12,468, 14,382, 13,653, 13,541, and 24,699 for the seven traits, respectively (Table 1).

Genome-wide association study (GWAS)

A mixed model GWAS approach was performed using MMAP [45]. Thus, the additive effect was divided into a random polygenic effect and a fixed effect of the candidate SNP. The variance components were estimated using REML. The model can be generally presented as:

$$\mathbf{y} = \mu + \mathbf{m}\mathbf{b} + \mathbf{a} + \mathbf{e}$$

where \mathbf{y} is a vector with de-regressed PTAs; μ is the global mean; \mathbf{m} is the candidate SNP genotype (allelic dosage coded as 0, 1 or 2) for each animal; \mathbf{b} is the solution effect of the candidate SNP; \mathbf{a} is a solution vector of polygenic effect accounting for the population structure assuming $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$, where \mathbf{G} is a relationship matrix; and \mathbf{e} is a vector of residuals assuming $\mathbf{e} \sim N(0, \mathbf{R}\sigma_e^2)$, where \mathbf{R} is a diagonal matrix with diagonal elements weighted by the individual de-regressed reliability ($R_{ii} = 1/(r_i^2 - 1)$). For each candidate variant, a Wald test was applied to

evaluate the alternative hypothesis, $H_1: b \neq 0$, against the null hypothesis $H_0: b = 0$. Bonferroni correction for multiple comparisons was applied to control the type-I error rate. Gene coordinates in the UMD v3.1 assembly [46] were obtained from the *Ensembl Genes 90* database using the *BioMart tool* [47]. The *cattle QTLdb* database [16] was examined to check if any associated genomic region was previously reported as a cattle quantitative trait locus (QTL).

Fine-mapping association study

In order to identify potential candidate genes and their causal variants, GWAS signals were investigated through a fine-mapping procedure using a Bayesian approach with the software BFMAP v.1 [19]. Independent association signals within the same QTL region as well as a posterior probability of causality (PPC) to each variant and its causality *p-value* were assigned. The minimal region covered by each lead variant was determined as ± 1 Mb upstream and downstream (candidate region ≥ 2 Mb). This extension allowed the region to cover most variants that have an LD r^2 of >0.3 with the lead variants. The employed fine-mapping approach included three steps: forward selection to add independent signals in the additive Bayesian model, repositioning signals, and generating credible variant sets for each signal. Details about the BFMAP algorithm and its procedure are described by Jiang et al., (2019).

Tissue-specific expression of candidate genes

We have collected and analyzed RNA-seq data of 723 samples from 91 tissues and cell types in Holstein cattle from previous studies on NCBI GEO database. Details of the processing and analyses have been described in a submitted manuscript [48]. We processed all the 732 RNA-seq data uniformly using a rigorous bioinformatics pipeline with stringent quality control procedures. We then calculated the *t*-statistics for differential expression for each gene in each tissue in the RNA-seq dataset using a previous method [49]. Specifically, the log₂-transformed

expression (i.e., log2FPKM) of genes was standardized with mean of 0 and variance of 1 within each tissue or cell type,

$$y_i = \mu_i + x_{is} + x_{iage} + x_{isex} + x_{istudy} + e_i$$

where y_i is the standardized log2-transformed expression level (i.e., log2FPKM) of i th gene; μ_i is the overall mean of the i th gene; x_{is} is the tissue effect, where samples of the tested tissue were denoted as '1', while other samples as '-1'; x_{iage} , x_{isex} , x_{istudy} were age, sex, and study effects for the i th gene, respectively; e_i is residual effect. We fitted this model for each gene in each tissue using the ordinary least-square approach and then obtained the t-statistics for the tissue effect to measure the expression specificity of this gene in the corresponding tissue. Using this approach, we evaluated the expression levels for each of the candidate genes that were fine-mapped in this study across the 91 tissues and cell types and identified the most relevant tissue or cell type for a disease trait of interest.

List of abbreviations

GWAS: genome-wide association study. **HO:** Holstein. **SNP:** single nucleotide polymorphism.

QTL: quantitative trait locus. **BTA:** Bos taurus chromosome. **CALC:** hypocalcemia. **LD:**

Linkage disequilibrium. **MAF:** minor allelic frequency. **PPC:** probability of causality. **PTA:**

Predicted Transmitting Ability. **CALC:** hypocalcemia. **DSAB:** displaced abomasum. **KETO:**

Ketosis. **MAST:** Mastitis. **METR:** Metritis. **RETP:** retained placenta.

Declarations

Ethics approval and consent to participate

The Council on Dairy Cattle Breeding (CDCB) approved data access and the research, which was conducted on USDA-AGIL computers using the national database shared jointly by CDCB and AGIL under a non-funded cooperative agreement:

http://aipl.arsusda.gov/reference/CDCB_NFCA.pdf.

Consent for publication

Not applicable.

Availability of data and material

The original performance and pedigree data are owned by CDCB. A request to CDCB to access the data may be sent to: João Dürr, CDCB Chief Executive Officer (joao.durr@cdcb.us). Bull genotypes are controlled by the Collaborative Dairy DNA Repository (CDDR; Verona, WI), and a request to access those data must be made to Jay Weiker, CDDR Administrator (jweiker@naab-css.org). The bovine transcriptome data can be directly downloaded from NCBI GEO database with accession numbers SRP042639, PRJNA177791, PRJNA379574, PRJNA416150, PRJNA305942, SRP111067, PRJNA392196, PRJNA428884, PRJNA298914, PRJEB27455, PRJNA268096, and PRJNA446068. All other data and results are included in the published article.

