Genome-wide associations for udder and teat conformational risk factors for mastitis in Holstein cows

Asha M. Miles
Cornell University College of Agriculture and Life Sciences
Christian Posbergh
Cornell University College of Agriculture and Life Sciences
Heather Jay Huson (hjh3@cornell.edu)
Cornell University College of Agriculture and Life Sciences https://orcid.org/0000-0001-8299-0447

Research article

Keywords: genome wide association, mastitis, principal component analysis, udder conformation, teat conformation

Posted Date: November 4th, 2019

DOI: https://doi.org/10.21203/rs.2.16763/v1

Abstract

BACKGROUND The objective of our study was to conduct high-density genome-wide association studies of dairy cow udder and teat conformation with direct phenotyping. We identified and compared quantitative trait loci (QTL) for a novel composite mastitis risk trait and considered environmental impact of milking by comparing primiparous cows only. Cows (N = 471) were genotyped on the Illumina BovineHD 777K beadchip and scored for front and rear teat length, width, end shape, and placement, fore udder attachment, udder cleft, udder depth, rear udder height, and rear udder width. Principal component analysis was performed on fore udder attachment, rear teat end shape, rear teat width, and rear udder height, to create a single new phenotype describing mastitis susceptibility based on these high-risk traits.

RESULTS Over all 14 traits of interest, a total of 56 genome-wide associations were performed and 28 significantly associated (Bonferroni multiple testing correction < 0.05) QTL were identified. The linkage disequilibrium (LD) block surrounding the associated QTL or a 1 Mb window in the absence of LD was interrogated for candidate genes, resulting in the identification of genes with functions related to both cell proliferation and immune signaling, including ZNF683, DHX9, CUX1, TNNT1 , and SPRY1 . We assessed a primiparous only subset of cows (n = 144) to account for the possibility that the genetic variance component of the phenotype is greater for cows who have had less exposure to the environment, and observed different associated QTL and inheritance patterns for udder depth in primiparous cows compared to the total cohort.

CONCLUSION Special focus was given to the aforementioned mastitis risk traits, and candidate gene investigation revealed both immune function and cell proliferation related genes in the areas surrounding significantly associated QTL, suggesting that selecting for mastitis resistant cows based on these traits would be an effective method for increasing mastitis resiliency in a herd.

Background

Mastitis, a condition characterized by inflamed mammary tissue and udder gland, is the costliest disease facing U.S. dairy producers, accounting for an estimated $2 billion in annual losses and 11% of total milk lost according to a recent market analysis (1). While mastitis has been historically considered a management problem, genetic correlations among milk yield, mastitis susceptibility, and udder morphology encouraged selection for udder and teat type traits as early as the 1950s (2). In 2009, udder composite values were incorporated into official national genomic evaluation systems to account for the influence of cow conformation on health traits (3, 4). However, the lack of standardized reporting in the United States has led to a deficit in health-related phenotypes, hindering genetic improvement in U.S. dairy herds, and there is little consensus in the literature regarding mastitis and udder and teat trait relationships or their respective heritabilities (5, 6). Most extant research has relied on pedigree information to calculate relationship matrices for estimation of heritability, genetic correlation, and variance, given the strong correlations between health, production, and conformation traits (7-10). Focus has been also given to sire transmitting abilities, and while these studies comprehensively evaluate udder

morphology, teat length, and teat placement, other teat characteristics such as width and shape are neglected (8, 11).

As genotyping technologies have become increasingly cost-effective more molecular-based studies are emerging. A large study utilized publically available whole genome sequences to impute medium and high-density single nucleotide polymorphism (**SNP**) data up to over 20 million sequence variants to identify quantitative trait loci (**QTL**) for a number of udder and teat type traits in Fleckvieh cattle (12). While Pausch et al. had extremely high density genetic data, they relied on bull genotypes and the indirect phenotyping of their daughters, raising external validity concerns. Currently, only one study has directly associated udder traits with cow genotypes, relating beef cow teat length, teat diameter, and a composite "udder support" score to genomic data from the low-density Bovine Illumina50k chip (13). Large-scale studies using genotype imputation from low density SNP arrays may maximize genome coverage and make large sample sizes more practical, but they introduce bias towards population averages and limit researcher ability to detect rare variants (14).

A high-density genome wide association (**GWA**) study of dairy cow udder and teat conformation with direct phenotyping and no reliance on genotype imputation has yet to be conducted. Thus, we previously associated udder and teat conformational traits with mastitis in a prospective cohort study, where flat teat ends and loose fore udder attachment significantly increased odds of both elevated somatic cell count and clinical mastitis, and low rear udder height and wider rear teats increased odds of clinical mastitis diagnosis alone (15). We posited that these risk factors may be effective criteria to include in culling protocols or inform mating strategies by selecting bulls with strong evaluations for udder and teat morphology without sacrificing milk yield. The purpose of this current study was to use GWA approaches to identify SNP markers associated with udder and teat type traits, with special focus on these four risk traits for incorporation into genomic selection marker panels for mastitis-resistant cows.

Results

Principal Component Analysis. Principal component (**PC**) analysis was used to create novel composite traits for udder conformation, and mastitis risk factors. Udder traits-only PC1 accounted for 41% of the variance in phenotype and described over-all udder size, loading towards deeper, lower, and wider rear udders, as well as looser fore udder attachment. Risk traits-only PC1 accounted for 35% of phenotypic variance, loading in the direction of the low risk phenotypes of thinner rear teats, tighter fore udders, and higher rear udders (Figure 1a). No other PCs were informative or considered for GWA.

Genome-wide association. After all quality assurance measures were applied, 458 cows with 581,663 SNPs remained for analysis. Phenotypes were first considered quantitatively to assess continuous variation in the trait, then on a case-control basis comparing extremes in morphology. Primiparous cows (n = 144) were only evaluated linearly to preserve the sample size and due to low representation of many udder and teat morphologies in first parity cows (15). Of the 56 GWA studies performed, 13 had significantly associated SNPs, minimally passing a multiple testing correction threshold of false

discovery rate (**FDR**) < 0.05 and no evidence of genomic inflation (Table 1). Quantile-Quantile plots are presented in Supplementary Figure 1. All QTL positions and candidate genes residing within an LD block or 1 Mb surrounding associated SNPs are reported in Supplementary Table 1. No significant associations (FDR < 0.05) with udder PC1, front teat end shape, rear teat placement, front teat placement, or udder cleft were identified in our total cohort of cows (N = 458). In the primiparous cow subset (N = 144), significant associations (FDR < 0.05) were identified for front teat placement and udder depth only.

Risk PC1. A linear GWA with dominant inheritance and covariates of farm, parity, rear teat length, udder depth, and udder width significantly associated 1 SNP on *Bos taurus* autosome (**BTA**) 15 (Bonferroni < 0.05; Figure 1b).

