Identifying Loci Associated with Susceptibility to Foot Lesions in Holstein Cattle

By

ELLEN LAI
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Anita M. Oberbauer, Chair

_____

Hao Cheng

_____

Terry W. Lehenbauer

Committee in Charge

2021

Acknowledgements

Although my name is listed as the author of this dissertation, the research summarized in this dissertation is truly the product of the combined efforts of a vast network of people who believed in me and the work that we have done. I would first like to thank my advisor Dr. Anita Oberbauer, whose steady, endlessly patient guidance and exceptionally kind and empathetic mentorship has allowed be to grow into the scientist I am today. Dr. O has been my scientific fairy godmother who always found a way to make sense of the unsensible and keep me motivated. I am grateful for Janelle Belanger-Sandoval, whose optimism never wavered in the face of my mistakes or exceedingly strange results and who has supplied me with truly exceptional olive oil at an even more exceptional price. I thank Dr. Liza Gershony, who always made time to help me in and out of lab, became one of my dearest friends, and remained one of my dearest friends even after I led her wildly astray on a hike that resulted in our hiking back to the trailhead in the dark with no cell service, no flashlight, and an aggressive bull prowling along the trail. I am indebted to Dr. Tom Famula, a one-of-a-kind stats magician who has taught me a couple tricks and given me the confidence to delve into more complex stats magic lore. I sincerely thank Dr. Terry Lehenbauer for providing the most up-to-date veterinary perspective and connecting me with Dr. Richard Van Vleck Pereira, who in turn helped me recruit another dairy to participate in our study. I wholeheartedly thank Dr. Hao Cheng for patiently helping me troubleshoot data analysis and coming up with creative ways to work around the imperfections of my dataset. I also thank Drs. Lehenbauer and Cheng for serving on my dissertation committee and providing thorough feedback on preliminary results and earlier drafts of this dissertation. I am also grateful for Dr. Alexa Danner, who began the sample collection and analysis that laid the groundwork for this project.

This work would not have been possible without the dairy producers and their personnel

who participated in our study and volunteered their time to collect blood samples, hoof trimming records, and health records for us. To these dairy producers and their personnel, thank you. I would also like to thank the two hoof trimmers who answered all my questions, connected me with additional dairies to recruit, and whose consistent, detailed phenotyping of foot disorders made this project possible.

I would like to thank my family and friends, who have kept me afloat during the more frustrating times and celebrated my accomplishments with me over these past five years. To my parents, Ming-Lai Lai and Susan Tzou, who passed down their love of academia to me, and to my sisters, Jean, Michelle, Kristine, and Chelsea, who are always willing to help me whenever I need it, thank you. To my dearest friend, Jessica Lin, who always knows how best to support me and whose sass and dog pictures always keep me entertained, thank you. Lastly, I would like to thank my pooch and most steadfast companion, Harley, whose begrudging affection and opinions have enriched my life for 13 years, who kept my spirits up during the COVID pandemic, and who watched me collect three (almost four) diplomas like he once collected leaves on his tail. To all of you and those I have inevitably forgotten to list here, thank you, thank you, thank you!

Table of contents

Abstract

Lameness imposes concerns for dairy cattle welfare and farmer profitability. After mastitis, lameness is the second most prevalent disease in dairy cattle and is commonly caused by digital dermatitis (DD, also known as foot warts), sole ulcers (SU), and white line disease (WLD). Digital dermatitis is an infectious foot lesion, whereas SU and WLD are noninfectious lesions that arise due to compromised horn production. Genomic selection against these foot lesions and its potential impact on other health traits (mastitis, hypocalcemia, displaced abomasum, ketosis, and metritis) requires the identification of loci associated with these foot lesions and assessment of the genetic correlation of foot lesions with other health traits. To detect susceptibility loci, a genome-wide association study (GWAS) was performed using genotypes from the high density SNP array (777K SNPs) and case/control phenotypes for DD (controls n =129, DD n = 85), SU (controls n = 102, SU n = 152), WLD (controls = 102, WLD n = 117), SU and/or WLD (SU and WLD, controls n = 102, n = 198), and any type of noninfectious foot lesion (controls n = 102, cases = 217). GWAS was performed using linear mixed model (LMM) and random forest (RF) approaches, and effect sizes of top SNPs were estimated using Bayesian regression. For the LMM GWAS, the number of effective SNPs (NES) was calculated as the number of SNPs that were not in linkage disequilibrium and used as the denominator to define Bonferroni-corrected p-value thresholds of genome-wide statistical significance ($p \leq 0.05/NES$) and suggestive significance ($p \leq 0.2/NES$). Genetic correlation among foot lesions and health traits was estimated using bivariate genome-based restricted maximum likelihood (GREML) analysis, and a multi-trait GWAS was conducted to identify genomic regions contributing to genetic correlation. Top SNPs identified in the GWAS were in or near genes that were functionally relevant to foot lesion etiology. For DD, both the LMM and RF analyses identified regions of association on *Bos taurus* autosome (BTA) 1 and 2, with one of the regions on BTA2

containing candidate genes related to immune function. The LMM GWAS revealed an associated region on BTA 8 for SU and BTA13 for WLD, SU and WLD, and noninfectious foot lesions. These associated regions contained genes related to wound healing, skin lesions, bone growth and mineralization, adipose tissue, and keratinization. Furthermore, the region on BTA8 included a SNP previously associated with SU susceptibility. The RF GWAS for SU, WLD, SU and WLD, and noninfectious lesions were overfitted, suggesting that the SNP effects were very small and prevented detection of susceptibility loci using this approach. Estimated effect sizes of top SNPs were small, and though significant genetic correlation was detected among lameness and health traits, the sample size prevented detection of loci contributing to multiple traits. The small effect sizes and the limited ability to detect pleiotropic loci reinforces that the environment plays a nontrivial role in disease susceptibility, and the remaining genetic component is likely governed by many loci. Larger sample sizes are necessary to identify small effect loci and their association with individual or multiple lameness and health traits amidst a strong environmental effect.

Simple summary

Lameness is the second most prevalent disease in dairy cattle after mastitis and the third most common reason for culling after mastitis and infertility. Lameness indicates cow discomfort and is often caused by painful foot disorders, most commonly digital dermatitis (DD, also known as foot warts), sole ulcers (SU), and white line disease (WLD). These foot disorders are an animal welfare issue, incur substantial financial losses for the producer, and inflate the environmental footprint per unit of milk due to losses in efficiency of resource use. Risk of developing DD, SU, or WLD has a small but significant genetic component, meaning that genetic selection against these foot disorders is possible once the location of the genetic markers (i.e., what position on which chromosome), or the genes, for these claw disorders is known. However, the genetic cause for DD, SU, and WLD are not definitively known. The objective of our study was to identify genetic markers associated with risk of developing DD, SU, WLD, and any noninfectious foot lesion. For each foot disorder, we compared genetic markers between cows that had at least one episode of the foot disorder, and sound cows to find which genetic markers were overrepresented in the lame cows. The markers that were overrepresented in the lame cows were considered associated with risk of developing that foot disorder. We found genetic markers for risk on chromosome 2, 7, and 20 for DD; chromosome 8 for SU; and chromosome13 for WLD. Notably, the genetic markers that we found for SU were in the same chromosomal region as those found in a previous study of SU. Exploring the regions in the genome adjacent to these identified risk markers revealed candidate genes that could plausibly play a role in the development of DD, SU, and WLD. Although these genetic markers, and genes, for risk are promising, they collectively have a very small influence on risk compared to non-genetic means of control (e.g., medicated foot baths to prevent DD, minimizing standing time to prevent SU and WLD). Accordingly, the most effective method of reducing the

prevalence of these claw disorders is through a combination of non-genetic methods and genetically selecting for cows at lower risk of developing foot lesions. Importantly, we found significant positive genetic correlation between SU and DD and between SU and WLD meaning that genetic selection against one of these three foot disorders will also select against the other two foot disorders Though no national genomic evaluation for lameness exists to date, USDA recently announced its intention to develop a lameness index, based upon markers such as those identified in this work, to reduce lameness because the current feet and leg conformation scoring does not reliably predict lameness

Chapter I. Introduction

1 Overview of dairy production

Dairy production relies on the meticulous management of the lactation cycle. Lactation begins after a cow has calved and consists of three phases: early, mid, and late lactation, with each phase lasting ~120 days. Lactation is followed by a ~65 day dry period such that the entire lactation cycle, comprised of the three lactation phases and the dry period, amounts to approximately one year. Three months into lactation, that is, during early lactation, the cow is bred so that she can conceive in a reasonable time frame and initiate the next lactation cycle. Based on this schedule, cows calve and undergo one lactation cycle per year. The more lactation cycles a cow undergoes, the more milk she produces in her lifetime, which offsets the financial and environmental resources used to raise her before she began lactating. Consequently, preventing premature culling elongates the productive lifetime and could translate to a lower financial and environmental impact per unit of milk, assuming heifer replacement costs are high and milk production is low. On average, cows undergo 2.69 lactations before they are culled from the herd (Van Raden et al., 2021), most often due to reproductive issues, mastitis, or lameness (USDA, 2018); less frequently a cow may be culled and replaced with heifers to introduce superior genetics. According to the USDA, 51.9% of cows are removed in the second to fourth lactation, and 25.9% of culled cows were in their fifth or higher lactations (USDA, 2018). One of the most common reasons for premature culling is lameness, with 7.2% of culled cows removed due to lameness (USDA, 2018).

2 What is lameness?

Lameness describes abnormal gait or posture and is an indicator of cow discomfort due to foot or leg disorders. Cows are typically screened for lameness by dairy personnel during daily

health observations and as milking cows are walking to and from the parlor. Personnel look for lameness and other indicators of foot discomfort, including an arched back, uneven gait, standing with hind legs in an elevated bed to relieve pressure on hind legs, frequently shifting their weight from foot to foot, a shift in the order of cows as a pen of cows walks through a gate, changes in the time spent laying down, reduced feeding time, or reduced standing time; all of which impact cow welfare (Hassall et al., 1993; Read and Walker, 1998; Van Nuffel et al., 2015; Sadiq et al., 2017). Upon diagnosis, lame cows are treated by the owner, hoof trimmer, or veterinarian, incurring expenses associated with lameness.

3 Lameness is a sustainability issue

To date, the USDA-National Animal Health Monitoring System (NAHMS) has conducted four national studies to estimate the annual prevalence of lameness in 1996, 2002, 2007, and 2014. These reports demonstrate that lameness has been a longstanding concern in the dairy industry, as it is the second most prevalent disease after mastitis (USDA, 2018). Among the four years that the reports were generated, annual lameness prevalence (i.e., percentage of cows/heifers that had at least one episode of lameness during the year) has ranged from 3.2 to 11.4% for bred heifers and 16.8% to 23.9% for cows. In 2014 (the most recent report to date), 97.7% of US dairies had at least one bred heifer or cow that was lame, with more dairies having lame cows (89.7%) than lame bred heifers (55.2%) (USDA, 2018).

Lameness is not only a prevalent welfare issue, but also presents a financial burden for dairy producers due to direct costs from prevention and treatment and indirect costs from reduced milk production, decreased fertility, and increased labor (Dolecheck and Bewley, 2018). Costs per case of each foot disorder range from $64 to $153 for digital dermatitis (DD), $181 to $258 for sole ulcers (SU), and $107 to 252 for white line disease (WLD) (adjusted to 2020 US

dollars) (Cha et al., 2010; Dolecheck et al., 2019). Reduction in milk production due to lameness decreases the efficiency of resource use, as alluded to above, and translates to a higher environmental impact per unit of milk. Losses in production efficiency due to extended calving interval, higher cull rate, and decreased milk production inflate the greenhouse gas emissions for each case of DD, SU, and WLD by 39 (4.3%), 33 (3.6%), and 39 (4.3%) kg CO_2 equivalents, respectively, per ton of fat-and-protein-corrected milk (Mostert et al., 2018).

Because lameness affects animal welfare, producer profit, and the environment, which collectively comprise the three pillars of sustainability, reducing lameness would improve the sustainability of dairy production.