Competing interests

The authors declare that they have no competing interests.

Funding

This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2016-67015-24886 and 2018-67015-28128 from the USDA National Institute of Food and Agriculture, MAES Competitive Grants from the Maryland Experimental Station 2017 and 2019, and the BARD Grant US-4997-17 from the US-Israel Binational Agricultural Research and Development Fund. JBC and PMV was supported by appropriated project 8042-31000-002-00-D, “Improving Dairy Animals by Increasing Accuracy of Genomic Prediction, Evaluating New Traits, and Redefining Selection Goals”, and GEL was supported by appropriated project 8042-31000-001-00-D, “Enhancing Genetic Merit of Ruminants Through Improved Genome Assembly, Annotation, and Selection”, of the Agricultural Research Service (ARS) of the United States Department of Agriculture. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors’ contributions

LM, JC, and CM conceived the study. EF, DS, LF, JJ analyzed and interpreted data. EF, DS, CM and LM wrote the manuscript. GEL, KPG, JC, PMV contributed tools and materials. All authors read and approved the final manuscript.

Acknowledgements

We thank the 1000 Bull Genomes Project for providing reference sequence data for imputation. The Council on Dairy Cattle Breeding (CDCB; Bowie, MD), Cooperative Dairy DNA Repository (Verona, WI), and dairy industry contributors are thanked for providing phenotypic, pedigree, and genomic data. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or

endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer.

References

1. Liang D, Arnold L, Stowe C, Harmon R, Bewley J: **Estimating US dairy clinical disease costs with a stochastic simulation model.** *Journal of dairy science* 2017, **100**(2):1472-1486.
2. Duffield T: **Subclinical ketosis in lactating dairy cattle.** *Veterinary clinics of north america: Food animal practice* 2000, **16**(2):231-253.
3. Seegers H, Fourichon C, Beaudeau F: **Production effects related to mastitis and mastitis economics in dairy cattle herds.** *Veterinary research* 2003, **34**(5):475-491.
4. Reinhardt TA, Lippolis JD, McCluskey BJ, Goff JP, Horst RL: **Prevalence of subclinical hypocalcemia in dairy herds.** *The Veterinary Journal* 2011, **188**(1):122-124.
5. Bartlett PC, Kirk JH, Wilke MA, Kaneene JB, Mather EC: **Metritis complex in Michigan Holstein-Friesian cattle: incidence, descriptive epidemiology and estimated economic impact.** *Preventive veterinary medicine* 1986, **4**(3):235-248.
6. Laven R, Peters A: **Bovine retained placenta: aetiology, pathogenesis and economic loss.** *Veterinary Record* 1996, **139**(19):465-471.
7. Ma L, Cole J, Da Y, VanRaden P: **Symposium review: Genetics, genome-wide association study, and genetic improvement of dairy fertility traits.** *Journal of dairy science* 2018.
8. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK: **Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows.** *BMC genomics* 2011, **12**(1):408.
9. Jiang J, Cole JB, Da Y, VanRaden PM, Ma L: **Fast Bayesian fine-mapping of 35 production, reproduction and body conformation traits with imputed sequences of 27K Holstein bulls.** *BioRxiv* 2018:428227.
10. Gallagher MD, Chen-Plotkin AS: **The post-GWAS era: from association to function.** *The American Journal of Human Genetics* 2018, **102**(5):717-730.
11. Fang L, Jiang J, Li B, Zhou Y, Freebern E, Vanraden PM, Cole JB, Liu GE, Ma L: **Genetic and epigenetic architecture of paternal origin contribute to gestation length in cattle.** *Communications biology* 2019, **2**(1):100.
12. Wright J, VanRaden P: **Genetic evaluation of dairy cow livability.** *Journal of Animal Science* 2016, **94**:178-178.
13. Parker Gaddis K, Tooker M, Wright J, Megonigal J, Clay J, Cole J, VanRaden P: **Development of national genomic evaluations for health traits in U.S. Holsteins.** *Proc 11th World Congr Genet Appl Livest Prod, Auckland, New Zealand, Feb 11–16 2018, Vol. Biol. & Species–Bovine (dairy) 1, p. 594.*
14. Garrick DJ, Taylor JF, Fernando RL: **Deregressing estimated breeding values and weighting information for genomic regression analyses.** *Genetics Selection Evolution* 2009, **41**(1):55.
15. TAKESHIMA SN, AIDA Y: **Structure, function and disease susceptibility of the bovine major histocompatibility complex.** *Animal Science Journal* 2006, **77**(2):138-150.
16. Hu Z-L, Park CA, Wu X-L, Reecy JM: **Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era.** *Nucleic acids research* 2012, **41**(D1):D871-D879.
17. Olsen HG, Knutsen TM, Lewandowska-Sabat AM, Grove H, Nome T, Svendsen M, Arnyasi M, Sodeland M, Sundsaasen KK, Dahl SR: **Fine mapping of a QTL on bovine chromosome 6 using**

- imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production.** *Genetics Selection Evolution* 2016, **48**(1):79.
18. Sahana G, Gulbrandsen B, Thomsen B, Lund MS: **Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle.** *Animal genetics* 2013, **44**(6):620-626.
 19. Jiang J, Cole JB, Freebern E, Da Y, VanRaden PM, Ma L: **Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls.** *Communications Biology* 2019, **2**(1):212.
 20. Wu X, Lund MS, Sahana G, Gulbrandsen B, Sun D, Zhang Q, Su G: **Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins.** *Genetics Selection Evolution* 2015, **47**(1):50.
 21. Nayeri S, Sargolzaei M, Abo-Ismael M, Miller S, Schenkel F, Moore S, Stothard P: **Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle.** *Journal of dairy science* 2017, **100**(2):1246-1258.
 22. Snelling W, Allan M, Keele J, Kuehn L, Mcdanel T, Smith T, Sonstegard T, Thallman R, Bennett G: **Genome-wide association study of growth in crossbred beef cattle.** *Journal of animal science* 2010, **88**(3):837-848.
 23. Pryce JE, Hayes BJ, Bolormaa S, Goddard ME: **Polymorphic regions affecting human height also control stature in cattle.** *Genetics* 2011, **187**(3):981-984.
 24. Nalaila S, Stothard P, Moore S, Li C, Wang Z: **Whole-genome QTL scan for ultrasound and carcass merit traits in beef cattle using Bayesian shrinkage method.** *Journal of Animal Breeding and Genetics* 2012, **129**(2):107-119.
 25. Olsen H, Hayes B, Kent M, Nome T, Svendsen M, Lien S: **A genome wide association study for QTL affecting direct and maternal effects of stillbirth and dystocia in cattle.** *Animal genetics* 2010, **41**(3):273-280.
 26. Martinez N, Risco C, Lima F, Bisinotto R, Greco L, Ribeiro E, Maunsell F, Galvão K, Santos J: **Evaluation of periparturient calcium status, energetic profile, and neutrophil function in dairy cows at low or high risk of developing uterine disease.** *Journal of dairy science* 2012, **95**(12):7158-7172.
 27. Gittleman I, Pincus J, Schmertzler E, Saito M: **Hypocalcemia occurring on the first day of life in mature and premature infants.** *Pediatrics* 1956, **18**:721-729.
 28. Cole J, VanRaden P, O'Connell J, Van Tassell C, Sonstegard T, Schnabel R, Taylor J, Wiggans G: **Distribution and location of genetic effects for dairy traits.** *Journal of dairy science* 2009, **92**(6):2931-2946.
 29. VanRaden PM, Tooker ME, O'Connell JR, Cole JB, Bickhart DM: **Selecting sequence variants to improve genomic predictions for dairy cattle.** *Genet Sel Evol* 2017, **49**(1):32.
 30. Tetens J, Seidenspinner T, Buttchereit N, Thaller G: **Whole-genome association study for energy balance and fat/protein ratio in German Holstein bull dams.** *Animal genetics* 2013, **44**(1):1-8.
 31. Toyofuku T, Yoshida J, Sugimoto T, Yamamoto M, Makino N, Takamatsu H, Takegahara N, Suto F, Hori M, Fujisawa H: **Repulsive and attractive semaphorins cooperate to direct the navigation of cardiac neural crest cells.** *Developmental biology* 2008, **321**(1):251-262.
 32. Barbaux S, Gascoin-Lachambre G, Buffat C, Monnier P, Mondon F, Tonanny M-B, Pinard A, Auer J, Bessières B, Barlier A: **A genome-wide approach reveals novel imprinted genes expressed in the human placenta.** *Epigenetics* 2012, **7**(9):1079-1090.
 33. Marete AG, Gulbrandsen B, Lund MS, Fritz S, Sahana G, Boichard D: **A meta-analysis including pre-selected sequence variants associated with seven traits in three French dairy cattle populations.** *Frontiers in genetics* 2018, **9**:522.