Udder traits. An additive mixed model with no covariates comparing loose versus tight fore udder attachment significantly associated 1 SNP on BTA 2 and 2 SNPs on the X chromosome passing Bonferroni correction and 1 additional SNP on BTA 2 passing FDR (Figure 2). Significant associations were made for udder depth in both the total cohort and primiparous subset populations of cows. In a case-control comparison of deep versus high udders among all cows, a dominant mixed model with parity and udder width as fixed effects significantly associated 1 SNP on BTA 5 passing both FDR and Bonferroni correction (Figure 3a). In a linear scoring of udder depth among the subset of primiparous cows only, a recessive mixed model significantly associated 10 SNPs on BTA 17 (FDR < 0.05; Figure 3b). In a case-control comparison of low udder heights to all others (intermediate and high udders combined), a recessive mixed model with covariates udder depth, udder width, and fore udder attachment identified 129 SNPs which passed FDR on 20 chromosomes spanning the genome, and 32 SNPs passing Bonferroni correction on BTAs 6, 14, 15, 18, and 22 (Figure 4). In a case-control comparison of wide versus all other widths (intermediate and narrow combined), a dominant mixed model with no covariates significantly associated 1 SNP passing both FDR and Bonferroni correction on BTA 15.

Teat Traits. In a quantitative representation of front teat length, a recessive mixed model with farm, parity, front teat width, and rear teat length and width covariates significantly associated 5 SNPs passing FDR on BTAs 9, 10, and 12, and 2 SNPs passing Bonferroni correction on BTA 10. The most appropriate model for a case-control comparison of wide versus narrow front teats was a recessive mixed model with farm, parity, and front teat length included as fixed effects, yielding 4 SNPs passing Bonferroni correction on BTA 23. A case-control comparison of short versus long rear teats significantly associated (Bonferroni < 0.05) 1 SNP on BTA 2 in an additive mixed model with no covariates. In a case-control comparison of flat versus pointed rear teat ends, a recessive mixed model significantly associated (Bonferroni < 0.05) 1 SNP on BTA 26 (Figure 5). In an examination of rear teat width, two recessive mixed models with no covariates yielded significantly associated SNPs. In the first model, a case-control comparison of narrow versus wide rear teats yielded 3 SNPs passing FDR on BTAs 18, 25, and 28, and 1 SNP passing Bonferroni correction on BTA 25 (Figure 6a). Rear teat width was represented quantitatively in the second model and 50 SNPs were significantly associated (FDR < 0.05) on BTAs 1, 4, 5, 10, 11, 15, 16, 18, 19, 25, 26, 27, and the X chromosome. A total of 23 SNPs passed Bonferroni correction on BTAs 10, 11, 16, 18, 19, and 25

(Figure 6b). Examining primiparous cows only, a linear representation of front teat placement with a recessive mixed model yielded 10 SNPs passing FDR on BTA 9.

Discussion

We performed genome-wide association studies to associate QTL with udder and teat conformation traits, with special focus on previously identified mastitis-risk traits of fore udder attachment, rear teat end shape, rear teat width, and udder height (15). These studies provide insight into the genetic regulation of teat and udder confirmation and mastitis susceptibility as well as novel traits and specific markers that may be used in genetic selection. A comprehensive list of candidate genes may be found in Supplementary Table 1; here we discuss those most biologically relevant to mastitis risk associated traits.

Our goal in performing PCA on these risk traits was to identify a single measure which may describe mastitis risk based on udder and teat conformation. Risk PC1 described this risk (excepting rear teat end shape), and GWA for this new composite measure significantly associated novel QTL (15:7287030-7311314) not identified by individual assessment of those risk traits (Figure 1). Candidate genes identified in this region were related to both cell division (*Centrosomal Protein 126 (CEP126)*) and immune cell progenitor differentiation (*Angiopoietin Like 5 (ANGPTL5)*), suggesting that Risk PC1 does indeed reflect both mastitis and udder and teat morphology (16, 17).

In this study cohort, we previously associated loose fore udder attachment with high odds of elevated milk somatic cell count and clinical mastitis diagnosis, making fore udder attachment among these cows a relevant criterion on which to base culling and management decisions for mastitis control (Miles et al., 2019). A case-control GWA for extremes in fore udder attachment (loose versus tight) identified a number of genes related to both immune function and cell proliferation near associated markers at 2:126359098-126364670 (Figure 2). *Wiskott-Aldrich Syndrome Protein Family 2 (WASF2)* is a cytoplasmic protein implicated in cell migration, phagocytosis, and immune synapse formation (18). Similarly, *Nuclear Distribution C (NUDC)* is critical to cytokinesis, *Stratifin (SFN)* may regulate cell cycle progression, and *Keratinocyte Differentiation Factor 1 (KDF1)* serves as an essential regulator of epidermis formation (19-21). In contrast, this same genomic region houses genes related to immune function, including *Ficolin 3 (FCN3)* which is an essential component of the lectin complement pathway, *Ribosomal Protein S6 Kinase A1 (RPS6KA1)* which has been implicated in activated Toll Like Receptor 4 signaling, *High Mobility Group Nucleosomal Binding Domain 2 (HMGN2)* which is known to have antimicrobial activity against bacteria, viruses, and fungi, as well as *Zinc Finger 683 (ZNF683)*, a tissue-resident T-cell transcription regulator (22-25). The variety in function among candidate genes associated with fore udder attachment, relating to both the physical trait and immune function, reinforce the role of this udder type trait as a major risk factor for mastitis.

Similarly, we previously associated rear teat end shape with increased odds of both elevated somatic cell count and clinical mastitis in this cohort of cows, and in this study investigated the potential genetic

regulation of this trait (Miles et al., 2019). Upon examining rear teat end shape (Figure 5), a similar candidate gene pattern emerged with gene functions related to both cell division (explaining teat morphology variation) and immune response regulation (reinforcing teat end shape as an appropriate indicator of mastitis risk). Significantly associated SNP BovineHD2600014687 positioned at 26:50630351 resides within *Kinase Non-Catalytic C-Lobe Domain Containing 1 (KNDC1)*, which has been implicated in the regulation of cellular senescence and cell cycle progression (26). In addition, *Adhesion G Protein-Coupled Receptor A1 (ADGRA1)* was identified within a 1 Mb window of this SNP and belongs to a family of receptors known to regulate immune signaling (27). A pattern of candidate gene function emerged like that of fore udder attachment, reinforcing that it may be appropriate to select for mastitis-resilient cows based on udder and teat conformational traits.