4 Etiology of digital dermatitis, sole ulcers, and white line disease

Though lameness is sometimes described as a disease itself, it is rather a symptom of a wide variety of injuries or disorders affecting the bones, hoof, or skin and soft tissues of the foot as well as other anatomical components of the front and hind limbs. Among the various causes of lameness, DD, SU, and WLD are the most pervasive (Green et al., 2002; Shearer and van Amstel, 2017; USDA, 2018), and are thus the focus of this dissertation. Digital dermatitis has historically comprised a large share of lameness cases, with the most recent NAHMS report documenting that 70.9% and 36.0% of lameness cases were due to DD in bred heifers and cows, respectively, in 2014 (USDA, 2018). Sole ulcers and WLD were not reported on a nationwide scale, but are consistently identified as a common cause of lameness (Shearer and van Amstel, 2017). All three of these lesions are most commonly observed on one or both hind feet (Murray et al., 2002; Greenough, 2007) and can be extremely painful (Bruijnis et al., 2012).

During the early lactation phase, the cow is in negative energy balance as her body transitions from investing energy into the calf during pregnancy to putting energy into milk

production. Due to this energy deficit and related factors, the early lactation phase is the period of highest disease incidence (USDA, 2018). The cow shunts energy into milk production that would otherwise be used to ward off disease, including the foot lesions that lead to lameness (Hassall et al., 1993; Bicalho et al., 2009; Bicalho and Oikonomou, 2013). Some have observed higher incidence of claw or foot disorders during early lactation and attribute this to the severe negative energy balance during this phase (Collard et al., 2000; Gernand et al., 2013), whereas others have not found significant differences in foot lesion incidence between early and late lactation (van der Spek et al., 2015a).

Digital dermatitis is a lesion typically found on the skin at the heel resulting from a polymicrobial infection predominated by multiple phylotypes of bacteria from the *Treponema* genus (Brandt et al., 2011). Though the etiology of DD is unclear, it is postulated that treponemes invade the skin through hair follicles (Evans et al., 2009) and elicit a strong initial innate immune response (Trott et al., 2003; Watts et al., 2018). The treponemes attenuate this initial immune response (Refaai et al., 2013), leading to prolonged inflammation, a delayed adaptive immune response, and impaired wound healing (Zuerner et al., 2007; Refaai et al., 2013). Consequently, fibroblasts proliferate excessively and develop into a raised erosive lesion with hypertrophied hairs (Read and Walker, 1998). The inability of the immune response to efficiently clear the infection is postulated to contribute not only to the long duration of DD but also to the recurrence of DD (Trott et al., 2003). Some have suggested that heifers have a less developed immune response than cows, which could make heifers more susceptible to DD and explains the higher prevalence of DD in heifers; however, the high rate of new lesion episodes or recurrence in cows (48%) suggests that cows are no more immune to DD than heifers (Read and Walker, 1998).

In contrast to DD, which is an infectious lesion, sole ulcers and white line disease are two types of noninfectious lesions that arise secondarily due to improper weight bearing or metabolic, enzymatic, and hormonal changes that in turn compromise horn production (Shearer and van Amstel, 2017). Accordingly, SU and WLD are considered two types of claw horn disruption lesions (CHDL). Normally, a suspensory system of tendons connected to the third phalanx and the digital cushion beneath the third phalanx stabilize this bone within the claw. Due to hormonal and/or nutritional changes, increased laxity of the suspensory system and thinning of the digital cushion allow the distal phalanx to move excessively in the claw and sink, crushing the underlying corium (Lischer et al., 2002; Bicalho et al., 2009; Newsome et al., 2017a, 2017b; Shearer and van Amstel, 2017; Stambuk et al., 2019). Increased pressure compromises horn production from the corium, leading to defective (WLD) or complete cessation (SU) of horn production at the pressure site (Shearer and van Amstel, 2017). For WLD, defective horn production along the white line makes this region more susceptible to debris and bacterial infiltration that, upon reaching the corium, lead to abscess development (i.e. white line abscess) (Shearer and van Amstel, 2017). Sole ulcers are thought to result from the distal phalanx rotating within the claw such that the heel of the bone sinks and crushes the underlying corium leading to sole hemorrhage, and eventually a hole in the solar horn through which the corium protrudes and sometimes becomes infected (Greenough, 2007; Shearer et al., 2015).

Though it is generally accepted that hormonal and nutritional changes can lead to SU and WLD, the exact mechanism of how these changes predispose a cow to SU and WLD is still debated. One proposed mechanism is that the severe negative energy balance after calving causes loss of body condition as the cow mobilizes fat from body stores, including the digital cushion, which in turn increases risk of lameness (Hassall et al., 1993; Bicalho et al., 2009). Though

reduced digital cushion thickness is strongly associated with higher incidence of SU, WLD, and lameness in general (Lischer et al., 2002; Bicalho et al., 2009; Newsome et al., 2017a, 2017b; Stambuk et al., 2019), it is unclear whether thinning of the digital cushion is a cause or effect of claw disorders. Bicalho et al. (2009) concluded that thinning of the digital cushion after calving was a risk factor for CHDLs and recapitulated these conclusions in their review (Bicalho and Oikonomou, 2013). In contrast, Newsome et al. (2017b) found that thin sole soft tissue (i.e. corium and digital cushion thickness) is associated with higher future CHDL incidence; however, thinning of the sole soft tissues after calving is not associated with higher CHDL incidence in the future.

Because claw horn disruption lesions are classically associated with rumen acidosis, some have postulated that nutritional imbalance also contributes to risk. Specifically, carbohydrate overloading causes rumen acidosis, which is thought to increase laxity in the suspensory system that in turn leads to faulty horn production (Nocek, 1997; Nordlund et al., 2004; Stone, 2004). However, a study testing carbohydrate overloading and laxity of the suspensory system did not support this hypothesis (Danscher et al., 2010), although that study only tested one carbohydrate overload dosage and measured laxity at two time points after carbohydrate overloading. More evidence is needed to support or refute the association between subacute rumen acidosis, whether from feeding high concentrate diets or other reason, and the development of SU and WLD.

5 Prevention and treatment

Because DD is an infectious disease, the main method for prevention (among dairies that implement prevention) is the use of medicated foot baths because vaccines that had been developed were not effective, likely because of the polytreponemal nature of the infection, and

were discontinued (Ertze et al., 2006). Although foot baths are prophylactic for DD, the majority of dairies in the US did not use a footbath for cows in 2013, 12.5% used footbaths seasonally or occasionally, and 30.6% used footbaths throughout the year (USDA, 2018). Of the dairies that used footbaths regularly, 80.9% used them weekly or more frequently, with 25.8% of dairies using them daily (USDA, 2018). The most common disinfectant ingredients used in foot baths are copper sulfate and formaldehyde, which are environmental and health hazards, respectively (Epperson and Midla, 2007; NTP, 2016), which may discourage dairy farmers from implementing foot baths and emphasizes the need for additional, less toxic prevention methods, such as genomic selection.

When prevention strategies fail, DD is treated through topical antibiotics, most commonly oxytetracycline and lincomycin, under a bandage (Berry et al., 2010, 2012); however, these antibiotics are not as effective as penicillin, penicillin derivatives, erythromycin, azithromycin and gamithromycin based on laboratory testing (Evans et al., 2012). The common use of oxytetracycline may be because oxytetracycline does not enter the milk when used topically to treat DD (Britt et al., 1999), but the widespread use of these less effective antibiotics may contribute to the high rate of recurrence (Berry et al., 2012; Evans et al., 2016).

Because improper weight bearing is the primary cause of SU and WLD, prevention of these claw horn disruption lesions is achieved through regular claw trimming to maintain the proper foot angle and even wearing between the lateral and medial claws (Shearer and van Amstel, 2001). Preventing excessive and/or uneven wear to the claws to encourage proper weight bearing can also be achieved through minimizing walking distance, minimizing standing time, and using rubber flooring in alleyways (Vanegas et al., 2006; Fjeldaas et al., 2011; Eicher et al., 2013). Additionally, close monitoring of nutrition may prevent metabolic imbalances that

could theoretically lead to poor horn production (Shearer and van Amstel, 2017).

Similar to prevention methods for SU and WLD, treatment for SU and WLD aims at restoring proper weight bearing through corrective trimming. The necrotic horn around the lesion and, if present, protruding granulated corium tissue are removed. Antibiotics such as tetracycline or oxytetracycline and copper sulfate or ichthammol ointment are sometimes applied to the affected area to prevent contamination and reduce inflammation during healing (Kleinhenz et al., 2014), though some suggest that these topical antiseptics may actually delay wound healing (Shearer et al., 2015). Because corrective claw trimming typically causes bleeding, some advise the area should be bandaged (Shearer and van Amstel, 2001), though other studies have shown limited improvement in, or worse, healing rates with bandages compared to no bandages for SU and WLD (White et al., 1981; Pyman, 1997; Klawitter et al., 2019). After corrective trimming, a claw block is often added to the sound claw to transfer weight to the sound claw and improve healing of the affected claw (Greenough, 2007).

These prevention and treatment methods negatively impact the producers' profit margins, the environment, and in the case of corrective claw trimming to treat an advanced SU or WLD abscess, can cause additional pain and discomfort to cows. Methods to control lameness prevalence can potentially be enhanced by genomic selection against these foot lesions and associated lameness, thereby reducing incidence and the need for non-genetic means of prevention and treatment.

6 Genetic selection to reduce lameness incidence

As attention around lameness has grown in the industry, the USDA (Animal Improvement Programs Laboratory, AIPL, now part of the Animal Genomic Improvement Laboratory, AGIL) and private companies have likewise placed more emphasis on feet and leg

traits and, more recently, lameness. In 1996, the Holstein Association USA (HAUSA) introduced the feet and legs composite index (FLC) to enable selection for optimal feet and leg conformation, which would elongate productive life (HAUSA, 2017). Since 2000, the USDA-AIPL has incorporated FLC into the net merit index (NM$), which also includes traits for production, reproduction, and health and wellness. The HAUSA has since updated the FLC multiple times, and in the most recent update in 2020, FLC is comprised of foot angle, rear legs-rear view, rear legs-side view, feet and legs score, and stature. Although the NM$ includes the FLC and other conformation traits, it does not include an index for lameness specifically, let alone indices for individual claw disorders. On its most recent report update for NM$, the USDA-AGIL recognized that selection for lameness cannot be achieved via selection on FLC because FLC is a poor indicator of lameness (Van Raden et al., 2021).

To target lameness directly, private companies have developed selection indices that include lameness in addition to other feet and leg traits, with Zoetis releasing the first wellness trait evaluations that included a lameness component in 2016 in the US (Vukasinovic et al., 2017), and the Agriculture and Horticulture Development Board in the United Kingdom releasing the Lameness Advantage index in 2018 (Winters, 2018). The first national effort for the genetic evaluation of hoof health and lameness was proposed in 2016 in Canada, where the Canadian Dairy Network led a centralized effort to collect hoof health data and use it to develop genetic evaluations for lameness (CDN , 2016). Only very recently has the US followed suit, with the Council for Dairy Cattle Breeding announcing its intention in February 2021 to collaborate with public and private stakeholders to create the infrastructure for standardized recording of hoof health data and develop genetic evaluations specifically for hoof health traits (Burchard et al., 2021). This will facilitate direct selection on hoof health traits in the US instead

of hoof health traits being combined with other production, fertility, and conformation traits in selection indices like the net merit index.

Because noninfectious claw lesions are thought to result from uneven weight bearing between and within claws, some have proposed that selection for favorable feet and leg conformation can reduce the incidence of noninfectious claw lesions (Van der Waaij et al., 2005; Häggman et al., 2013) and therefore feet and leg conformation selection indices are sufficient. The underlying assumption for using feet and leg conformation traits to indirectly select on noninfectious claw lesion susceptibility is that the genetic correlation between the conformation and lesion traits must be strong. However, studies have shown that the genetic correlation between feet and leg conformation traits and noninfectious claw lesion susceptibility is low at best (Häggman and Juga, 2013; Malchiodi et al., 2017; Ring et al., 2018), which would prevent efficient indirect selection. The low genetic correlation may be partially because abnormal hind leg conformation can be both a cause and effect of claw lesions, creating a vicious cycle of poor conformation and lameness susceptibility (Capion et al., 2008). Consequently, in cases where abnormal hind leg conformation is the effect of claw lesions, selection on these conformation traits would not necessarily reduce claw lesion incidence because the conformation trait is an effect of, rather than genetically correlated with, the lesion. Because indirect selection using feet and leg conformation traits is not efficient, research has shifted to finding the genetic basis of lameness traits.