34. Purfield DC, Bradley DG, Evans RD, Kearney FJ, Berry DP: **Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle.** *Genetics Selection Evolution* 2015, **47**(1):47.
35. Drielsma A, Jalas C, Simonis N, Désir J, Simanovsky N, Pirson I, Elpeleg O, Abramowicz M, Edvardson S: **Two novel CCDC88C mutations confirm the role of DAPLE in autosomal recessive congenital hydrocephalus.** *Journal of medical genetics* 2012, **49**(11):708-712.
36. Tabuchi S, Uno T: **Hydrocephalus with panventricular enlargement as the primary manifestation of neurosarcoidosis: a case report.** *Journal of medical case reports* 2013, **7**(1):240.
37. Gowane G, Vandre R, Nangre M, Sharma A: **Major histocompatibility complex (MHC) of bovines: an insight into infectious disease resistance.** *Livestock Research International* 2013, **1**(2):46-57.
38. Hiroki T, Liebhaber SA, Cooke NE: **An intronic locus control region plays an essential role in the establishment of an autonomous hepatic chromatin domain for the human vitamin D-binding protein gene.** *Molecular and cellular biology* 2007, **27**(21):7365-7380.
39. Horst R, Goff J, Reinhardt T: **Role of vitamin D in calcium homeostasis and its use in prevention of bovine periparturient paresis.** *ACTA VETERINARIA SCANDINAVICA SUPPLEMENTUM* 2003, **97**:35-50.
40. Zadoks RN, Middleton JR, McDougall S, Katholm J, Schukken YH: **Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans.** *Journal of mammary gland biology and neoplasia* 2011, **16**(4):357-372.
41. Buzanskas ME, do Amaral Grossi D, Ventura RV, Schenkel FS, Chud TCS, Stafuzza NB, Rola LD, Meirelles SLC, Mokry FB, de Alvarenga Mudadu M: **Candidate genes for male and female reproductive traits in Canchim beef cattle.** *Journal of animal science and biotechnology* 2017, **8**(1):67.
42. Wu Y, Smas CM: **Expression and regulation of transcript for the novel transmembrane protein Tmem182 in the adipocyte and muscle lineage.** *BMC research notes* 2008, **1**(1):85.
43. Schlafer D, Fisher P, Davies C: **The bovine placenta before and after birth: placental development and function in health and disease.** *Animal reproduction science* 2000, **60**:145-160.
44. VanRaden PM, Sun C: **Fast Imputation Using Medium- or Low-Coverage Sequence Data.** *Proceedings, 10th World Congress of Genetics Applied to Livestock Production* 2014.
45. O'Connell JR: **MMAF User Guide.** Available: <http://edn.som.umaryland.edu/mmap/index.php>. Accessed 8 October 2015. 2015.
46. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS: **A whole-genome assembly of the domestic cow, *Bos taurus*.** *Genome biology* 2009, **10**(4):R42.
47. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A: **Ensembl BioMarts: a hub for data retrieval across taxonomic space.** *Database* 2011, **2011**.
48. Fang L, Cai W, Liu SL, Jiang J, Li B, Schroder SG, Rosen BD, Li C, Sonstegard TS, Van Tassell CP *et al*: **Comprehensive Analyses of 723 Transcriptomes Enhance GWAS Biological Interpretation and Genomic Prediction for Complex Traits in Cattle.** (Submitted).
49. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh P-R, Lareau C, Shores N: **Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types.** *Nature genetics* 2018, **50**(4):621.

Tables

Table 1. Number of Holstein bulls, reliability of PTA, and heritability (h^2) for six disease traits and cow livability.

Trait	N	h^2	Average Reliability
Hypocalcemia	11,880	0.006	0.228
Displaced Abomasum	13,229	0.011	0.269
Ketosis	12,468	0.012	0.260
Mastitis	14,382	0.031	0.338
Metritis	13,653	0.014	0.281
Retained Placenta	13,541	0.001	0.266
Livability	24,699	0.040	0.397

Table 2. Top SNPs and candidate genes associated with hypocalcemia (CALC), displaced abomasum (DSAB), ketosis (KETO), mastitis (MAST), metritis (METR), retained placenta (RETP) and cow livability.

Trait	Chr	Position	MAF	P-value	Genes Nearby	Traits Previously Associated#
CALC	6	10,521,824	0.014	8.3×10^{-10} *	<i>TRAM1L1</i> , <i>NDST4</i>	Subcutaneous fat
DSAB	4	97,101,981	0.021	4.4×10^{-7}	<i>PLXNA4</i> , <i>CHCHD3</i>	Milk protein yield
DSAB	8	83,052,202	0.109	1.3×10^{-7}	<i>FANCC</i>	Stature
DSAB	29	35,977,236	0.073	1.3×10^{-7}	<i>NTM</i>	Milk kappa-casein percentage
KETO	14	2,762,595	0.033	1.8×10^{-9} *	<i>LY6K</i>	Milk protein percentage
KETO	16	7,048,452	0.019	1.9×10^{-8}	<i>KCNT2</i>	Milk fat percentage
MAST	6	88,868,886	0.460	4.2×10^{-8}	<i>GC</i>	Clinical mastitis
METR	4	3,662,486	0.011	2.7×10^{-7}	<i>RF00322</i>	Milk protein yield
RETP	4	32,578,298	0.218	7.4×10^{-7}	<i>RUNDC3B</i>	Calving ease
RETP	6	117,620,548	0.026	7.2×10^{-7}	<i>QDPR</i>	Milk kappa-casein percentage
RETP	11	7,465,110	0.060	9.1×10^{-8}	<i>TMEM182</i>	Abomasum displacement
RETP	18	64,492,219	0.012	1.6×10^{-7}	<i>ZFP28</i>	Still birth
Livability	5	88,823,164	0.472	1.5×10^{-10} *	<i>ABCC9</i>	Productive life
Livability	6	88,801,999	0.454	1.7×10^{-18} *	<i>GC</i>	Clinical mastitis
Livability	14	8,536,538	0.020	5.3×10^{-10} *	<i>ZFAT</i>	Productive life
Livability	18	58,194,319	0.075	1.1×10^{-20} *	<i>ZNF614</i>	Bovine respiratory disease
Livability	21	56,700,449	0.013	8.6×10^{-11} *	<i>CCDC88C</i>	Type
Livability	23	26,131,593	0.017	3.8×10^{-9} *	<i>BLA-DQB</i>	Antibody-mediated immune response

*Genome-wide significance after Bonferroni correction

#Information obtained from the Animal QTLdb for cattle [16]

Table 3. Association results of the top SNPs associated with cow livability for hypocalcemia, displaced abomasum (DSAB), ketosis, mastitis, and metritis. *P*-values larger than 0.05 and their Beta coefficients were excluded.