In this cohort of cows, we previously associated increasing rear teat width with increased odds of clinical mastitis (Miles et al., 2019). We first assessed this trait quantitatively to account for continuous variation in rear teat width, and a linear GWA for rear teat width also revealed candidate genes related to both cell division and immune function at a number of different QTL spanning the genome (Figure 6). Near a significantly associated SNP at BTA 11:104129366, we identified *Caspase Recruitment Domain-Containing Protein 9 (CARD9)*, a key modulator of immune response related to TLR and NOD2 signaling pathways as well as NF-kB activation (28, 29). Furthermore, genes related to cell differentiation and survival were identified including *Notch Receptor 1 (NOTCH1)*, a highly conserved protein with an extracellular domain containing many epidermal growth factor repeats and whose signaling is heavily involved in cell fate specification, and *Epidermal Growth Factor- Like 7 (EGFL7)* which is involved in Notch binding (30). In addition, a *Myomaker Myoblast Fusion Factor (MYMK)* resides in this region and has been associated with muscle hypertrophy, making it a strong candidate for impacting variation in teat morphology (31). Genes were also investigated in various regions on BTA 16, including *Centrosomal Protein 350 (CEP350)* which plays a critical role in microtubule binding and spindle integrity during cell replication (32). This region is also home to *Laminin Subunit Gamma 1 and 2 (LAMC1/2)*, which are thought to regulate cell organization into tissues, potentially contributing to variation in teat width (33). In regards to immune function, nearby genes were investigated including *Major Histocompatibility Complex Class I-Related (MR1)* which is critical to adaptive immune response, *Ribonuclease L (RNASEL)* is involved in interferon regulation, and *DExH-Box Helicase 9 (DHX9)*, which has been found to control TLR-stimulated immune responses (34-36). A significantly associated QTL at BTA 19:29058547-29063744 is located within the *Growth Arrest Specific 7 (GAS7)* gene, which while previously understood to influence neuron differentiation, was recently found to be abundantly expressed in murine alveolar macrophages, though its exact roles in immune responses are still unknown (37). Furthermore, *HIC ZBTB Transcriptional Repressor (HIC1)* and *Tyrosine 3-Monooxygenase Activation Protein (YWHAE)* have both been associated with the regulation of cell proliferation, and in the case of *HIC1* its function as a transcriptional regulator has been specifically tied to immune homeostasis (38, 39). A significantly associated QTL on BTA 25:40126743-40190566 resides within the *Sidekick Cell Adhesion Molecule 1 (SDK1)* gene, an adhesion molecule isoform in the immunoglobulin superfamily primarily known for synapse formation in the retina, though Sidekicks in general are expressed in many different tissue types

(40). A significantly associated SNP at 25:35208040 was interrogated for candidate genes in the surrounding area, which included *Cut Like Homeobox 1 (CUX1)*, known for its role in morphogenesis as well as regulation of antigen presenting cells (41, 42). *Myosin Light Chain 10 (MYL10)*, implicated in immune cell transmigration, and *Tripartite Motif Containing 56 (TRIM56)*, an ubiquitin-ligase with a role in antiviral innate immunity, were also found in this region (43, 44). A case-control GWA study was also performed to identify genes which may potentially drive extremes in morphology, and significantly associated a single SNP at BTA 25:38568564. Candidate genes near this SNP included *Polypeptide N-Acetylgalactosaminyltransferase 2* (*GALNT2*) which is believed to be involved in O-linked glycosylation of the immunoglobulin A1 hinge region and *PiggyBac Transposable Element Derived 5* (*PGBD5*), a transposase suspected to mediate genomic rearrangements (45, 46). The large number of QTL identified spanning the genome suggests rear teat width may be highly polygenic.

We previously associated low rear udder height with increased odds of clinical mastitis (Miles et al., 2019), and a case-control comparison of low udder height versus all other udder height types also identified several significantly associated QTL spanning the genome (Figure 4). A significantly associated SNP at 14:27024015 lay within the *Aspartate Beta-Hydroxylase (ASPH)* gene, known for hydroxylating epidermal growth factors and contributing to dysmorphic features (47). Furthermore, a significantly associated QTL on BTA 18 was near *Troponin T1, Slow Skeletal Type (TNNT1)*, a sarcomere regulatory complex associated with muscle weakness, which could feasibly contribute to weak rear udder attachments, and consequently lower udder height (48). In addition, *NLR Family Pyrin Domain Containing 2 (NLRP2)*, a member of the NLR family known to be regulators of immune responses, was in this same region (49). While we considered low rear udder height as a risk factor for clinical mastitis, the majority of candidate genes identified in this area seem more related to the physical trait itself rather than the indirect trait of “mastitis-risk”, making rear udder height a less compelling trait by which to base breeding and culling decisions for mastitis control.

We hypothesized that the genotype by environment interaction may be greater in multiparous cows who have had greater mastitis exposure and mechanical manipulation of the udder and teats via milking, and thus evaluated a primiparous-only subset of cows (n = 144) for each trait. Only linear GWAs assessing continuous variation in the trait resulted in significant associations; the lack of success with the case-control approach is likely explained by few risk phenotypes (< 20%) observed in primiparous cows. In these linear GWA, QTL were only significantly associated with two traits (front teat placement and udder depth) likely due to low power from a smaller sample size. For udder depth, different inheritance patterns and QTL were identified for each population (Figure 3), which is likely explained by farm culling resulting in different genotypic frequencies among primiparous and multiparous cows. The total cohort GWA associated a single SNP at BTA 5:113268242 located within *Transcription Factor 20 (TCF20)* which is primarily associated with human neurodevelopmental disorders (50). The primiparous only subset significantly associated one QTL on BTA 17. Antisense genes *Nudix Hydrolase 6 (NUDT6)* and *Fibroblast Growth Factor 2 (FGF2)* were in linkage with this QTL, and are hypothesized to regulate cell proliferation (51). Also in linkage with this QTL lay *Sprouty RTK Signaling Antagonist 1 (SPRY1)*, which has been shown to influence mammary epithelial morphogenesis during post-natal development by negatively

regulating epidermal growth factor signaling in the murine mammary gland (52). We hypothesize that this QTL was masked in our analysis of the total cohort due to the differing genotypic frequencies in multiparous versus primiparous populations, suggesting to truly elucidate the genetic mechanism underlying these morphological characteristics primiparous populations with minimal exposure to selective pressures must be evaluated.

Conclusions

This is the first study to use high density SNP chip data with direct phenotyping and no reliance on imputation to investigate the genetic mechanisms regulating bovine udder and teat conformation. Potential biases due to indirect phenotyping and genotype imputation are mitigated by assessing high density genotypes, researcher-controlled phenotypes, and disease risk all within the same population of cows. For many traits, we found significantly associated QTL spanning the genome, suggesting that udder and teat morphologies are complex traits with many causal variants with small effect sizes. A noteworthy aspect of this study was the examination of a primiparous only subset of cows, which demonstrated different inheritance patterns and associated QTL than the total multiparous cohort. In particular, the objective of this study was to identify genetic mechanisms which may underlie our previously identified mastitis risk factors of loose fore udder attachment, flat rear teat ends, wide rear teats, and low udder height in this cohort of cows. In the case of these risk factors, candidate genes surrounding significantly associated QTL were identified with functions relating to both the immune response and cell proliferation and tissue morphology, suggesting these traits are indeed representative of mastitis susceptibility.

Methods

Phenotyping. We conducted a prospective cohort study of 523 Holstein cows on 2 commercial herds in New York State involving direct udder and teat phenotype determination as previously described (15). Udder and teat traits included fore udder attachment, udder cleft, udder depth, rear udder height, rear udder width, as well as front and rear teat placement, end shape, length, and width. Front and rear teat length and width were scored quantitatively, while the remainder were scored on 3 categorical levels according to the U.S. Holstein Association's linear descriptive traits (Holstein Association USA, 2014). Phenotypes were first considered quantitatively to assess continuous variation in the trait, but also on a case-control basis comparing extremes in morphology (53).