7 Previous studies about the genetic basis of hoof health

Susceptibility to DD, SU and WLD has a nontrivial genetic component, as implied by the breed predisposition of Holstein cattle to DD and SU (Holzhauer et al., 2006, 2008), and the low to moderate heritability estimates for these foot lesions. Holstein cattle are the most prevalent

breed of dairy cattle in the US (86% of US dairy cattle) (USDA, 2016), which underscores the large economic potential of developing genetic evaluation tools to reduce susceptibility to these foot lesions. The genetic component of susceptibility to DD, SU, and WLD has been quantified by using pedigree information (i.e., pedigree relations and the phenotypes of relatives) to estimate heritability, the proportion of phenotypic variation attributable to genetic variation. Heritability estimates of susceptibility to foot disorders vary widely in the literature, ranging from 0.01 to 0.4 for DD, 0.01 to 0.3 for SU, and 0.017 to 0.26 for WLD (Van der Waaij et al., 2005; Onyiro et al., 2008; van der Linde et al., 2010; Häggman and Juga, 2013; Oberbauer et al., 2013; van der Spek et al., 2013, 2015a; Malchiodi et al., 2015a; Biemans et al., 2018), thereby indicating the notable genetic contribution to susceptibility.

The low to moderate heritability of DD, SU, and WLD estimated from pedigree analyses implies that certain loci were contributing to susceptibility, leading to association studies to identify which loci were associated with susceptibility. Markers for loci were initially microsatellite markers (e.g., Buitenhuls et al. 2007) and, with the development of high-throughput genotyping, have since progressed to SNP panel genotypes (e.g., van der Spek et al. 2015b). SNP panel genotyping enabled genome-wide association studies (GWAS), in which SNPs or variants across the entire genome are tested for association with susceptibility for these foot conditions. This genome-wide approach contrasts with candidate gene approaches, in which only genes that might be functionally relevant are tested for association with phenotype (e.g., El-Shafaey 2017). In these GWAS, hoof health is commonly analyzed as binary traits for individual foot disorders that cause lameness [yes/no the cow was lame or had a foot lesion(s)], as composite binary traits that include certain groups of foot disorders, or as a composite trait consisting of any lameness event.

To date, the results of association studies are discordant; no overlap in detected susceptibility loci has been found among the studies within or across specific foot lesions or composite hoof health traits. Genome-wide association studies for DD susceptibility have found significant and suggestive SNP associations on *Bos taurus* autosomes (BTAs) 1, 3, 5, 6, 8, 9, 10, 14, and 26 (Scholey et al., 2013; Malchiodi et al., 2015b; Naderi et al., 2018; Biemans et al., 2019; Sánchez-Molano et al., 2019), or no significant or suggestive SNPs (van der Spek et al., 2015b). The two published GWAS for SU found suggestive SNPs on BTA8, 10, 11, 18, and 22 (van der Spek et al., 2015b) as well as BTA12 and 25 (Sánchez-Molano et al., 2019). Though few GWAS specifically for SU or WLD have been published to date, GWAS for other traits related to SU and WLD have been performed for digital cushion thickness (Sánchez-Molano et al., 2019; Stambuk et al., 2020), sole hemorrhage (van der Spek et al., 2015b; Sánchez-Molano et al., 2019), and laminitis (Naderi et al., 2018). Digital cushion thickness is strongly associated with SU, WLD, and lameness in general (Bicalho et al., 2009), and sole hemorrhage and laminitis are thought to be precursors to SU and WLD (Shearer and van Amstel, 2017). Other GWAS for hoof health traits examined trimming status (i.e., the need to claw trimming), laminitis-related claw disorders (double sole, sole hemorrhage, sole ulcers, white line separation), infectious foot disorders (DD and interdigital dermatitis) and heel erosion (van der Spek et al., 2015b), hoof health status determined by veterinarians and claw trimmer, total number of hoof disorders (Suchocki et al., 2020), and a composite feet and leg disorders index (Wu et al., 2016). The wide variation in susceptibility loci detected imply that susceptibility to individual foot lesions and composite lameness indices are complex traits strongly affected by the environment, and the remaining genetic component is comprised of many small effect loci contributing to susceptibility.

8 Genetic correlation between lameness and other health traits, milk production traits, and reproduction traits

The genetic correlation between lameness traits and other health traits (e.g., clinical mastitis, retained placenta, metritis) must also be considered during genomic evaluations so that unfavorable correlations can be minimized. Genetic correlation estimates for individual foot disorders and clinical mastitis or somatic cell count range from significantly different from zero (0.15 to 0.35) (Koenig et al., 2005; Buch et al., 2011) or close to zero (Gernand et al., 2012). To date, no other studies have estimated the genetic correlation between claw disorders and health traits other than mastitis traits. Also of particular interest is the genetic correlation of lameness with milk production traits (e.g., milk yield, fat/protein yield) and reproduction traits, though these traits are beyond the scope of this dissertation. Genetic correlation estimates between individual foot disorders and milk production traits (i.e., milk yield, fat and protein yield) range from unfavorable positive estimates (Pryce et al., 1997; Koenig et al., 2005, 2008; Buch et al., 2011), to estimates close to zero (Gernand et al., 2012), to favorable negative estimates (Onyiro et al., 2008). The wide range of estimates of genetic correlation between milk production traits and lameness traits reflects the disparity in lameness incidence in high-producing cows compared to that in low-producing cows. Some have found that lameness appears more frequently in high-producing cows during early lactation compared to low-producing cows (Collard et al., 2000; Green et al., 2014). In contrast, previous work in our lab has found that high-and low-producing cows have the same probability of hoof lesions (Oberbauer et al., 2013).

9 Objective

Lameness has been a longstanding prevalent issue caused primarily by DD, SU, and WLD. Though it is known that genetics plays a role in the susceptibility to these three disorders,

which loci influencing susceptibility is still unclear. Thus, the objective of this dissertation research is to identify genomic regions associated with the three most common causes of lameness: DD, SU, and WLD. Loci identified from this research could inform selective breeding programs by specifically targeting the loci associated with these foot diseases. This would reduce the incidence of DD, SU, and WLD and in turn improve animal welfare, producer profit, and the environmental impact of dairy production.

10 Overview of methodology

The first two experimental chapters detail investigations to identify loci contributing to DD (Chapter 2), SU, and WLD (Chapter 3). Because the genetics underlying complex traits are rarely exclusive and loci often affect multiple traits, the fourth chapter explores how the susceptibility loci for lameness traits fit within the broader context of other health traits by estimating the genetic correlation between lameness traits and other common health traits and identifying loci contributing to correlated traits. Knowledge of how lameness traits covary with other health traits will help avoid inadvertently selecting for increased susceptibility to one disease while selecting for lower susceptibility to lameness.

Chapter II. Genome-wide association studies reveal susceptibility loci for digital dermatitis in Holstein cattle

This chapter has been published in the MDPI Animals journal.

Citation: Lai, E., A.L. Danner, T.R. Famula, and A.M. Oberbauer. 2020. Genome-Wide Association Studies Reveal Susceptibility Loci for Digital Dermatitis in Holstein Cattle. Animals 10:2009. doi:10.3390/ani10112009.

Simple Summary: Digital dermatitis (DD), a leading cause of foot problems in dairy cattle, is a welfare concern and causes financial losses due to treatment and reduced milk production. Foot warts, or the technically correct term of digital dermatitis, result from a bacterial infection followed by delayed healing due to both genetic and environmental factors. Dairy farmers are already combatting DD through environmental control, but they do not have genetic selection tools because the genetics influencing DD susceptibility are largely unknown. We sought to identify the genetics associated with DD which can be incorporated into genetic selection tools. Farmers can then use these genetic selection tools to breed cows that are less susceptible to DD. We identified promising genes that play a role in the immune response and wound healing—immune functions that, if impaired, could increase a cow's susceptibility to DD. Though these genes were promising, their associated genetic markers had very little influence on DD susceptibility when compared to environmental management. Thus, the findings imply that the best approach for reducing DD prevalence is likely through combining a genetics approach with environmental management.

Abstract: Digital dermatitis (DD) causes lameness in dairy cattle. To detect the quantitative trait loci (QTL) associated with DD, genome-wide association studies (GWAS) were performed using high-density single nucleotide polymorphism (SNP) genotypes and binary case/control, quantitative (average number of DD per hoof trimming record) and recurrent (cases with ≥2 DD episodes vs. controls) phenotypes from cows across four dairies (controls $n = 129$ vs. DD $n = 85$). Linear mixed model (LMM) and random forest (RF) approaches identified the top SNPs, which were used as predictors in Bayesian regression models to assess the SNP predictive value. The LMM and RF analyses identified QTL regions containing candidate genes on *Bos taurus* autosome (BTA) 2 for the binary and recurrent phenotypes and BTA7 and 20 for

the quantitative phenotype that related to epidermal integrity, immune function, and wound healing. Although larger sample sizes are necessary to reaffirm these small effect loci amidst a strong environmental effect, the sample cohort used in this study was sufficient for estimating SNP effects with a high predictive value.

Keywords: digital dermatitis; foot warts; lameness; genome-wide association study; linear mixed model; random forest; Bayesian estimation; sustainability; animal welfare

1 INTRODUCTION

Lameness, or abnormal gait, affects 16% of dairy cows in the United States, making lameness the second most prevalent disease in dairy cattle after mastitis (USDA, 2018). Digital dermatitis (DD) is a common cause of lameness, comprising 70.9% and 36.0% of lameness cases in heifers and cows, respectively (USDA, 2018). The economic impacts of DD are $64 to $153 per episode due to reduced milk production, discarded milk, treatment costs, and additional labor (Cha et al., 2010; Dolecheck et al., 2019). Furthermore, premature culling obligates producers to expand their replacement heifer herd. Because heifers consume inputs without contributing to milk production, a larger replacement heifer herd inflates the economic cost (Hadley et al., 2006) and carbon footprint (Ratwan et al., 2015) per unit of milk. Thus, reducing the incidence of DD and associated lameness has great potential to benefit animal welfare, the producer's profit margin, and the environment, bolstering the three pillars of sustainability.

Heritability estimates for DD range from 0.01 to 0.4 (Onyiro et al., 2008; Oberbauer et al., 2013; Biemans et al., 2019), indicating genetic contributions to DD susceptibility along with a strong environmental influence. Reducing DD incidence, therefore, will likely be achieved through a combination of management and genetic approaches informed by the etiology of DD. Although the etiology of DD has not been completely elucidated, multiple bacterial phylotypes

belonging to the genus *Treponema* are consistently found in DD lesions (Brandt et al., 2011). Accordingly, the main environmental management method for reducing DD incidence is medicated foot baths (Holzhauer et al., 2012; Holzhauer, 2017), though this treatment is expensive, with estimates of ~$42 per cow per year (Cook, 2017). Additionally, the disinfectant compounds commonly used in foot baths raise environmental and health concerns, as the primary ingredients, copper sulfate and formaldehyde, are environmental pollutants (Epperson and Midla, 2007) and carcinogenic (2016), respectively. To alleviate these issues and improve DD prevention, some producers emphasize feet and leg conformation scores or indices that include claw health when selecting sires and, increasingly, rely upon genetic testing for heifers. However, the low genetic correlation between conformation traits and foot lesions impairs efficient indirect selection against foot lesions when using selection on conformation traits (van der Linde et al., 2010).

Currently, no selection index exists specifically for DD susceptibility. Targeted genetic selection against DD and associated lameness requires the identification of genomic regions influencing DD susceptibility. To find the contributing quantitative trait loci (QTL), genome-wide association studies (GWAS) have been undertaken, although the results to date have been discordant. Previous studies found significant and suggestive single nucleotide polymorphisms (SNPs) on *Bos taurus* autosomes (BTAs) 1, 3, 5, 6, 8, 9, 10, 14, and 26 (Scholey et al., 2013; Malchiodi et al., 2015b; Naderi et al., 2018; Biemans et al., 2019; Sánchez-Molano et al., 2019), or no suggestive or significant SNPs (van der Spek et al., 2015b).

In an effort to improve upon and refine past studies, the present study used strict phenotyping, dairies with similar management practices, and a high-density SNP genotyping array to identify the associations between DD and genomic regions. We hypothesized that certain

genetic markers would be associated with DD susceptibility, and that those markers would have small effects. Our results revealed multiple small-effect SNPs were associated with DD and defined QTL that contained candidate genes related to immune function and wound healing, supporting our hypothesis.