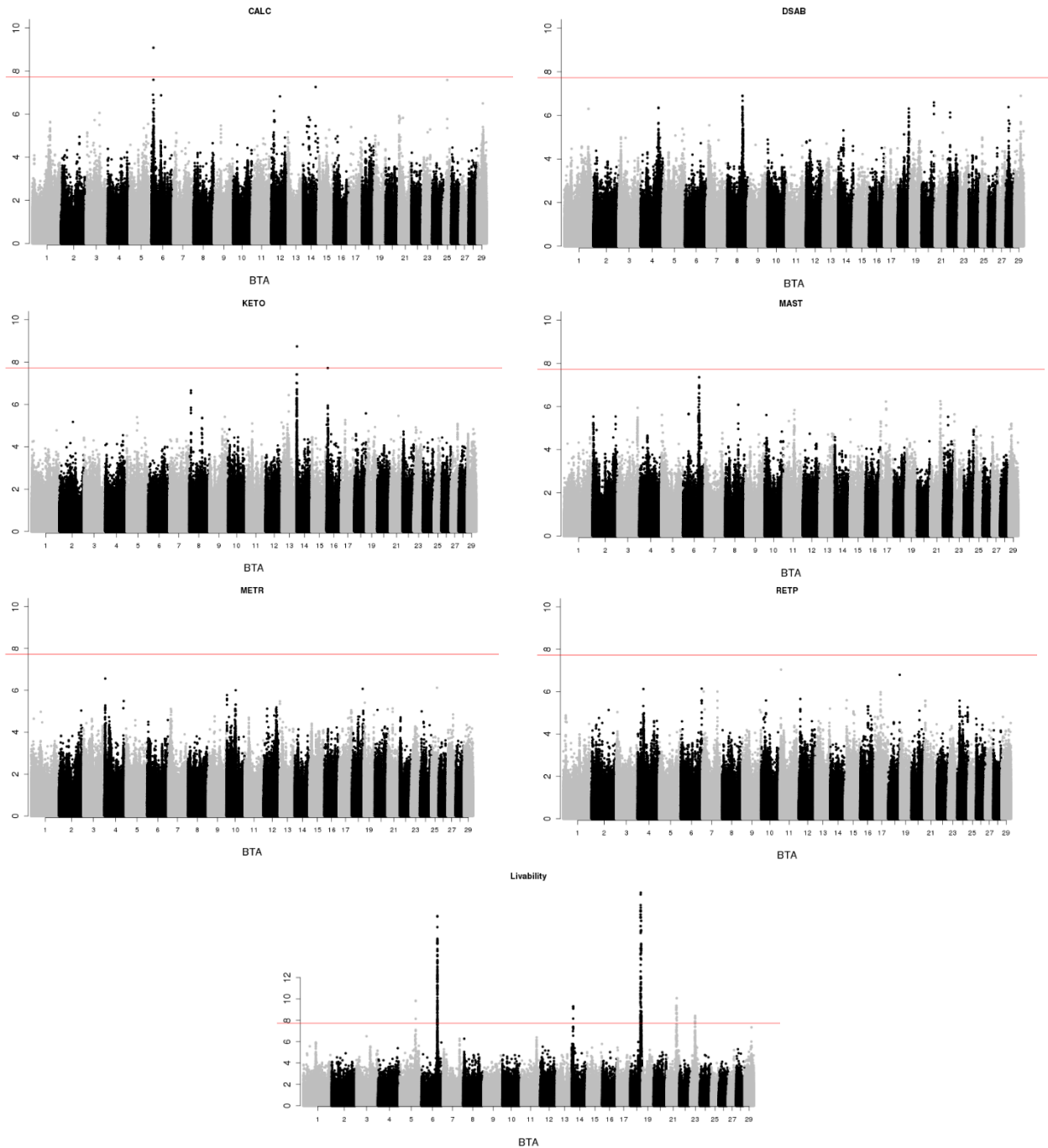
Chr	Position	Livability		Hypocalcemia		DSAB		Ketosis		Mastitis		Metritis	
		<i>P</i> -value	Beta	<i>P</i> -value	Beta	<i>P</i> -value	Beta	<i>P</i> -value	Beta	<i>P</i> -value	Beta	<i>P</i> -value	Beta
5	88,823,164	1.5×10 ⁻¹⁰	-0.43	-	-	0.04	-0.14	0.04	-0.21	-	-	-	-
6	88,801,999	1.7×10 ⁻¹⁸	-0.66	5.0×10 ⁻³	-0.2	-	-	2.1×10 ⁻³	-0.35	4.2×10 ⁻⁷	-0.75	-	-
14	8,536,538	5.3×10 ⁻¹⁰	-1.1	-	-	-	-	-	-	-	-	-	-
18	58,194,319	1.1×10 ⁻²⁰	-1.0	-	-	1.1×10 ⁻⁴	-0.47	-	-	-	-	0.01	-0.51
21	56,700,449	8.6×10 ⁻¹¹	-1.5	0.03	-0.58	-	-	-	-	9.1×10 ⁻³	-1.43	-	-
23	26,131,593	3.8×10 ⁻⁹	0.71	-	-	-	-	-	-	-	-	-	-

Table 4. List of candidate genes with highest posterior probability of causality (PPC) and their minimum *P*-values for casualty (M_Casualty) and GWAS (M_GWAS) associated with hypocalcemia (CALC), displaced abomasum (DSAB), ketosis (KETO), mastitis (MAST), metritis (METR), retained placenta (RETP) and cow livability and their tissue specific expression.