Genotyping and Quality Control. A whole blood sample from each cow was taken via the coccygeal vein, collected in 10 mL K_2EDTA anticoagulant vacutainers, and stored at 4ºC or -20ºC until DNA extraction. Genomic DNA was extracted from whole blood according to the Gentra Puregene Blood Kit protocol (Gentra Systems, Inc. Minneapolis, MN, USA) using laboratory-made buffers. In sum, 471 cows were submitted to GeneSeek (Neogen Genomics, Lincoln, NE) for SNP genotyping on the Illumina BovineHD 777K beadchip (Illumina, Inc., San Diego, CA). Quality control filtering was applied to all genotypes via Golden Helix SNP & Variation Suite software v8.3.4 (Golden Helix, Bozeman, MT). Genotypes were

retained if SNP and individual call rate ≥ 0.9, minor allele frequency ≥ 0.05, and allele number ≤ 2. Identity-by-descent (**IBD**) estimates were calculated for all pairs of individuals based on available genotype data and individuals with IBD estimate ≥ 0.9 were removed.

Principal Component Analysis. PCA of phenotype data was performed using RStudio version 3.2.5 (54) to identify components that may describe multiple udder and teat traits. Four PCAs were performed: teat traits only (front and rear teat placement, end shape, length, and width), udder traits only (fore udder attachment, udder cleft, udder depth, rear udder height, and rear udder width), teat and udder traits together, and "risk" traits only (rear teat end shape, rear teat width, fore udder attachment, and rear udder height).

Genome-Wide Association. Efficient Mixed Model Linear analysis (**EMMAX**) models were employed in Golden Helix SVS allowing the inclusion of the IBD matrix to correct for any population structure among this cohort of cows (55). Additive, dominant, and recessive inheritance models were considered along with the variables of farm, parity, batch, rear teat length, udder depth, and rear udder width as potential covariates given their potential correlation and confounding effects on traits of interest. To address the possibility that the genetic variance component of the phenotype is greater for cows who have had less exposure to the environment, a primiparous-only subset of the larger cohort of cows was also evaluated (n = 144). *P*-values were adjusted for multiple testing using a Bonferroni correction and False Discovery Rate (**FDR**).

Model Selection and Candidate Gene Investigation. Quantile–quantile (**QQ**) plots of the $\log_{10}$ (Expected *P*-values) and $\log_{10}$ (Observed *P*-values) and a genomic inflation factor lambda (median observed *P*-value divided by the median expected *P*-value) were used to assess goodness of fit. A stringent multiple testing correction was applied and only regions with SNPs passing Bonferroni correction (adjusted *P*-value < 0.05) were interrogated for candidate genes. Any gene in linkage disequilibrium (**LD**) with an associated marker, or in a 500 kb upstream or downstream range of said marker if LD was not present, was identified using the NCBI RefSeq Database (56). Gene functions were investigated using the GeneCards database (57). All genome coordinates given use the most recent ARS UCD 1.2 bovine genome assembly.

Abbreviations

BTA – *Bos taurus* Autosome

DNA – Deoxyribonucleic Acid

EMMAX – Efficient Mixed Model Linear Analysis

FDR – False Discovery Rate

FUA – Fore Udder Attachment

GWA – Genome Wide Association

IBD – Identity By Descent

K_2EDTA – Dipotassium Ethylenediaminetetraacetic Acid

LD – Linkage Disequilibrium

NCBI – National Center for Biotechnology Information

NLR – Nucleotide-binding domain and Leucine-rich Repeat

PC – Principal Component

PCA – Principal Component Analysis

QQ – Quantile-Quantile

QTL – Quantitative Trait Loci

RTS – Rear Teat End Shape

RTW – Rear Teat Width

SNP – Single Nucleotide Polymorphism

TLR – Toll-like Receptor

UH – Rear Udder Height

Declarations

Ethics approval and consent to participate: Farm owner signed consent was obtained prior to sampling and this study was approved under the Cornell University Institutional Animal Care and Use Committee Protocol #2014-0121 at Cornell University. This was a non-intervention study and all animals remained in the farm-owned herd after study completion.

Consent for publication. Not applicable.

Availability of data and materials. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests. The authors declare that they have no competing interests.

Funding. This work was made possible through funding by the National Institute of Food and Agriculture Animal Health Project #NYC-127898.

Authors' contributions. AM performed on-farm data collection, DNA extraction, sample preparation for genotyping, statistical and genetic analyses, interpretation, and manuscript preparation. CP aided in on-farm data collection, DNA extraction, and interpretation of genetic analyses. HH provided funding and study design. All authors read and approved the final manuscript.

Acknowledgements. We thank commercial dairy farms for their invaluable participation, and many Cornell University graduate and undergraduate students associated with the Huson lab for their assistance with on farm sampling and laboratory work.

References

1. iGEM C. Market Analysis: Bovine Mastitis. 2016.
2. O'Bleness GV, van Vleck LD, Henderson CR. Heritabilities of Some Type Appraisal Traits and Their Genetic and Phenotypic Correlations with Production. Journal of Dairy Science. 1960;43(10):1490-8.
3. Wiggans GR, VanRaden PM, Cooper TA. The genomic evaluation system in the United States: Past, present, future. Journal of Dairy Science. 2011;94(6):3202-11.
4. Interbull. Description of National Genetic Evaluation Systems: Conformation 2013 [Available from: https://queries.uscdcb.com/reference/Form_GENO_Conformation_1302.pdf.
5. Parker Gaddis KL, Cole JB, Clay JS, Maltecca C. Genomic selection for producer-recorded health event data in US dairy cattle. Journal of Dairy Science. 2014;97(5):3190-9.
6. Seykora AJ, McDaniel BT. Udder and teat morphology related to mastitis resistance: a review. J Dairy Sci. 1985;68(8):2087-93.
7. Rupp R, Boichard D. Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. J Dairy Sci. 1999;82(10):2198-204.
8. DeGroot BJ, Keown JF, Van Vleck LD, Marotz EL. Genetic Parameters and Responses of Linear Type, Yield Traits, and Somatic Cell Scores to Divergent Selection for Predicted Transmitting Ability for Type in Holsteins1. Journal of Dairy Science. 2002;85(6):1578-85.
9. Seykora AJ, McDaniel BT. Genetics Statistics and Relationships of Teat and Udder Traits, Somatic Cell Counts, and Milk Production1. Journal of Dairy Science. 1986;69(9):2395-407.
10. Chrystal MA, Seykora AJ, Hansen LB, Freeman AE, Kelley DH, Healey MH. Heritability of Teat-End Shape and the Relationship of Teat-End Shape with Somatic Cell Score for an Experimental Herd of Cows1,2. Journal of Dairy Science. 2001;84(11):2549-54.
11. Nash DL, Rogers GW, Cooper JB, Hargrove GL, Keown JF, Hansen LB. Heritability of Clinical Mastitis Incidence and Relationships with Sire Transmitting Abilities for Somatic Cell Score, Udder Type Traits, Productive Life, and Protein Yield. Journal of Dairy Science. 2000;83(10):2350-60.
12. Pausch H, Emmerling R, Schwarzenbacher H, Fries R. A multi-trait meta-analysis with imputed sequence variants reveals twelve QTL for mammary gland morphology in Fleckvieh cattle. Genetics Selection Evolution. 2016;48(1):14.