2 MATERIALS AND METHODS

Four commercial dairies in the Central Valley of California, two of which had participated in our previous heritability study (Oberbauer et al., 2013), provided hoof trimming records and blood samples from which to generate genotypic data. All the procedures were conducted in accordance with the ethical standards set by the University of California, Davis, and approved by the Institutional Animal Care and Use Committee.

2.1 Phenotypic data

Hoof trimming records were used to generate binary and quantitative phenotypes. Foot lesions and lameness issues were diagnosed by a single hoof trimmer servicing three of the dairies (dairies A, B, and C), and a different hoof trimmer servicing the fourth dairy (dairy D). The hoof trimmer servicing three dairies was trained by Dr. Steven Berry, a veterinarian specializing in foot lesions who offered hoof trimming training workshops to the industry and was a coauthor of our earlier paper (Oberbauer et al., 2013), and the other trimmer shadowed trained trimmers to standardize the diagnostics. Foot lesions were diagnosed and recorded while the cow was restrained. Foot lesion types and the foot with DD lesions were recorded into the cow's electronic record (dairies A and D) or maintained in a hard copy format (dairies B and C). Each type of foot lesion (e.g., DD, sole ulcer, laminitis, white line disease, foot rot, etc.) and miscellaneous lameness event (e.g., rock, cut, etc.) was tallied for each cow.

Both hoof trimmers utilized similar criteria for defining instances of DD in the cattle to

reduce the phenotypic classification variability; specifically, DD was recorded for cows with raised, inflamed lesions on the skin above the heel of the foot or above the interdigital space on the front of the foot. Routine hoof trimming varied by dairy: cows were trimmed at the beginning and middle of lactation, during dry off, and when exhibiting altered gait (dairy A); during dry off and when exhibiting altered gait (dairy B and C); and only when exhibiting altered gait (dairy D). Cases were defined as cows who had exhibited at least one DD lesion, whereas controls had no DD or other lameness records and were 6.5 years of age or older to avoid misphenotyping younger cows who had insufficient time to develop lameness events. Cows may have multiple bouts of DD over their lifetime and, in some cases, the first instance of DD may have occurred before the cows were in milk (i.e., as heifers). Because DD lesions typically last for four to six months (Read and Walker, 1998; Krull et al., 2016), we defined independent DD lesion episodes as those that were separated by at least six months. Cows with two or more independent DD episodes were considered recurrent cases. Digital dermatitis records that were less than six months apart were considered repeated records of one persistent DD episode. Digital dermatitis was analyzed as a binary phenotype to identify loci influencing general susceptibility to DD and as a quantitative phenotype calculated as the total number of independent DD lesions a cow had divided by the total number of hoof trimming records to standardize the number of lesions by the number of hoof trimming records for each cow. Consequently, the quantitative phenotype for a control cow was zero. Digital dermatitis was also analyzed as a recurrent phenotype (cases with ≥2 DD episodes vs. controls) to identify the loci contributing to reoccurring DD episodes.

2.2 Genome-wide association and linear mixed model analyses

Genomic DNA was extracted from whole blood samples using the QIAGEN QIAamp DNA Blood Mini Kit (QIAGEN Inc., Valencia, CA, USA) and quantified using the NanoDrop

(ND-2000 v3.2.1) spectrophotometer (Thermo Scientific, Wilmington, DE, USA). DNA samples were genotyped on the BovineHD BeadChip (777962 SNPs, Illumina Inc., San Diego, CA, USA) by GeneSeek (Lincoln, NE, USA). Raw and processed microarray data were submitted to the NCBI Gene Expression Omnibus database (GEO series record GSE159157). Illumina's GenCall algorithm was used to call genotypes.

GWAS were performed using the binary, quantitative, and recurrent phenotypes using the SNP coordinates from the ARS-UCD1.2 map (accessed August 2020 from the National Animal Genome Research Project's Cattle Genome Analysis Data Repository (https://www.animalgenome.org/repository/cattle/UMC_bovine_coordinates/), version last modified 11 September 2018). The quality filtering of SNP genotypes was performed using PLINK 1.9 (Chang et al., 2015; Purcell and Chang, 2015) to remove from further analysis any cows having less than 5% of all SNPs genotyped and SNPs missing genotypes in more than 5% of the cows. SNPs with a minor allele frequency of less than 0.05 were removed to exclude rare variants, and SNPs that deviated significantly from the Hardy–Weinberg equilibrium ($p < 1 \times 10^{-6}$) in controls were removed to exclude systematic genotyping errors.

Family structure is extremely prevalent in the dairy population from breeding elite bulls to hundreds to tens of thousands of cows. To visualize the genetic similarity among cows at this initial dairy, a multi-dimensional scaling (MDS) analysis was performed and the first two dimensions were plotted. The GWA analyses were performed using the genetic analysis program Genome-wide Complex Trait Analysis (GCTA) (Yang et al., 2011) to fit a linear mixed model (LMM) that tests for the association of SNP genotypes with binary and quantitative DD phenotypes. An LMM was selected for its ability to incorporate a genetic relatedness matrix to correct for familial relatedness and population structure. Linear mixed models are designed for

quantitative phenotypes, as LMMs assume that phenotypes are normally distributed; however, LMMs have also been routinely used to analyze binary traits (Lippert et al., 2011; Zhou and Stephens, 2012). A genetic relatedness matrix was computed and included along with farm as a covariate in the LMM. When fitting the LMM for each SNP, the LMM included the chromosome of the candidate SNP being tested. To reduce false positive associations due to multiple testing across many loci without being overly stringent, the effective number of independent SNPs (M_e) after linkage disequilibrium (LD) pruning was determined using the Genetic Type I error calculator (GEC) and used as the denominator for Bonferroni-corrected thresholds (Li et al., 2012). Significant SNPs were defined as those with $p < 0.05/M_e$, whereas suggestive SNPs were defined as having $p < 1/M_e$ (Lander and Kruglyak, 1995). To calculate the genomic inflation factors (λ_{GC}), chi-squared test statistics were first generated from association *p*-values, and the median of the resulting chi-squared distribution was divided by the median of the expected chi-squared distribution. Quantile-quantile plots (qqplots) and Manhattan plots were plotted in R (R Development Core Team, 2010) using the package qqman (Turner, 2014).

2.3 Random forest analysis

Random forest (RF) analysis was performed as an additional method for identifying SNPs that appeared to importantly contribute to disease phenotypes. Random forests do not make any assumptions about the inheritance model (additive, dominant, recessive) and are able to test multiple SNPs jointly for association with phenotype. Additionally, the RF approach is unaffected by an uneven farm distribution of cases and controls because RF builds decision trees and estimates the importance of each feature by the frequency it appears in the decision trees, rather than estimating parameters for a model. Consequently, RFs avoid estimating parameters for which there are no data. These properties make RFs well equipped to identify structure within

complex genetic architectures like DD susceptibility. Specifically, RF can accommodate data despite uneven sampling across farms, in which contributing SNPs may have different modes of inheritance and where epistasis is likely prevalent.

After converting quality-filtered binary PLINK files into VCF files split by chromosome in PLINK 1.9 (Purcell et al., 2007; Chang et al., 2015), all the missing genotypes were imputed using BEAGLE 5.1 (Browning et al., 2018) because the RF analysis cannot handle missing genotypes. The resulting VCF files were converted back to binary PLINK files, which were LD-pruned using a threshold of $R^2 \geq 0.90$ to avoid diluting the importance of SNPs in strong LD during the RF analysis (Goldstein et al., 2010) and recoded to additive and dominant component files suitable for importing into R. The additive component (i.e., genotypes coded as 0/1/2 minor alleles) was used as input for the RF analysis in R using the caret package (Kuhn, 2008; R Development Core Team, 2010). For binary and quantitative phenotypes, RF analysis was implemented with all genome-wide SNPs in one run to estimate the relative importance of explainers, comprised of SNP genotypes and farm. For both runs, the same random sample of two thirds of the cows was used to train the model and calculate variables of importance for each explainer. The RF run for each phenotype built 500 decision trees that included three values of *mtry,* the number of predictors considered at each node of the tree. The value of *mtry* that yielded the most accurate model was used as the final model. The most important explainer was assigned an importance variable of 100, and the other explainers were assigned importance variables relative to the most important explainer (e.g., an explainer with an importance of 50 is 50% as important as the most important explainer). To assess the accuracy of the final model, the remaining third of cows was used as the test population, using the explainers and their relative importance to predict phenotype.

After evaluating the model accuracy using the test population, a threshold of importance was determined by ranking and plotting the SNPs the RF identified as important for each chromosome in a scree plot and finding the rank of the second-order point of inflection using the d2uik option in the inflection package in R (Christopoulos, 2016, 2017b). SNPs ranking equally as or more important than this threshold were considered important and included in further analyses.

2.4 Bayesian regression to assess model predictability and validation

To assess the collective predictive ability of the top SNPs identified in the LMM and RF analyses, the top SNPs from each analysis (i.e., significant and suggestive SNPs from LMM analyses, important SNPs from RF analyses) were tested for association with phenotype using Bayesian regression. Bayesian regression was selected because of its ability to fit multiple SNPs simultaneously while also recognizing that the majority of SNPs have small effects on DD susceptibility (van der Spek et al., 2015b; Biemans et al., 2018), that some SNPs are likely correlated due to LD, and that not all farms contributed controls to the analyses. Additionally, Bayesian regression enables the thorough evaluation of model fit through leave-one-out (LOO) validation and posterior predictive checking (PPC), the latter of which is a uniquely Bayesian feature.

Suggestive and significant SNPs from the LMM GWAS and important SNPs from the RF analysis were used as predictors along with farm in each Bayesian regression model. Similar to the RF analyses, SNP genotypes were coded as 0/1/2 minor alleles. A Bayesian regression model was fitted for each combination of GWAS method (LMM and RF) and phenotype (binary and quantitative), such that four models were fitted: LMM-binary and RF-binary were fitted using a Bayesian logistic regression model, and LMM-quantitative and RF-quantitative were fitted using

a Bayesian generalized linear model for continuous data. Susceptibility to DD appears to be complex and the majority of SNP effects are likely to be small (van der Spek et al., 2015b; Biemans et al., 2018). To reflect this distribution of SNP effects, a normal prior with a small-scale N (0,1) was used for the distribution of predictors for all four models. Each of the four models was fitted by sampling from the posterior distribution using the Hamiltonian Monte Carlo algorithm, a Markov chain Monte Carlo (MCMC) algorithm using the rstanarm package in R (Goodrich et al., 2020). Four parallel chains sampled the posterior distribution, and each chain was run for 10,000 iterations with a warmup of 2500 iterations, keeping every 25th iteration to avoid autocorrelation.

Unlike frequentist regression, which would output a point estimate of each SNP effect, Bayesian regression outputted a distribution of where the true value of each SNP effects fell, defined by the Bayesian uncertainty interval (UI). SNPs with 95% UIs that did not include zero were considered significantly associated with DD susceptibility. For each significant SNP, the probability of disease given a genotype at the significant SNP (coded as 0/1/2 minor alleles) and a 0 genotype at all other SNPs was calculated using the median of SNP effect estimates as point estimates in the inverse logit equation using the R package arm (Gelman et al., 2020). Diagnostic and Bayesian UI plots for the posterior medians of SNP effects were plotted using the bayesplot package. Leave one out cross validation was performed using the loo package (Vehtari et al., 2017, 2020) in R to predict the phenotype of each cow using the SNP effects estimated from all other cows. The reliability of prediction was assessed using the Pareto *k* diagnostic values outputted from the LOO analysis. Posterior predictive checking (PPC) from the bayesplot package (Gabry et al., 2019b) was used to assess the goodness of fit of the model. Posterior predictive checking assessed how well the estimated predictor effects were able to simulate

phenotypes with a similar distribution to that of the observed phenotypes.

2.5 Defining and annotating QTL regions

For the significant and suggestive SNPs identified in the LMM analyses and the important SNPs identified in the RF analyses, the QTL boundaries and regions were defined and annotated. Because SNPs are more likely to be in LD with causal variants than be causal themselves, the linkage disequilibrium in the regions flanking these top SNPs was used to define the boundaries of QTL, per the methods used in previous GWAS studies (Richardson et al., 2016; Twomey et al., 2019). Specifically, SNPs within 5 Mb of significant and suggestive SNPs that were also in LD ($r^2 \geq 0.5$) were considered as belonging to the same QTL. The SNPs furthest upstream and downstream that were in LD with the target suggestive or significant SNP defined the boundaries of the QTL. Overlapping QTL were combined into one QTL. QTL from the LMM and RF analyses were compared to discern whether the two analyses found the same QTL. QTL regions that were identified in both LMM and RF analyses were explored for candidate genes. Additionally, QTL defined by SNPs that were significant in the Bayesian regression analyses were also explored for candidate genes. Candidate genes were defined as genes falling in QTL regions identified in both LMM and RF analyses or in QTL defined by SNPs that were significant in Bayesian regression and were functionally relevant to DD etiology.