Trait	Gene	Chr	Start	End	M_GWAS P-value	M_Casualty P-value	PPC	Type	Tissue-Specific Expression
DSAB	<i>PLXNA4</i>	4	96,574,369	97,120,718	4.5×10^{-7}	6.5×10^{-7}	0.49	protein_coding	Aorta, Liver, Trachea
DSAB	<i>FANCC</i>	8	83,022,522	83,228,696	1.3×10^{-7}	2.1×10^{-9}	0.79	protein_coding	Thyroid
DSAB	<i>NTM</i>	29	35,153,012	36,117,726	1.3×10^{-7}	1.7×10^{-7}	0.99	protein_coding	Central Nervous System
KETO	<i>LOC107133096</i>	14	2,760,093	2,762,878	2.0×10^{-9}	5.9×10^{-8}	0.92	lncRNA	-
	<i>PARP10</i>	14	2,024,509	2,031,477	7.0×10^{-7}	1.7×10^{-5}	0.16	protein_coding	-
KETO	<i>DGATI</i>	14	1,795,425	1,804,838	1.0×10^{-6}	1.7×10^{-5}	0.08	protein_coding	Bone Marrow
	<i>LOC783947</i>	16	7,050,445	7,055,021	1.9×10^{-8}	1.3×10^{-8}	1.00	lncRNA	-
MAST	<i>GC</i>	6	88,687,845	88,739,292	2.0×10^{-7}	1.2×10^{-7}	0.15	protein_coding	Kidney, Cortex, Liver
METR	<i>COBL</i>	4	4,494,925	4,795,904	4.3×10^{-3}	7.7×10^{-4}	1.00	protein_coding	-
	<i>LOC100296627</i>	4	32,573,079	32,613,237	7.6×10^{-7}	4.0×10^{-13}	1.00	protein_coding	-
RETP	<i>MALSU1</i>	4	32,051,590	32,077,036	7.5×10^{-4}	1.1×10^{-13}	0.98	protein_coding	-
	<i>ABCB1</i>	4	33,013,208	33,095,708	6.3×10^{-1}	8.4×10^{-3}	0.28	protein_coding	-
RETP	<i>TMEM182</i>	11	7,449,519	7,492,871	9.0×10^{-8}	9.9×10^{-8}	0.96	protein_coding	Heart, Muscle, Tongue
RETP	<i>LOC783493</i>	18	63,799,608	63,803,213	8.3×10^{-3}	1.2×10^{-5}	0.94	Pseudogene	-
Livability	<i>ABCC9</i>	5	8,867,2047	88,834,491	1.5×10^{-10}	1.5×10^{-10}	1.00	protein_coding	Aorta, Atrium, Lung, Muscle Uterine myometrium, Ventricle
Livability	<i>GC</i>	6	88,687,845	88,739,292	1.9×10^{-17}	1.4×10^{-19}	0.03	protein_coding	Kidney, Cortex, Liver
Livability	<i>ZFAT</i>	14	8,144,774	8,305,775	2.1×10^{-5}	3.2×10^{-5}	0.23	protein_coding	-
Livability	<i>LOC618463</i>	18	57,587,990	57,594,549	1.7×10^{-20}	3.1×10^{-20}	0.20	protein_coding	-
Livability	<i>CCDC88C</i>	21	56,645,629	56,773,438	8.6×10^{-11}	8.9×10^{-11}	0.95	protein_coding	CD8_cell
Livability	<i>LOC101908667</i>	23	25,904,084	25,909,461	2.1×10^{-8}	7.9×10^{-9}	0.31	lncRNA	-