13. Tolleson MW, Gill CA, Herring AD, Riggs PK, Sawyer JE, Sanders JO, et al. Association of udder traits with single nucleotide polymorphisms in crossbred Bos indicus–Bos taurus cows1,2. Journal of Animal Science. 2017;95(6):2399-407.
14. Marete A, Lund MS, Boichard D, Ramayo-Caldas Y. A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. PLOS ONE. 2018;13(7):e0199931.
15. Miles AM, McArt JAA, Leal Yepes FA, Stambuk CR, Virkler PD, Huson HJ. Udder and teat conformational risk factors for elevated somatic cell count and clinical mastitis in New York Holsteins. Preventive Veterinary Medicine. 2019;163:7-13.
16. Bonavita R, Walas D, Brown AK, Luini A, Stephens DJ, Colanzi A. Cep126 is required for pericentriolar satellite localisation to the centrosome and for primary cilium formation. Biology of the cell. 2014;106(8):254-67.
17. Drake AC, Khoury M, Leskov I, Iliopoulou BP, Fragoso M, Lodish H, et al. Human CD34+ CD133+ hematopoietic stem cells cultured with growth factors including Angptl5 efficiently engraft adult NOD-SCID Il2rgamma-/- (NSG) mice. PLoS One. 2011;6(4):e18382.
18. Thrasher AJ, Burns SO. WASP: a key immunological multitasker. Nature reviews Immunology. 2010;10(3):182-92.
19. Zhang C, Zhang W, Lu Y, Yan X, Yan X, Zhu X, et al. NudC regulates actin dynamics and ciliogenesis by stabilizing cofilin 1. Cell research. 2016;26(2):239-53.
20. Ghahary A, Karimi-Busheri F, Marcoux Y, Li Y, Tredget EE, Taghi Kilani R, et al. Keratinocyte-releasable stratifin functions as a potent collagenase-stimulating factor in fibroblasts. J Invest Dermatol. 2004;122(5):1188-97.
21. Lee S, Kong Y, Weatherbee SD. Forward genetics identifies Kdf1/1810019J16Rik as an essential regulator of the proliferation-differentiation decision in epidermal progenitor cells. Developmental biology. 2013;383(2):201-13.
22. Plovsing RR, Berg RM, Munthe-Fog L, Konge L, Iversen M, Moller K, et al. Alveolar recruitment of ficolin-3 in response to acute pulmonary inflammation in humans. Immunobiology. 2016;221(5):690-7.
23. Moller DE, Xia CH, Tang W, Zhu AX, Jakubowski M. Human rsk isoforms: cloning and characterization of tissue-specific expression. Am J Physiol. 1994;266(2 Pt 1):C351-9.
24. Tian H, Miao J, Zhang F, Xiong F, Zhu F, Li J, et al. Non-histone nuclear protein HMGN2 differently regulates the urothelium barrier function by altering expression of antimicrobial peptides and tight junction protein genes in UPEC J96-infected bladder epithelial cell monolayer. Acta biochimica Polonica. 2018;65(1):93-100.
25. Vieira Braga FA, Hertoghs KM, Kragten NA, Doody GM, Barnes NA, Remmerswaal EB, et al. Blimp-1 homolog Hobit identifies effector-type lymphocytes in humans. European journal of immunology. 2015;45(10):2945-58.

26. Zhang C, Zhen YZ, Lin YJ, Liu J, Wei J, Xu R, et al. KNDC1 knockdown protects human umbilical vein endothelial cells from senescence. Molecular medicine reports. 2014;10(1):82-8.
27. Knapp B, Wolfrum U. Adhesion GPCR-Related Protein Networks. Handbook of experimental pharmacology. 2016;234:147-78.
28. Lamas B, Michel ML, Waldschmitt N, Pham HP, Zacharioudaki V, Dupraz L, et al. Card9 mediates susceptibility to intestinal pathogens through microbiota modulation and control of bacterial virulence. Gut. 2018;67(10):1836-44.
29. Bertin J, Guo Y, Wang L, Srinivasula SM, Jacobson MD, Poyet JL, et al. CARD9 is a novel caspase recruitment domain-containing protein that interacts with BCL10/CLAP and activates NF-kappa B. J Biol Chem. 2000;275(52):41082-6.
30. Siebel C, Lendahl U. Notch Signaling in Development, Tissue Homeostasis, and Disease. Physiological reviews. 2017;97(4):1235-94.
31. Si Y, Wen H, Du S. Genetic Mutations in jamb, jamc, and myomaker Revealed Different Roles on Myoblast Fusion and Muscle Growth. Marine biotechnology (New York, NY). 2019;21(1):111-23.
32. Yan X, Habedanck R, Nigg EA. A complex of two centrosomal proteins, CAP350 and FOP, cooperates with EB1 in microtubule anchoring. Molecular biology of the cell. 2006;17(2):634-44.
33. Nakad B, Fares F, Azzam N, Feiner B, Zilberlicht A, Abramov Y. Estrogen receptor and laminin genetic polymorphism among women with pelvic organ prolapse. Taiwanese journal of obstetrics & gynecology. 2017;56(6):750-4.
34. Yamaguchi H, Kurosawa Y, Hashimoto K. Expanded genomic organization of conserved mammalian MHC class I-related genes, human MR1 and its murine ortholog. Biochem Biophys Res Commun. 1998;250(3):558-64.
35. Squire J, Zhou A, Hassel BA, Nie H, Silverman RH. Localization of the interferon-induced, 2-5A-dependent RNase gene (RNS4) to human chromosome 1q25. Genomics. 1994;19(1):174-5.
36. Dempsey A, Keating SE, Carty M, Bowie AG. Poxviral protein E3-altered cytokine production reveals that DExD/H-box helicase 9 controls Toll-like receptor-stimulated immune responses. J Biol Chem. 2018;293(39):14989-5001.
37. Xu Q, Liu X, Wang X, Hua Y, Wang X, Chen J, et al. Growth arrest-specific protein 7 regulates the murine M1 alveolar macrophage polarization. Immunologic research. 2017;65(5):1065-73.
38. Leal MF, Ribeiro HF, Rey JA, Pinto GR, Smith MC, Moreira-Nunes CA, et al. YWHAE silencing induces cell proliferation, invasion and migration through the up-regulation of CDC25B and MYC in gastric cancer cells: new insights about YWHAE role in the tumor development and metastasis process. Oncotarget. 2016;7(51):85393-410.
39. Burrows K, Antignano F, Bramhall M, Chenery A, Scheer S, Korinek V, et al. The transcriptional repressor HIC1 regulates intestinal immune homeostasis. Mucosal immunology. 2017;10(6):1518-28.
40. Yamagata M, Sanes JR. Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. Nature. 2008;451(7177):465-9.