To annotate the QTL regions, the genomic regions search in FAANGMine v1.1 (Functional Annotation of Animal Genomes (FAANG) Consortium, 2019) using the ARS_UCD1.2 assembly was implemented to find genes within the QTL regions. The RefSeq identifiers of genes within the QTL were used in a gene ontology and pathway enrichment analysis in FAANGMine to discern whether the genes belonged to higher-order functions and pathways related to DD etiology. For the gene ontology and pathway enrichment analyses, the

Benjamini Hochberg test correction was used to correct for multiple testing, and all the RefSeq genes in *B. taurus* were used as the background population. To identify the functions of individual genes, protein coding genes in QTL defined by SNPs that were significant in two analyses (i.e., LMM, RF, and/or Bayesian regression) were searched in the Mouse Genome Informatics batch query database (http://www.informatics.jax.org/batch) using the mammalian phenotype option (Smith and Eppig, 2009).

3 RESULTS

3.1 Descriptive data

Hoof trimming records for 1382 DD-affected cows at dairies A, B, and D from 2002 to 2019 were used to calculate the age of onset statistics. Dairy C did not have hoof trimming records from the beginning of the cows' lives and was thus excluded from calculating the age of onset statistics. The average age of onset for the first episode of DD observed was 3.7 (SD = 1.6) years old and the median was 3.5 years old, indicating a minimum age of 6.5 years old for controls was sufficiently stringent to avoid misphenotyping younger cows. The cases and controls were sampled from 2013 to 2020. Cases were sampled from all four dairies, whereas only dairies A and D had control cows that met our stringent age and soundness requirements (Table II-1). In total, 222 cows were genotyped (cases n = 90, controls n = 132), of which six were removed during quality filtering (cases n = 3, controls n = 3), leaving 216 cows for analysis (cases n = 87, controls n = 129). Of the 87 cases, 24 had recurrent DD episodes and were used in the GWAS of controls vs. recurrent DD cases. Forty-seven percent of the DD cases had no other foot lesions other than DD during their lifetime. The remaining cases had, in addition to clearly identifiable DD, abscesses, sole fracture, sole ulcers, or bruising. One cow also had foot rot in addition to DD. Of these other foot lesions, only foot rot was considered infectious, whereas the

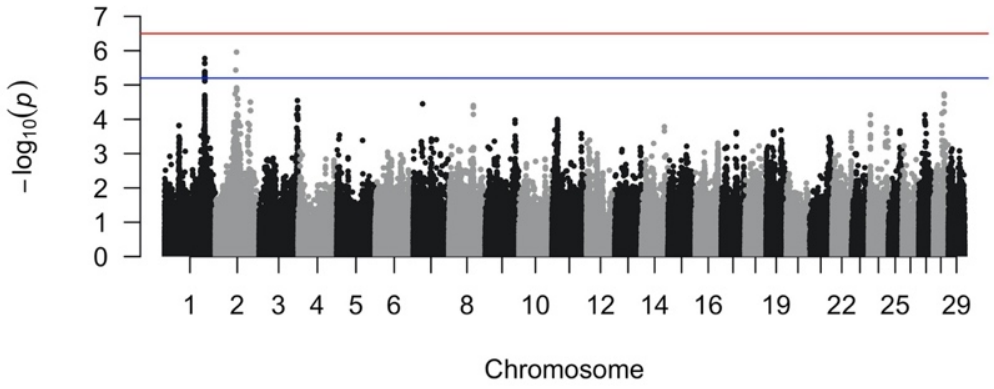
other concomitant lesions were noninfectious and associated with excessive wear of the claw due to hard flooring and/or metabolic issues (Shearer and van Amstel, 2017).

Table II-1. Distribution of digital dermatitis cases and non-lame controls across the four dairies.

Farm	Case	Control	Total
A	19	112	131
B	22	0	22
C	30	0	30
D	16	17	33
Total	87	129	216

After quality control filtering, 560,277 SNPs remained for the LMM analysis, and 222,060 SNPs (40%) for the RF analyses remained after LD pruning ($r^2 > 0.90$). The MDS analysis indicated no obvious population stratification (Figure II-S1). The effective number of SNPs (i.e., SNPs that were not in LD) was approximately 158,000 SNPs, yielding a cutoff of significance at 3.2×10^{-7} or 6.5 on the $-\log_{10}(p)$ scale and a suggestive cutoff at 6.3×10^{-6} or 5.2 on the $-\log_{10}(p)$ scale. Manhattan plots for the LMM binary and quantitative analyses are shown in Figure II-1 and suggestive and significant SNPs, in Table II-2 and Table II-3. For the recurrent LMM GWAS, the Manhattan plot is depicted in Figure II-S2 and suggestive and significant SNPs in Table II-S1. The genomic inflation factors were 0.97 for the binary and quantitative GWASs and 1.0 for the recurrent GWAS; when considered in conjunction with the qqplots, the analyses sufficiently accounted for population structure (Figure II-S3). In separate analyses, we removed outlier control cows, defined as having a value < -0.10 in the first coordinate and a value < -0.08 in the second coordinate of the MDS plot, and the conclusions of association remained unchanged (Figure II-S4). Our method of correction for multiple testing (i.e., using the effective number of independent SNPs as the denominator for Bonferroni correction) resulted in more stringent significance thresholds than those based on false discovery rate that are used in other GWASs for DD (Malchiodi et al., 2015b; van der Spek et al., 2015b; Biemans et al., 2019).

(**a**)



(**b**)

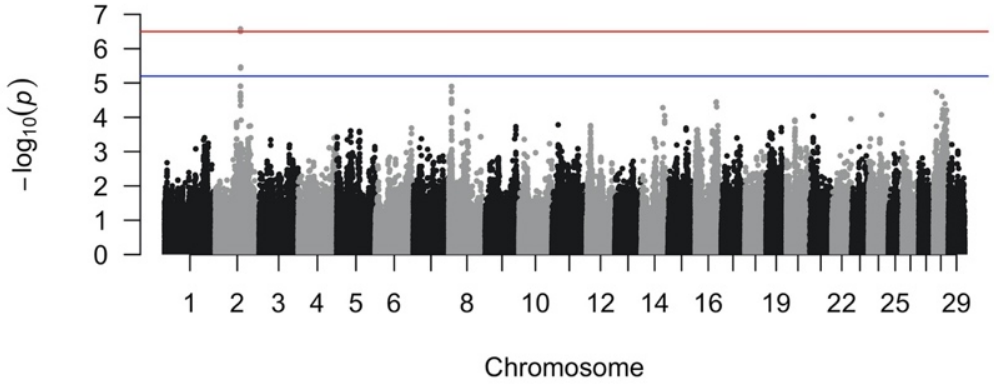


Figure II-1. Manhattan plots from the linear mixed model genome-wide association analyses using (a) binary phenotypes designating the presence of digital dermatitis (DD) lesions or the absence of any lameness issues and (b) quantitative phenotypes calculated by dividing the number of DD episodes by the total number of hoof trimming records. The red line indicates the threshold for genome-wide significance (Bonferroni-corrected using the number of independent SNPs at $p < 0.05$), and the blue line indicates the threshold for suggestive significance (Bonferroni-corrected using the number of independent SNPs at $p < 1$).

Table II-2. Suggestive SNPs detected from the linear mixed model genome-wide association analysis using binary phenotypes and their defined QTL.

				Minor Allele Count		MAF [a]							
SNP ID	BTA	SNP Position (bp)	Minor/ Major Allele	Cases (2n = 174)	Controls (2n = 258)	Cases	Controls	Effect Size (SE)	*p*	Significance in Bayesian Regression	QTL Start Position (bp)	QTL End Position (bp)	QTL Size (kb)
BovineHD0100035768	1	125563251	A/G	63	54	0.362	0.209	0.178 (0.037)	1.68×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035771	1	125565548	G/A	63	54	0.362	0.211	0.175 (0.037)	2.31×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035773	1	125567245	T/C	63	55	0.362	0.213	0.175 (0.037)	2.35×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035776	1	125570173	G/T	63	55	0.362	0.213	0.175 (0.037)	2.35×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035780	1	125573042	G/A	63	55	0.362	0.213	0.175 (0.037)	2.35×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035783	1	125576193	G/A	63	55	0.362	0.213	0.175 (0.037)	2.35×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035788	1	125598084	G/A	63	57	0.362	0.221	0.164 (0.036)	5.36×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD4100000712	1	125598643	T/C	63	57	0.362	0.223	0.163 (0.036)	6.31×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035789	1	125599413	C/T	63	57	0.362	0.221	0.164 (0.036)	5.36×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035796	1	125608174	A/G	64	56	0.368	0.217	0.163 (0.036)	4.59×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
ARS-BFGL-NGS-113021	1	125609019	C/T	64	56	0.368	0.217	0.163 (0.036)	4.59×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035797	1	125609959	C/T	64	56	0.368	0.217	0.163 (0.036)	4.59×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035802	1	125627579	C/T	64	56	0.368	0.217	0.163 (0.036)	4.59×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035803	1	125628401	A/C	64	56	0.368	0.217	0.163 (0.036)	4.59×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035828	1	125680990	G/A	87	88	0.500	0.341	0.155 (0.034)	4.11×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035829	1	125681850	C/A	87	88	0.500	0.341	0.155 (0.034)	4.11×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
ARS-BFGL-NGS-100109	1	125683184	C/T	87	88	0.500	0.341	0.155 (0.034)	4.11×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035833	1	125688941	C/T	87	88	0.500	0.341	0.155 (0.034)	4.11×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035841	1	125700410	A/G	87	89	0.500	0.345	0.152 (0.034)	5.85×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035842	1	125700857	C/T	87	89	0.500	0.345	0.152 (0.034)	5.85×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035844	1	125702010	C/T	87	89	0.500	0.345	0.152 (0.034)	5.85×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0100035845	1	125702906	G/T	87	89	0.500	0.345	0.152 (0.034)	5.85×10^{-6}	ns	125550933 [b]	125822143 [b]	271.21 [b]
BTA-47853-no-rs	2	63365256	A/G	78	64	0.448	0.248	0.167 (0.036)	3.69×10^{-6}	s	60971364	63389576	2418.2
BovineHD0200019142	2	65836042	G/A	41	32	0.236	0.124	0.224 (0.046)	1.10×10^{-6}	s	65836042	65836042	-

[a] MAF = minor allele frequency.

[b] This QTL is defined in both the linear mixed model and random forest analyses for the binary case-control phenotype. s = SNP effect

estimated from Bayesian regression was significantly different from zero, as defined by the 95% uncertainty interval. ns = SNP effect estimated from Bayesian regression was not significantly different from zero, as defined by the 95% uncertainty interval.

Table II-3. Important SNPs from random forest analysis using binary phenotypes and their defined QTL. Importance variables are expressed as % importance relative to farm (i.e., farm was 100% importance).