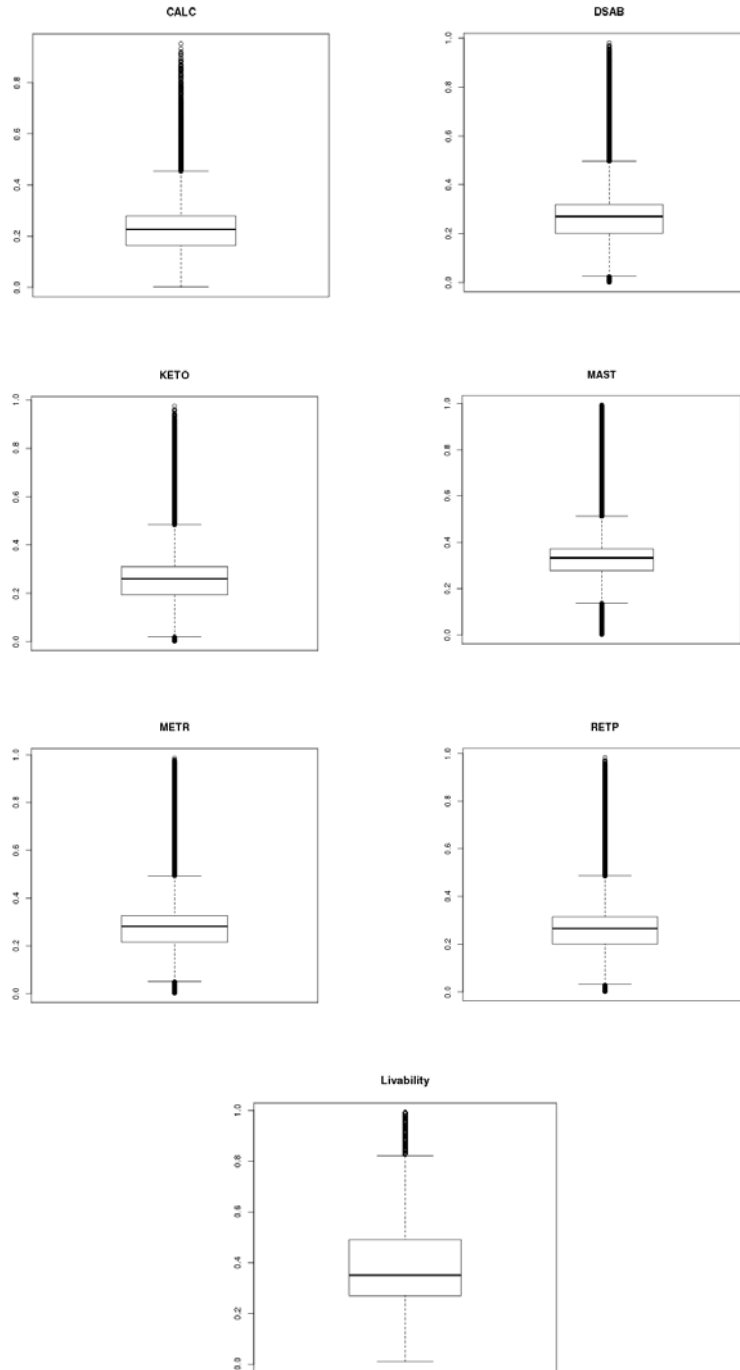
Figures

Figure 1. Manhattan plots for hypocalcemia (CALC), displaced abomasum (DSAB), ketosis (KETO), mastitis (MAST), metritis (METR), retained placenta (RETP) and cow livability. The genome-wide threshold (red line) corresponds to the Bonferroni correction.

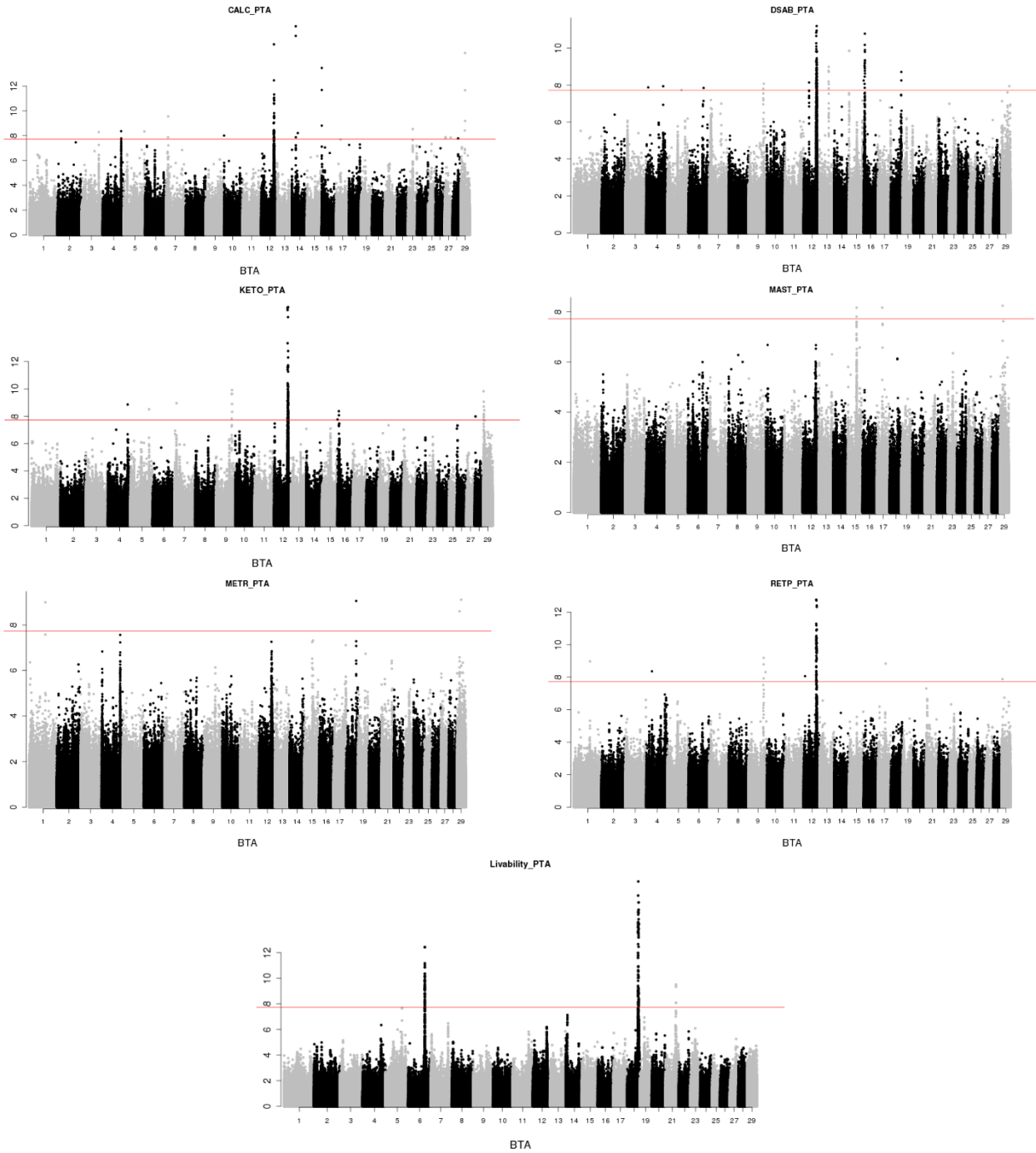


Additional Files

Additional File 1: Boxplot with PTA reliability for hypocalcemia (CALC), displaced abomasum (DSAB), ketosis (KETO), mastitis (MAST), metritis (METR), retained placenta (RETP) and cow livability.