41. Kuhnemuth B, Muhlberg L, Schipper M, Griesmann H, Neesse A, Milosevic N, et al. CUX1 modulates polarization of tumor-associated macrophages by antagonizing NF-kappaB signaling. Oncogene. 2015;34(2):177-87.
42. Xu H, He JH, Xu SJ, Xie SJ, Ma LM, Zhang Y, et al. A group of tissue-specific microRNAs contribute to the silencing of CUX1 in different cell lineages during development. Journal of cellular biochemistry. 2018;119(7):6238-48.
43. Oltz EM, Yancopoulos GD, Morrow MA, Rolink A, Lee G, Wong F, et al. A novel regulatory myosin light chain gene distinguishes pre-B cell subsets and is IL-7 inducible. Embo j. 1992;11(7):2759-67.
44. Chen Y, Zhao J, Li D, Hao J, He P, Wang H, et al. TRIM56 Suppresses Multiple Myeloma Progression by Activating TLR3/TRIF Signaling. Yonsei medical journal. 2018;59(1):43-50.
45. Iwasaki H, Zhang Y, Tachibana K, Gotoh M, Kikuchi N, Kwon YD, et al. Initiation of O-glycan synthesis in IgA1 hinge region is determined by a single enzyme, UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2. J Biol Chem. 2003;278(8):5613-
46. Henssen AG, Henaff E, Jiang E, Eisenberg AR, Carson JR, Villasante CM, et al. Genomic DNA transposition induced by human PGBD5. eLife. 2015;4.
47. Patel N, Khan AO, Mansour A, Mohamed JY, Al-Assiri A, Haddad R, et al. Mutations in ASPH cause facial dysmorphism, lens dislocation, anterior-segment abnormalities, and spontaneous filtering blebs, or Traboulsi syndrome. Am J Hum Genet. 2014;94(5):755-9.
48. Fox MD, Carson VJ, Feng HZ, Lawlor MW, Gray JT, Brigatti KW, et al. TNNT1 nemaline myopathy: natural history and therapeutic frontier. Hum Mol Genet. 2018;27(18):3272-82.
49. Jones JD, Vance RE, Dangl JL. Intracellular innate immune surveillance devices in plants and animals. Science. 2016;354(6316).
50. Schafgen J, Cremer K, Becker J, Wieland T, Zink AM, Kim S, et al. De novo nonsense and frameshift variants of TCF20 in individuals with intellectual disability and postnatal overgrowth. European journal of human genetics : EJHG. 2016;24(12):1739-45.
51. Asa SL, Ramyar L, Murphy PR, Li AW, Ezzat S. The endogenous fibroblast growth factor-2 antisense gene product regulates pituitary cell growth and hormone production. Molecular endocrinology (Baltimore, Md). 2001;15(4):589-99.
52. Koledova Z, Zhang X, Streuli C, Clarke RB, Klein OD, Werb Z, et al. SPRY1 regulates mammary epithelial morphogenesis by modulating EGFR-dependent stromal paracrine signaling and ECM remodeling. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(39):E5731-40.
53. Nazari-Ghadikolaei A, Mehrabani-Yeganeh H, Miarei-Aashtiani SR, Staiger EA, Rashidi A, Huson HJ. Genome-Wide Association Studies Identify Candidate Genes for Coat Color and Mohair Traits in the Iranian Markhoz Goat. Frontiers in genetics. 2018;9:105-.
54. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

55. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nature Genetics. 2010;42:348.
56. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733-45.
57. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Current Protocols in Bioinformatics. 2016;54(1):1.30.1-1..3.

Table

Table 1. Models selected. Summary of final models including sample size (N), model type, inheritance patterns, quality control measures, trait heritabilities, and total number of significantly associated SNPs.

Trait	N	Model Type	Inheritance	Pseudo-Lambda[1]	Pseudo-heritability	FDR[2]	Bonferroni[2]
Front Teat Length	458	Linear	Recessive	1.00	0.33	5	2
Front Teat Width	458	Case-control[3]	Recessive	1.01	0.06	4	4
Fore Udder Attachment	288	Case-control[4]	Additive	1.00	0.45	4	3
Risk PC1	458	Linear	Dominant	1.01	0.07	2	1
Rear Teat Length	458	Case-control[5]	Additive	1.00	0.42	1	1
Rear Teat End Shape	227	Case-control[6]	Recessive	1.02	0.50	1	1
Rear Teat Width	458	Case-control[3]	Recessive	1.02	0.03	3	1
Rear Teat Width	458	Linear	Recessive	0.98	0.18	50	23
Udder Depth	265	Case-control[7]	Dominant	0.99	0.99	1	1
Udder Height	458	Case-control[8]	Recessive	1.00	0.02	129	32
Udder Width	458	Case-control[9]	Dominant	1.00	0.30	1	1
Front Teat Placement[10]	144	Linear	Recessive	1.02	0.38	10	7
Udder Depth[10]	144	Linear	Recessive	1.05	0.62	10	0

[1]genomic inflation factor for model quality control, Quantile-Quantile plots in Supplementary Figure 1

[2]number of SNP associations passing either Bonferroni or False Discovery Rate (FDR) multiple testing corrections at $P < 0.05$

[3]narrow versus wide teats, split by median measurement

[4]loose versus tight fore udders, excluding all cows with intermediate fore udder attachment

[5]short versus long rear teats, split by median measurement

[6]flat versus pointed rear teat ends, excluding all cows with round rear teat end shape

[7]deep versus high udder depth, excluding all cows with intermediate udder depth

[8]low versus all other udder heights (intermediate and high rear udders combined)

[9]wide versus all other udder widths (intermediate and narrow udders combined)

[10]primiparous cow subset

Additional Files

Supplementary Figure 1. Quantile-Quantile plots showing the –log10(Expected *P* Values) plotted against the –log10(Observed *P* Values) for each final model A) Front Teat Length, B) Front Teat Width, C) Fore Udder Attachment, D) Risk PC1, E) Rear Teat Length, F) Rear Teat End Shape, G) Rear Teat Width (case-control), H) Rear Teat Width (linear), I) Udder Depth, J) Udder Height, K) Udder Width, L) Front Teat Placement (primiparous cows only), M) Udder Depth (primiparous cows only. The black line represents X = Y.

File Name: MilesA_BMC-Genomics_U&T-GWAS_Figures.ppt

Title: All Figures

Description: Per the submission guidelines, all figures to be included in body of paper

(Figures1-6) are on separate slides in this file. Supplementary Figure 1 is also included in

this file.

File Name: MilesA_BMC-Genomics_U&T-GWAS_Supplementary-Table-1.doc

Title: Supplementary Table 1

Description: A complete list of significantly associated QTL for all GWA models, their

position, and candidate genes.