				Minor Allele Count		MAF [a]						
SNP ID	**BTA**	**SNP Position (bp)**	**Minor/ Major Allele**	**Cases (2n = 174)**	**Controls (2n = 258)**	**Cases**	**Controls**	**SNP Importance (% Relative to Farm)**	**Significance in Bayesian Regression**	**QTL Start Position (bp)**	**QTL End Position (bp)**	**QTL Size (kb)**
BovineHD0100001686	1	5894509	G/A	54	75	0.310	0.291	70.9	ns	5894509	5901795	7.3
BovineHD0100013452	1	47090630	C/T	41	25	0.238	0.098	75.4	ns	43459206	49409839	5950.6
BovineHD0100013140	1	45742004	G/A	35	20	0.201	0.078	75.6	ns	43459206	49409839	5950.6
BovineHD0100013551	1	47618749	T/G	50	27	0.291	0.105	81.5	ns	43459206	49409839	5950.6
BovineHD0100033878	1	118845470	A/G	9	41	0.052	0.159	76.1	ns	114235013	119003717	4768.7
BovineHD0100035876	1	125811728	A/C	70	68	0.402	0.264	79.4	s	125550933 [b]	125822143 [b]	271.21 [b]
BovineHD0200017030	2	59626300	C/T	22	95	0.126	0.368	77.0	s	58016533	59967789	1951.3
BovineHD0200037724	2	129189118	T/C	35	73	0.201	0.283	83.0	ns	128495987	129671807	1175.8
BovineHD0300035231	3	119898047	T/G	52	53	0.299	0.205	76.7	s	119720909	119942789	221.9
BovineHD0400033808	4	115632631	A/G	84	92	0.483	0.357	76.5	ns	115461900	115812750	350.9
ARS-BFGL-NGS-111175	4	119082548	A/C	38	30	0.218	0.116	76.8	ns	116927673	119130213	2202.5
BovineHD0400034694	4	117654227	G/A	53	119	0.305	0.461	76.9	ns	116927673	119130213	2202.5
BovineHD0700005793	7	19675119	C/T	87	102	0.500	0.395	75.5	ns	17910021	19773720	1863.7
BovineHD0700016221	7	54331048	A/G	6	42	0.034	0.163	77.0	ns	49401649	54505899	5104.3
BovineHD1300007641	13	26082265	C/T	69	140	0.397	0.543	76.0	ns	22185154	26101077	3915.9
BovineHD1500016894	15	57724182	A/G	60	59	0.345	0.229	72.9	ns	56807906	58102169	1294.3
BovineHD1600016687	16	58237523	C/T	105	108	0.603	0.419	81.8	ns	56372228	62230342	5858.1
BovineHD1700012893	17	45209840	T/C	62	52	0.356	0.202	80.4	ns	44418753	45224548	805.8
BovineHD1800003369	18	9579005	T/C	100	102	0.575	0.395	79.6	ns	9510127	9582839	72.7
BovineHD1800012376	18	41782168	C/T	27	8	0.155	0.031	88.4	ns	41753915	41863187	109.3
ARS-BFGL-BAC-35025	18	47814171	G/A	32	84	0.184	0.326	79.8	s	47099464	47831459	732.0
BovineHD1900013252	19	46915144	C/T	27	90	0.155	0.349	86.2	ns	46871178	47070613	199.4
BovineHD2200002436	22	8104318	A/G	36	106	0.207	0.411	79.8	ns	7974675	8109630	135.0
BovineHD2200002746	22	9090720	A/G	17	77	0.098	0.298	85.0	ns	9068141	9090720	22.6
BovineHD2600011849	26	42398008	A/G	59	68	0.339	0.264	75.6	ns	40792161	43877138	3085.0
ARS-BFGL-NGS-117055	27	12656552	C/T	86	89	0.494	0.348	75.7	ns	12202138	12834272	632.1

[a] MAF = minor allele frequency. [b] This QTL is defined in both the linear mixed model and random forest analyses for the binary case-control phenotype. s = SNP effect estimated from Bayesian regression was significantly different from zero, as defined by the 95% uncertainty interval. ns = SNP effect estimated from Bayesian regression was not significantly different from zero, as defined by the 95% uncertainty interval.

3.2 SNPs associated with DD as a binary phenotype

The binary LMM GWAS detected 22 suggestive SNPs on BTA1 that fell in the last three introns of *SLC9A9* and two suggestive intergenic SNPs on BTA2 (Table II-2). When used to define QTL boundaries, the 22 suggestive SNPs on BTA1 were all in LD and defined one 271.2 kb QTL region at BTA1:125550933–125822143 containing three genes: a long-noncoding RNA gene (LOC112447746), a tRNA-CAU gene, and *SLC9A9*. The BTA2:63365256 (BTA-47853-no-rs) SNP on BTA 2 identified a 2.4 Mb QTL region at BTA2:60971364–63389576 containing 25 genes, whereas the other SNP identified on BTA 2, BTA2:65836042 (BovineHD0200019142), was not in LD with neighboring SNPs ($r^2 < 0.5$). Because the number of genes discovered from the LMM QTL was limited, no gene ontologies or pathways were overrepresented.

When suggestive SNPs from the LMM-binary GWAS were used as predictors in the Bayesian regression models, MCMC sampling was able to efficiently explore the posterior. Though the effects of SNPs on BTA1 were not significantly different from zero at 50% UI, the effects of the two SNPs on BTA2 (BTA-47853-no-rs and BovineHD0200019142) were significantly different from zero at 95% UI (Figure II-2, Table II-2). Unlike a frequentist 95% confidence interval, which defines the range within which the true value of the SNP effect falls 95% of the time in repeated sampling, a Bayesian 95% uncertainty interval indicates there is a 95% probability that the true value of the SNP effect falls within the range. For example, to give context for the impact of SNP effect size, each minor allele at BTA-47853-no-rs and BovineHD0200019142, respectively, increased the log odds of having DD by 1.3 and 1.5, using the median as the point estimate for SNP effect. A 1.3 increase in the log odds of having DD for each minor allele at BTA-47853-no-rs corresponded to an increase in the probability of having

DD by 22% and 54% for heterozygotes and homozygotes of the minor allele relative to homozygotes of the major allele. A 1.5 increase in the log odds of having DD for each minor allele at BovineHD0200019142 corresponded to a 25% and 60% increase in the probability of having DD for the heterozygotes and homozygotes of the minor allele, relative to the homozygotes of the major allele. The relatively large increases in the probability of having DD from each additional minor allele reflects the high minor allele frequency in cases (45%) relative to controls (25%) in this population. Additionally, the magnitude of increase in the probability of DD also depended upon the genotype of the cow at other SNPs. For instance, a cow with a genotype other than homozygous major for all SNPs could have a smaller increase in the probability of DD with each additional minor alelle at BTA-47853-no-rs or BovineHD0200019142.

Using the LMM-binary suggestive SNPs as predictors in the LOO analysis, Pareto k diagnostic values were acceptable ($k \leq 0.7$) for all cows, indicating that the estimated SNP effects were collectively predictive of phenotype within the original population. The LOO analysis indicated that the effective number of predictors in the model was 6.6, considerably lower than the 27 predictors that were actually in the model due to correlated predictors: the SNPs on BTA1 were in LD, and this correlation among predictors reduced the effective number of predictors. The PPC indicated that the observed and simulated data were similar to each other (Figure II-S5), supporting that the predictor estimates were collectively predictive of phenotype.

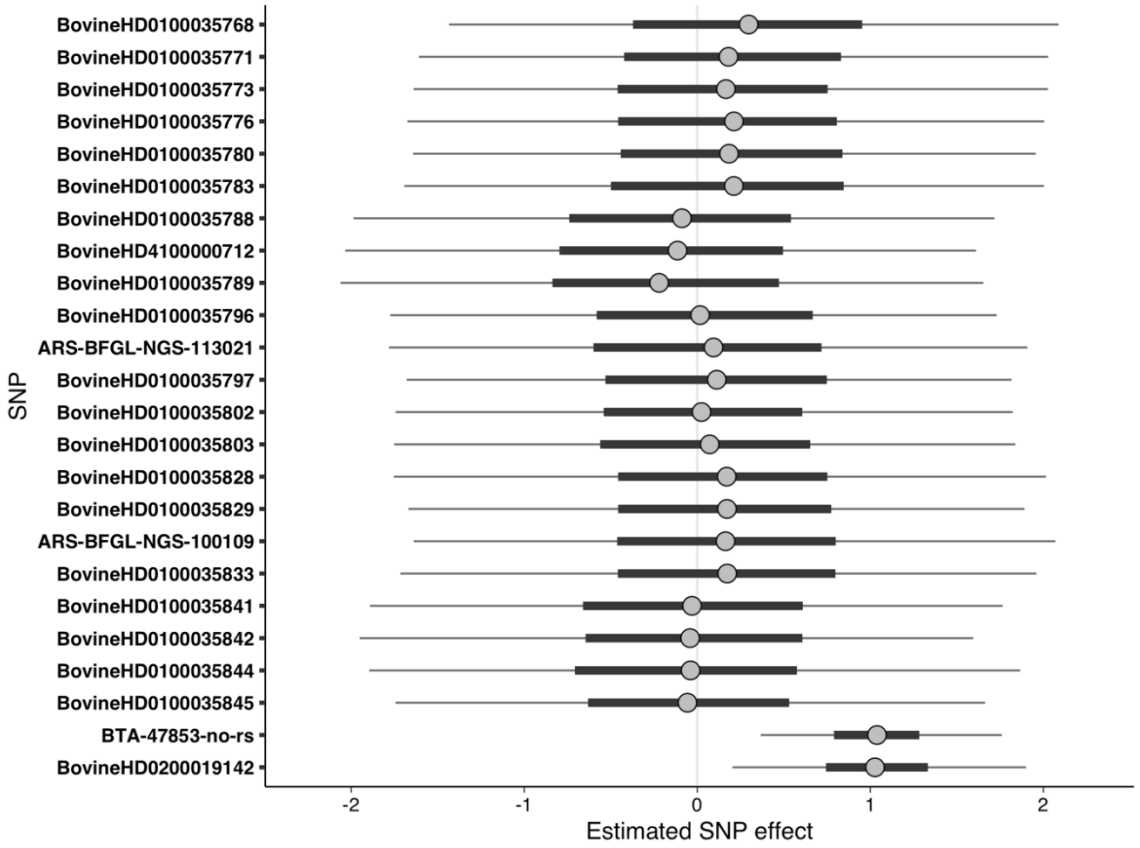


Figure II-2. Uncertainty interval (UI) plot for suggestive SNPs on BTA1 (BovineHD0100035768 through BovineHD0100035845) and BTA2 (BTA-47853-no-rs and BovineHD0200019142) from the linear mixed model GWAS using binary phenotypes. Dots represent the median of the SNP effect estimates from Markov chain Monte Carlo draws, thick bars indicate the 50% UI, and the thin lines indicate the 95% UI. SNPs with 95% UI not overlapping zero were considered significant. Positive values of predictor effect estimates indicate a higher risk of DD, whereas negative values indicate a lower risk of DD.

Random forest analysis revealed that farm was ranked as the most important explainer, and consequently the importance of SNPs was expressed as the percentage of importance relative to farm. Of the three values of *mtry* that were tested (6, 666, and 222,061), *mtry* = 666 yielded the most accurate model and was selected for further analyses. The accuracy of the selected model (0.69 with 95% CI 0.57–0.80) was not significantly different from the baseline no information rate (in this case, the proportion of controls: 0.64, $p = 0.20$), indicating that the model was unable to call case and control phenotypes more accurately than simply calling the more common phenotype. Random forest analyses found 26 important SNPs from the RF-binary,

and using LD to determine the QTL boundaries defined 23 QTL for the RF-binary dataset (Table II-3), one of which was the same QTL on BTA1:125550933–125822143 identified from the LMM-binary GWAS. Within the RF-binary QTL, FAANGMine found 566 genes, of which 129 and 188 were used in the pathway and gene ontology enrichment analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway herpes simplex virus 1 infection and the Reactome pathways P2Y receptors and nucleotide-like (purinergic) receptors were significantly enriched (Benjamini Hochberg $p = 0.003$, 0.021, and 0.035, respectively).

When important SNPs from the RF-binary analyses were used as predictors in the Bayesian logistic regression model, four SNPs had estimated effects that were significantly different from zero, including the SNP defining the QTL at BTA1:125550933–125822143 (Table II-3, Figure II-3). The important SNPs from the RF-binary analyses were not as predictive of phenotype within the population compared to the suggestive SNPs from the LMM-binary analysis, as evidenced by 13% of cows having high Pareto k diagnostic values ($k > 0.7$) from the LOO analysis. The lower predictability indicates that the RF was able to find small effect SNPs, but also found some noninformative SNPs. Though the PPC indicated that the observed and simulated data were similar to each other (Figure II-S6), this similarity was likely due to overfitting.