Additional File 2: Manhattan plots using the PTA as phenotype for hypocalcemia (CALC), displaced abomasum (DSAB), ketosis (KETO), mastitis (MAST), metritis (METR), retained placenta (RETP) and cow livability. The genome-wide threshold (red line) corresponds to the Bonferroni correction for a nominal P -value $\alpha = 0.05$.



Additional File 3: List of variants into genes with highest posterior probability of causality mostly associated with displaced abomasum (DSAB), ketosis (KETO), mastitis (MAST), metritis (METR), retained placenta (RETP) and cow livability.

Trait	Gene	Variant	MAF	GWAS <i>P</i> -value	Causality <i>P</i> -value	N_PPC	Annotation
DSAB	<i>PLXNA4</i>	4:97101981	0.02	4.5×10^{-7}	6.5×10^{-7}	0.26	Intergenic
DSAB	<i>FANCC</i>	8:83052202	0.11	1.3×10^{-7}	2.1×10^{-9}	0.19	Intron
DSAB	<i>FANCC</i>	8:83083934	0.11	1.3×10^{-7}	2.1×10^{-9}	0.19	Intron
DSAB	<i>FANCC</i>	8:83123780	0.11	1.3×10^{-7}	2.1×10^{-9}	0.19	Intron
DSAB	<i>NTM</i>	29:35977236	0.07	1.3×10^{-7}	1.7×10^{-7}	0.91	Intergenic
KETO	<i>LOC107133096</i>	14:2762595	0.03	2.0×10^{-9}	5.9×10^{-8}	0.89	Upstream/Intergenic
		14:2025096	0.23	7.0×10^{-7}	1.7×10^{-5}	0.02	Missense/Downstream
	<i>PARP10</i>	14:2026646	0.23	7.0×10^{-7}	1.7×10^{-5}	0.02	Upstream/Intron
		14:1802266	0.22	1.0×10^{-6}	2.1×10^{-5}	0.02	Missense/Downstream
	<i>DGATI</i>	14:1804647	0.22	1.0×10^{-6}	2.1×10^{-5}	0.02	Downstream
MAST	<i>GC</i>	6:88718227	0.45	2.0×10^{-7}	1.2×10^{-7}	0.13	Intron
METR	<i>COBL</i>	4:4643092	0.13	4.3×10^{-3}	7.7×10^{-4}	0.04	Intron
RETP	<i>LOC100296627</i>	4:32578298	0.22	7.6×10^{-7}	4.0×10^{-13}	1	Intron
		4:32057434	0.17	7.5×10^{-4}	1.1×10^{-13}	0	Upstream/Intergenic
RETP	<i>ABCB1</i>	4:33063807	0.12	6.3×10^{-1}	8.4×10^{-3}	0	Intron
RETP	<i>QDPR</i>	6:117727506	0.024	2.2×10^{-4}	2.3×10^{-5}	0	3_prime_UTR_variant
RETP	<i>QDPR</i>	6:117727635	0.024	2.2×10^{-4}	2.3×10^{-5}	0	3_prime_UTR_variant
RETP	<i>QDPR</i>	6:117727851	0.024	2.2×10^{-4}	2.3×10^{-5}	0	3_prime_UTR_variant
RETP	<i>QDPR</i>	6:117728023	0.024	2.2×10^{-4}	2.3×10^{-5}	0	Missense
RETP	<i>QDPR</i>	6:117743248	0.024	2.2×10^{-4}	2.3×10^{-5}	0	Missense
RETP	<i>TMEM182</i>	11:7465110	0.06	9.0×10^{-8}	9.9×10^{-8}	0.95	Intron
RETP	<i>LOC783493</i>	18:63802274	0.24	8.3×10^{-3}	2.3×10^{-3}	0	Intron
Livability	<i>ABCC9</i>	5:88823164	0.47	1.5×10^{-10}	1.5×10^{-10}	0.51	Splice/Intron
Livability	<i>GC</i>	6:88718227	0.45	1.9×10^{-17}	1.4×10^{-19}	0.03	Intergenic
Livability	<i>ZFAT</i>	14:8249314	0.04	2.1×10^{-5}	3.0×10^{-5}	0	Intron
Livability	<i>LOC618463</i>	18:57589121	0.07	1.7×10^{-20}	3.1×10^{-20}	0.19	Intron
Livability	<i>CCDC88C</i>	21:56700449	0.01	8.6×10^{-11}	8.9×10^{-11}	0.46	Intron
Livability	<i>LOC101908667</i>	23:25905667	0.02	2.1×10^{-8}	7.9×10^{-9}	0.08	Intergenic

N_PPC = Normalized posterior probability of causality