Figures

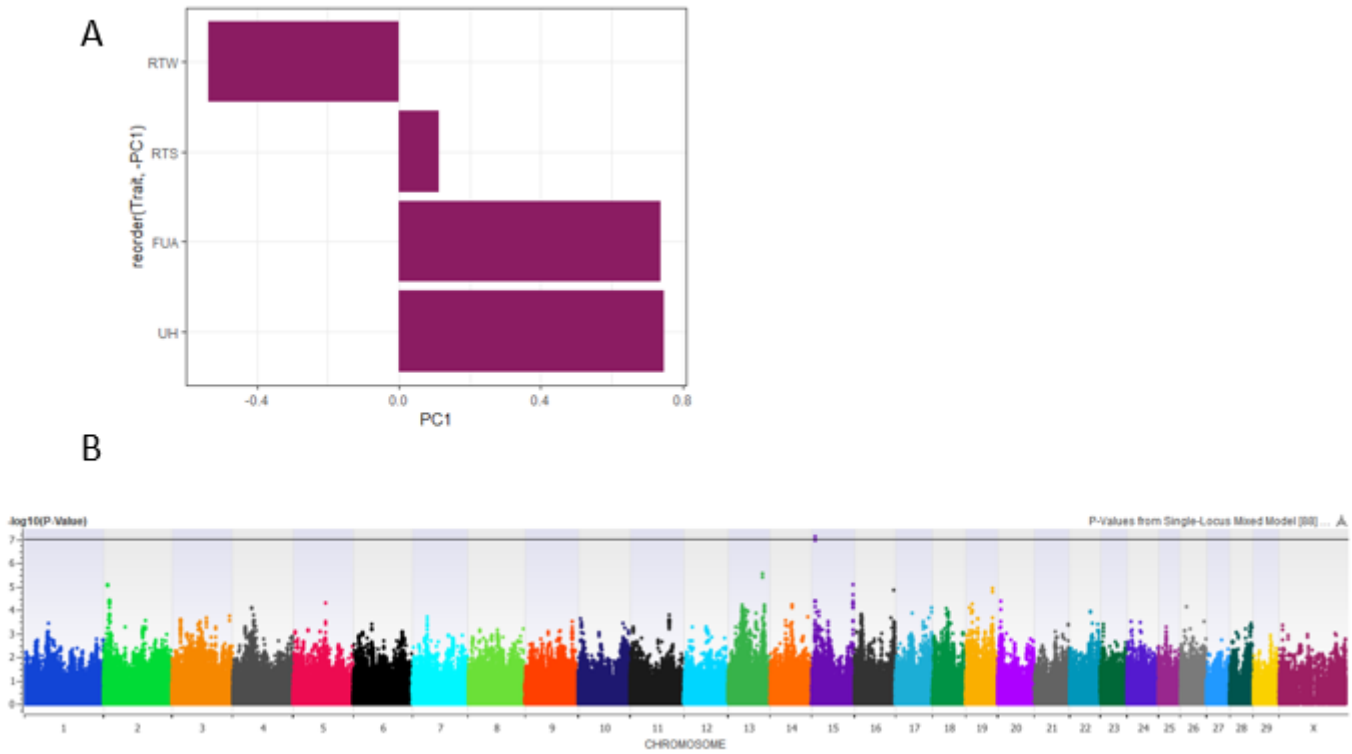

Figure 1

Creation of new “mastitis risk” trait for GWA using PCA. A) Loadings values for Risk PC1 by RTW (rear teat width), RTS (rear teat end shape), FUA (fore udder attachment), and UH (udder height). B) Manhattan plot showing –log10 P-values by chromosome in a linear GWA for Risk PC1 with the black line representing the Bonferroni multiple correction testing threshold.

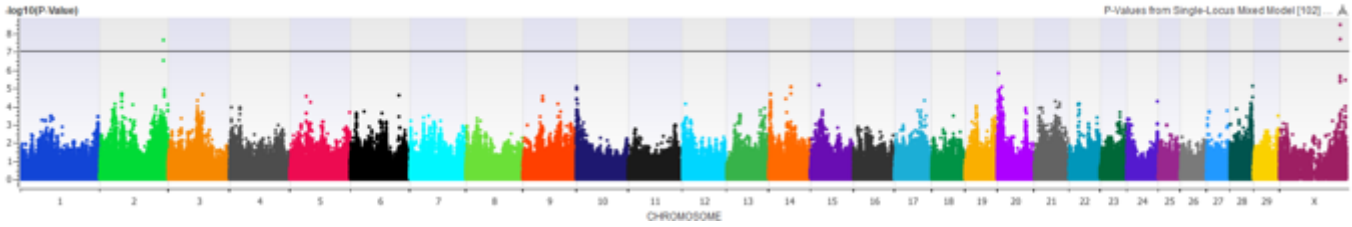

Figure 2

Manhattan Plot showing –log10 P-values by chromosome in a case-control GWA for fore udder attachment, comparing loose fore udders to tight fore udders (N = 288). The black line represents the Bonferroni multiple testing correction threshold.

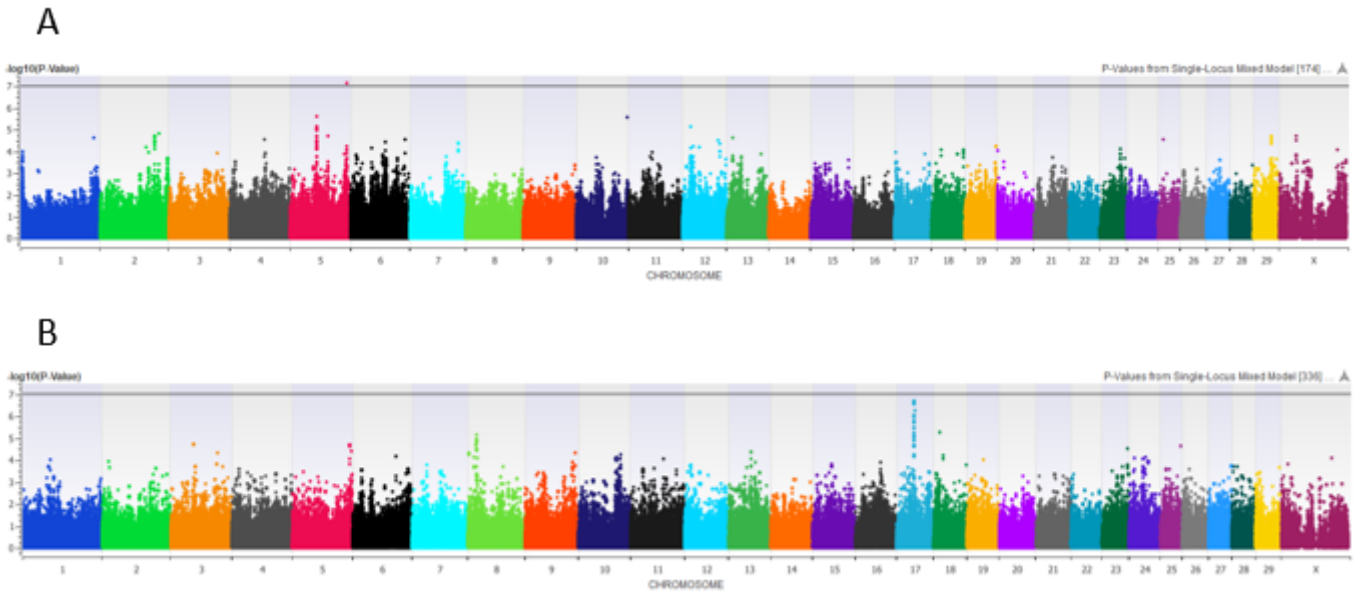

Figure 3

Manhattan Plot showing −log10 P-values by chromosome in a A) case-control GWA with dominant inheritance comparing deep versus high udders for all cows (N = 458), and B) a linear GWA with recessive inheritance using all udder depth scores (deep, intermediate, and high) for a subset of primiparous cows only (N = 144). The black line represents the Bonferroni multiple testing correction threshold.