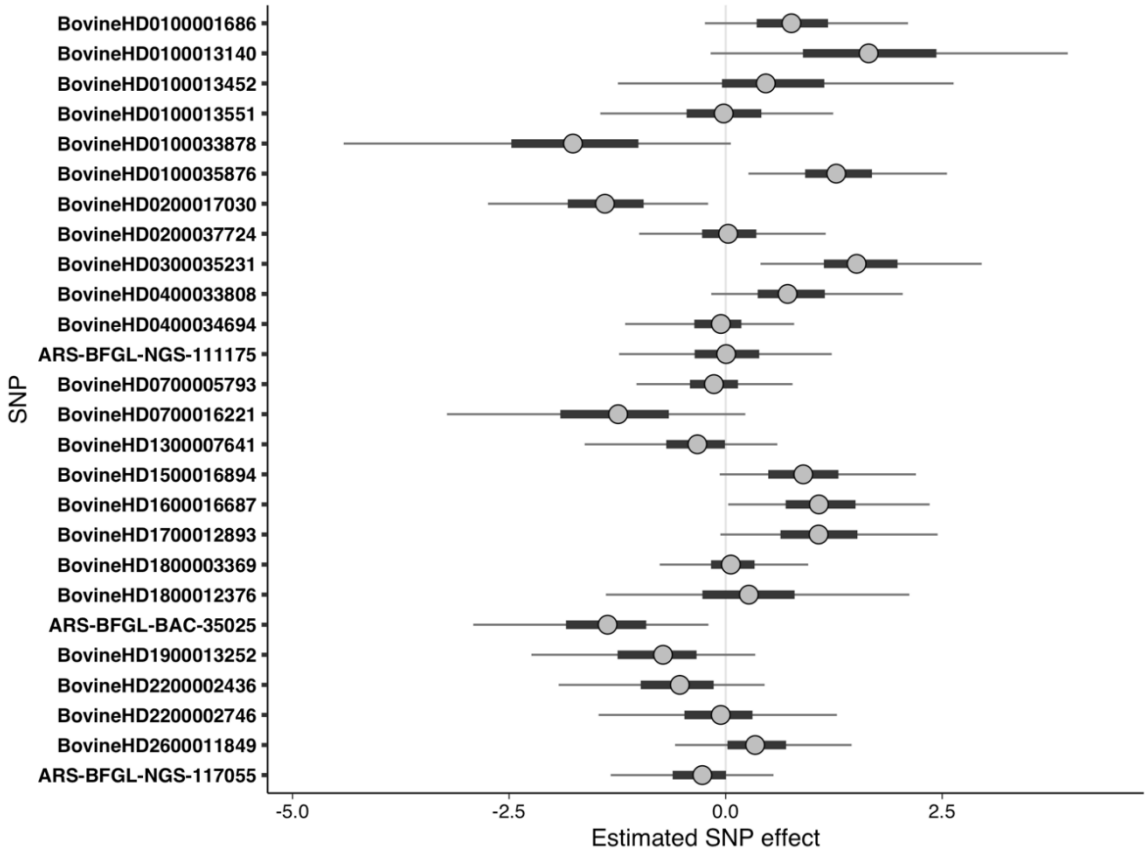


Figure II-3. Uncertainty interval (UI) plot for important SNPs from the random forest analysis using binary phenotypes. Dots represent the median of the SNP effect estimates from the Markov chain Monte Carlo draws, thick bars indicate the 50% UI, and the thin lines indicate the 95% UI. SNPs with 95% UI not overlapping zero were considered significant. Positive values of predictor effect estimates indicate a higher risk of DD, whereas negative values indicate a lower risk of DD.

3.3 SNPs associated with DD as a quantitative phenotype

The quantitative LMM GWAS identified seven significant and two suggestive SNPs, all of which were intergenic (Table II-4). The gene nearest to these nine SNPs was a suppressor of cytokine-signaling 6-like pseudogene (LOC615204) falling between the seven significant and two suggestive SNPs. When these nine SNPs were used to determine the QTL boundaries, all nine SNPs were in LD ($r^2 > 0.5$) and defined a 2 Mb QTL region at BTA2:77930065–79925981 (Table II-4). This 2 Mb QTL region included nine genes, including LOC615204. The recurrent DD cases vs. controls placed more emphasis on finding genetic differences between controls and

cases with more DD cases, similar to the LMM-quantitative GWAS; however, the LMM GWAS using recurrent DD cases vs. controls identified QTL regions in common with the LMM-binary and not the LMM-quantitative GWAS. In the recurrent GWAS, the same SNPs observed on BTA1 from the LMM-binary analyses formed a peak of association but did not reach suggestive significance, whereas three SNPs on BTA2 in addition to the two detected in the LMM-binary GWAS reached suggestive significance (Figure II-S2, Table II-S1). The three additional suggestive SNPs on BTA2 revealed by the recurrent analysis defined a 328 kb QTL at BTA2:65836042–66217730 that was not in LD with the QTL at BTA2:60971364–63389576 defined by BTA-47853-no-rs at BTA2:63365256 in both the binary and recurrent LMM GWASs (Table II-S1).

Table II-4. Significant and suggestive SNPs detected from the linear mixed model genome-wide association analysis using quantitative phenotypes and their defined QTL.

SNP ID	BTA	SNP Position (bp)	MAF [a]	Effect Size (SE)	*p*		QTL Start Position (bp)	QTL End Position (bp)	QTL Size (kb)
BovineHD0200022555	2	78069923	0.231	0.127 (0.025)	3.14×10^{-7}	*	77930065	79925981	1995.9
BovineHD0200022557	2	78080217	0.231	0.127 (0.025)	3.14×10^{-7}	*	77930065	79925981	1995.9
Hapmap43777-BTA-115985	2	78080944	0.233	0.128 (0.025)	2.66×10^{-7}	*	77930065	79925981	1995.9
BovineHD0200022559	2	78092854	0.231	0.127 (0.025)	3.14×10^{-7}	*	77930065	79925981	1995.9
BovineHD0200022560	2	78100071	0.231	0.127 (0.025)	3.14×10^{-7}	*	77930065	79925981	1995.9
BovineHD0200022562	2	78110140	0.231	0.127 (0.025)	3.14×10^{-7}	*	77930065	79925981	1995.9
BovineHD0200022563	2	78111523	0.231	0.127 (0.025)	3.14×10^{-7}	*	77930065	79925981	1995.9
BovineHD0200022605	2	78307821	0.28	0.107 (0.023)	3.68×10^{-7}	†	77930065	79925981	1995.9
BovineHD0200022737	2	78767889	0.278	0.108 (0.023)	3.43×10^{-7}	†	77930065	79925981	1995.9

[a] MAF = minor allele frequency. * = genome-wide significant. † = genome-wide suggestive significance.

When the significant and suggestive SNPs from the LMM-quantitative analysis were used as predictors in Bayesian regression, MCMC sampling to fit the model was unable to efficiently explore the posterior likely because the phenotypes did not follow a normal distribution as expected by the model. This resulted in unreliable results and thereby prevented further analyses. The limited number of genes within the LMM-quantitative QTL on BTA 2:77930065–79925981 prevented the detection of overrepresented gene ontologies or pathways.

Random forest analysis using quantitative phenotypes revealed that, similar to the RF-binary rankings, farm was ranked as the most important explainer. The 15 important SNPs identified from the RF-quantitative analysis defined 13 QTL distinct from those defined in the LMM-quantitative analysis (Table II-5). The RF-quantitative QTL contained 124 genes. The 28 and 13 genes that were used in pathway analysis using KEGG and Reactome pathways did not find enriched pathways. The 37 genes used in gene ontology enrichment analysis did not have significantly overrepresented gene ontologies after multiple testing correction.

Although no pathways or gene ontologies were enriched from the RF-quantitative dataset, the important SNPs detected were nonetheless predictive of phenotype when used as predictors in Bayesian regression. MCMC sampling to fit the Bayesian model was able to explore the posterior sufficiently, resulting in convergence. Three of the important SNPs had effect sizes significantly greater than zero at 95% UI (Figure II-4, Table II-5). The LOO analysis indicated that the 15 SNPs were predictive of quantitative phenotype, as all the cows had Pareto k diagnostic values that were acceptable ($k \leq 0.7$). The PPC demonstrated that the simulated data followed a similar distribution to the original data, though the frequency of more extreme phenotypes was dampened (Figure II-S7).

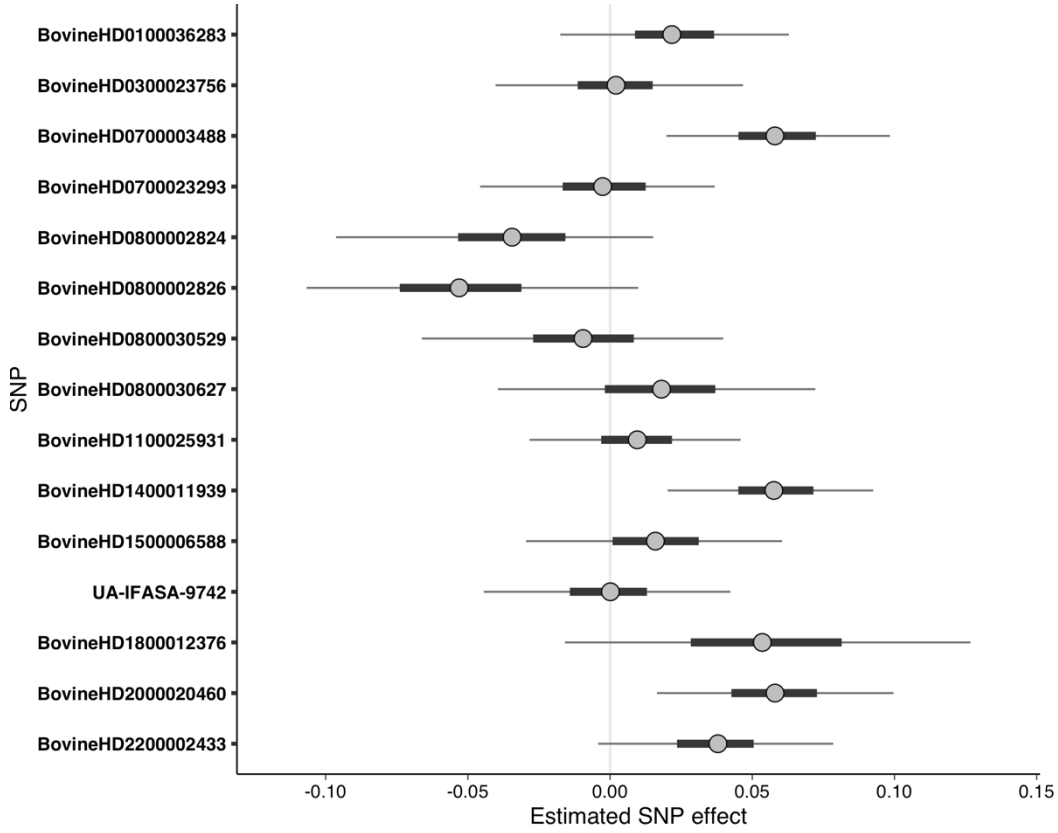


Figure II-4. Uncertainty interval (UI) plot for important SNPs from the random forest analysis using quantitative phenotypes. Dots represent the median of the SNP effect estimates from the Markov chain Monte Carlo draws, thick bars indicate the 50% UI, and the thin lines indicate the 95% UI. SNPs with 95% UI not overlapping zero were considered significant. Positive values of the predictor effect estimates indicate a phenotypic value for DD, whereas negative values indicate a phenotypic value for DD.

Table II-5. Important SNPs from random forest analysis using quantitative phenotypes and their defined QTL. Importance variables are expressed as % importance relative to farm (i.e., farm had 100% importance).

SNP ID	BTA	SNP Position (bp)	Minor/major allele	MAF[a]	SNP Importance (% Relative to Farm)	Significance in Bayesian Regression	QTL Start Position (bp)	QTL End Position (bp)	QTL Size (kb)
BovineHD0100036283	1	127408427	A/G	0.350	8.9	ns	127389567	127408427	18.9
BovineHD0300023756	3	82473975	A/G	0.391	8.5	ns	82468446	82480613	12.2
BovineHD0700003488	7	12238249	T/G	0.354	17.1	s	11979738	12261707	282.0
BovineHD0700023293	7	77533459	T/C	0.220	8.7	ns	77242189	78032023	789.8
BovineHD0800002826	8	8983282	C/T	0.463	8.7	ns	8671707	9806692	1135.0
BovineHD0800002824	8	8979816	G/A	0.373	10.6	ns	8671707	9806692	1135.0
BovineHD0800030529	8	100994105	C/T	0.402	10.5	ns	100412296	102353854	1941.6
BovineHD0800030627	8	101328029	G/A	0.350	9.5	ns	100412296	102353854	1941.6
BovineHD1100025931	11	89788438	C/A	0.387	9.9	ns	89375874	89788438	412.6
BovineHD1400011939	14	39785964	T/C	0.448	10.0	s	39785964	39818361	32.4
BovineHD1500006588	15	24668401	A/G	0.250	11.2	ns	24668401	24771237	102.8
UA-IFASA-9742	15	42081374	G/T	0.250	8.6	ns	42081374	42092689	11.3
BovineHD1800012376	18	41782168	C/T	0.081	9.6	ns	41753915	41863187	109.3
BovineHD2000020460	20	69870827	T/C	0.308	8.3	s	69696705	71850045	2153.3
BovineHD2200002433	22	8091674	T/C	0.205	9.1	ns	6375507	8317371	1941.9

[a] MAF = minor allele frequency. s = SNP effect estimated from Bayesian regression was significantly different from zero, as defined by the 95% uncertainty interval. ns = SNP effect estimated from Bayesian regression was not significantly different from zero, as defined by the 95% uncertainty interval.