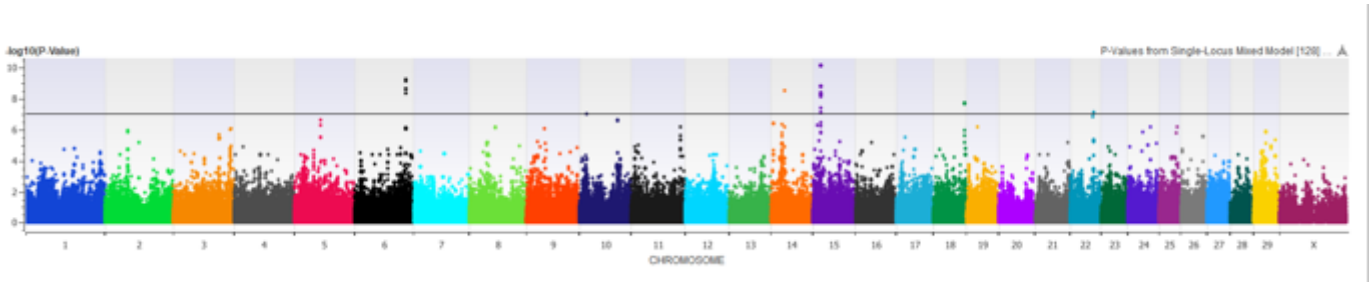

Figure 4

Manhattan Plot showing −log10 P-values by chromosome in a case-control GWA for udder height, comparing low rear udders to all other udder classifications (intermediate and high udder height). The black line represents the Bonferroni multiple testing correction threshold.

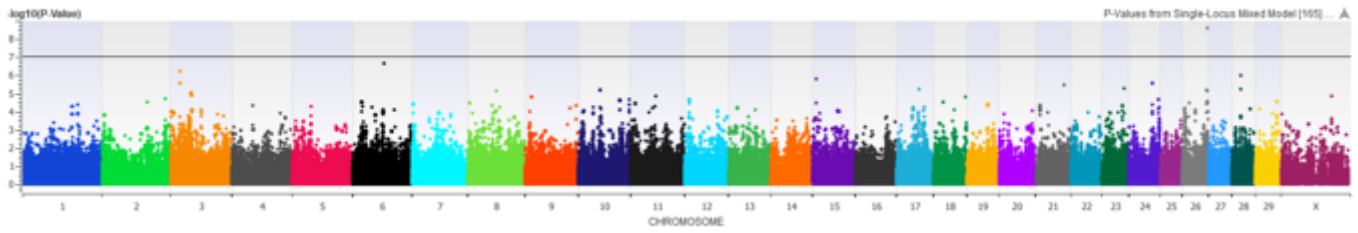

Figure 5

Manhattan Plot showing −log10 P-values by chromosome in a case-control GWA for rear teat end shape, comparing flat and pointed teat ends (N = 227). The black line represents the Bonferroni multiple testing correction threshold.

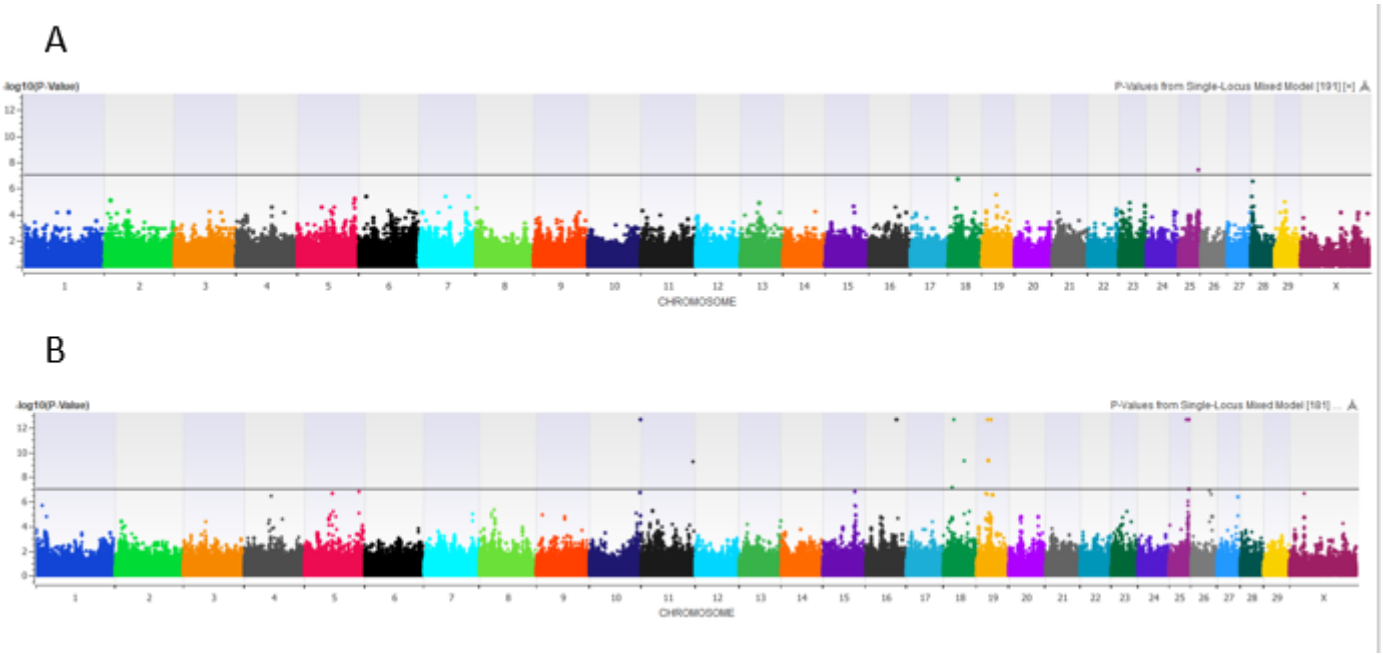

Figure 6

Manhattan Plot showing −log10 P-values by chromosome in a A) case-control GWA for rear teat width where cows are divided around the median value, and B) linear GWA for rear teat width using the quantitative measures. The black line represents the Bonferroni multiple testing correction threshold.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.jpg
- SupplementaryTable1.docx