4 DISCUSSION

The genetic component of DD susceptibility is highly complex and heterogeneous (van der Spek et al., 2015b; Biemans et al., 2018), as demonstrated by the numerous and varied QTL detected in previous studies (Scholey et al., 2013; Malchiodi et al., 2015b; van der Spek et al., 2015b; Naderi et al., 2018; Biemans et al., 2019; Sánchez-Molano et al., 2019). We sought to further identify the QTL contributing to DD susceptibility using a high-density SNP array and LMM and RF analytical approaches on well-phenotyped DD cases and controls. The LMM GWAS and RF analyses revealed suggestive, significant, and important SNPs that defined QTL regions in binary, quantitative, and recurrent DD phenotypes. The LMM GWAS using recurrent DD cases vs. controls indicated that the recurrent DD cases were contributing to the significance of association in the LMM-binary GWAS on BTA1 and BTA2, but not in the LMM-quantitative GWAS. Bayesian regression allowed for an intuitive estimate of SNP effects and the robust evaluation of model fit through the LOO and PPC analyses, providing additional distinctions of informative and noninformative SNPs among the top SNPs. QTL regions were explored for candidate genes if the QTL was defined by the top SNPs (i.e., significant or suggestive SNPs from LMM analyses or important SNPs in RF analyses) that were also significant in Bayesian regression or were top SNPs in both LMM and RF analyses. That is, nine QTL were investigated further (Table II-6).

Within these QTL regions, we identified likely candidate genes based on their relevance to DD etiology. DD is associated with Treponema bacteria invading the dermis and epidermis, likely through hair follicles, and results in a raised erosive lesion (Read and Walker, 1998; Evans et al., 2009). The infection elicits a strong initial activation of the innate immune response (Watts et al., 2018) that is then attenuated by the treponemes (Zuerner et al., 2007), leading to prolonged

inflammation and a delayed adaptive immune response (Refaai et al., 2013). Within the adaptive immune response, the antibody-mediated immune response is mainly responsible for defending the host against extracellular pathogens (Murphy and Weaver, 2016), such as treponemes. After the immune response, the skin then attempts to heal the wound, a process that is also impaired by treponemes (Zuerner et al., 2007). As such, changes in the sequence or expression of genes related to maintaining epidermal integrity, immune response, or wound repair could affect a cow's DD susceptibility and the persistence of a DD lesion. Previous work has indicated that genes related to these three functions were dysregulated in DD lesions (Zuerner et al., 2007; Scholey et al., 2013). Therefore, we considered candidate genes as those with associated phenotypes, as determined by MGI, that pertained to these functions and fell within the six QTL regions, resulting in six candidate genes: *CXCR4*, *MGAT5*, *CACNA1A*, *TERT*, *SLC9A3*, and *AHRR* (Table II-6, Table II-S2). All six candidate genes were related to immune function, and *TERT* was also associated with skin hyperplasia and wound healing (Table II-S2). Similarly, we defined functionally relevant gene ontologies and pathways as those related to these three functions. The QTL on BTA18 contained 16 zinc finger genes that were part of the herpes simplex virus 1 infection pathway, implying an immune function of these genes that could also play a role in DD infection (Table II-6).

A limitation of the study is the small sample size. Minimizing phenotypic variation and increasing sample size are both methods to improve the detection of small-effect SNPs, but often pursuing one of these approaches comes at the expense of the other—for example, in this study, large sample size. Our strict phenotypic criteria also caused the controls to be from only two dairies, which was partially accounted for in the LMM analyses by including a covariate term. While the uneven sampling of dairies can be problematic in frequentist methods such as LMM,

those issues were avoided in RF analyses and Bayesian estimation because these models account for parameters that did not exist (e.g., a control cow from Farm B or C). Furthermore, the SNPs that defined the QTL regions containing promising candidate genes were significant, suggestive, or important in the LMM and RF analyses, some of which also had nonzero effect sizes estimated from Bayesian regression despite the small sample size. For the quantitative phenotypes, a larger sample size might have more normally distributed phenotypes that the model expects, thereby improving the efficiency of MCMC sampling and more accurate SNP effect estimates. Although the sample size of this study was limited due to our intentionally reducing phenotypic variation, which may have prevented the detection of additional small SNP effects, the sample size was sufficient to very accurately predict the phenotype within the original population. Future replication studies are necessary to determine how well the SNP effects estimated in this study population can be extrapolated to larger populations in different geographical regions and other dairies.

Table II-6. Quantitative trait loci (QTL) defined by SNPs that were significant in at least two analyses: linear mixed model (LMM), random forest (RF), or Bayesian regression of top SNPs from linear mixed model (LMM-B) or random forest (RF-B) containing functionally relevant pathways or genes.

Phenotype	BTA	QTL Start Position (bp)	QTL End Position (bp)	QTL Size (kb)	Methodology Used in Defining the QTL	Relevant Pathways	Candidate Genes in QTL
Binary	1	125550933	125822143	271.2	LMM, LMM-B, RF, RF-B		
	2	60971364	63389576	2418.2	LMM, LMM-B		*CXCR*, *MGAT5*
	2	58016533	59967789	1951.3	RF, RF-B		
	2	65836042	65836042	-	LMM, LMM-B		
	3	119720909	119942789	221.9	RF, RF-B		
	18	47099464	47831459	732.0	RF, RF-B	Herpes simplex virus 1 infection	
Quantitative	7	11979738	12261707	282.0	RF, RF-B		*CACNA1A*
	14	39785964	39818361	32.4	RF, RF-B		
	20	69696705	71850045	2153.3	RF, RF-B		*TERT*, *SLC9A3*, *AHRR*

In addition to minimizing phenotyping variation, our GWAS used high-density SNP genotyping to increase the resolution of QTL detection. Previous studies (Scholey et al., 2012; Malchiodi et al., 2015b; van der Spek et al., 2015b; Biemans et al., 2018; Sánchez-Molano et al., 2019) had larger sample sizes than our study, achieved by using dairies across multiple geographic regions and various lower-density SNP panels (maximum 76 K SNPs). The lower-resolution SNP panels in those studies may have prevented the detection of smaller linkage disequilibrium blocks (<20 kb) in Holstein cattle (Pérez O'Brien et al., 2014) and contributed to the inconsistency of genomic regions detected. Although two previous studies found associated loci on BTA1, for one study the suggestive SNPs were in a different region (Biemans et al., 2019), while the other study did not provide SNP coordinates to permit comparisons (Malchiodi et al., 2015b). Similarly, other GWASs also detected the associated SNPs on BTA3 (Naderi et al., 2018; Sánchez-Molano et al., 2019) and BTA14 (Biemans et al., 2019), but in different regions. Other GWASs did not detect SNPs on the same chromosomes as our (Scholey et al., 2012) or did not detect any suggestive or significant SNPs (van der Spek et al., 2015b). The published GWASs with smaller sample sizes using the high-density SNP array were able to find SNPs associated for other traits in Holstein cattle, including digital cushion thickness (Stambuk et al., 2020), mastitis resistance (Kurz et al., 2019), and fat deposition (Lehner et al., 2018). Our study using tightly controlled cases and controls was the first to use high-density SNP genotypes in a GWAS for DD susceptibility for improved resolution and the first to find significant and suggestive SNPs on BTA2, 7, 18, and 20 in regions containing likely candidate genes or genes in relevant pathways. The multiplicity of associated chromosomal regions supports that the genetic component of DD susceptibility is heterogeneous and highly complex, such that different combinations of loci with small effects contribute to DD susceptibility, as suggested by previous

authors (van der Spek et al., 2015b; Biemans et al., 2018). The complex genetic architecture of DD susceptibility likely reflects multiple physiological systems (e.g., immune system, hair morphology, skin matrix remodeling) interacting in the etiology of DD.

The lack of congruence in the genomic regions associated with DD across published studies and the small effect sizes of those QTLs identified further supports that, in addition to many low-impact loci, non-genetic factors strongly influence DD susceptibility. The ranking of farm as the most important predictor in the RF analyses supports the concept that farm management (e.g., hoof trimming regiment, methods of preventing and treating DD) plays a significant role in reducing DD prevalence. Employing genetic selection in combination with environmental management will likely further reduce DD prevalence.

5 CONCLUSIONS

GWAS using LMM and RF approaches identified loci containing six genes on BTA1, 7, and 20 that regulate skin integrity, immune function, and wound repair: *CXCR4*, *MGAT5*, *CACNA1A*, *TERT*, *SLC9A3*, and *AHRR*. Bayesian estimation of SNP effects was used to additionally distinguish between informative and noninformative SNPs and indicated that the top SNPs from LMM-binary and RF-quantitative were collectively predictive of binary and quantitative phenotypes. Despite our identifying significant QTL, the absence of the congruency of associated SNPs in this study compared to other studies and the consistent ranking of the farm as the most important predictor in the RF analyses support the notion that DD susceptibility is heavily influenced by management, and the remaining genetic component is heterogeneous and highly complex. Thus, although farm management may be the most effective short-term method for reducing DD prevalence, combining genetic selection with management will likely be the most effective and sustainable long-term solution.

6 SUPPLEMENTARY MATERIALS

6.1 Supplementary tables

Supplementary tables are provided in the Excel workbook associated with this dissertation as well as online at http://www.mdpi.com/2076-2615/10/11/2009/s1

Table II-S1. Suggestive SNPs detected from the linear mixed model genome-wide association analysis using binary recurrent phenotypes and their defined QTL.

Table II-S2. Candidate genes found within the nine QTL defined by SNPs that were significant/important in at least two of the following analyses: linear mixed model, random forest, and/or Bayesian regression.

6.2 Supplementary figures

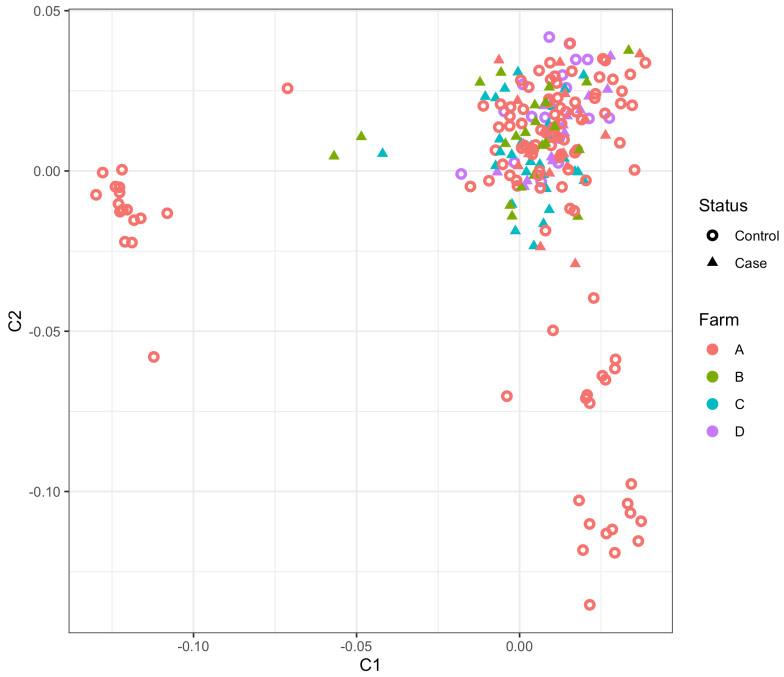


Figure II-S1. Multidimensional scaling plot depicting the first two dimensions. Each dot represents a cow, status is indicated by point shape, and farm is indicated by point color.

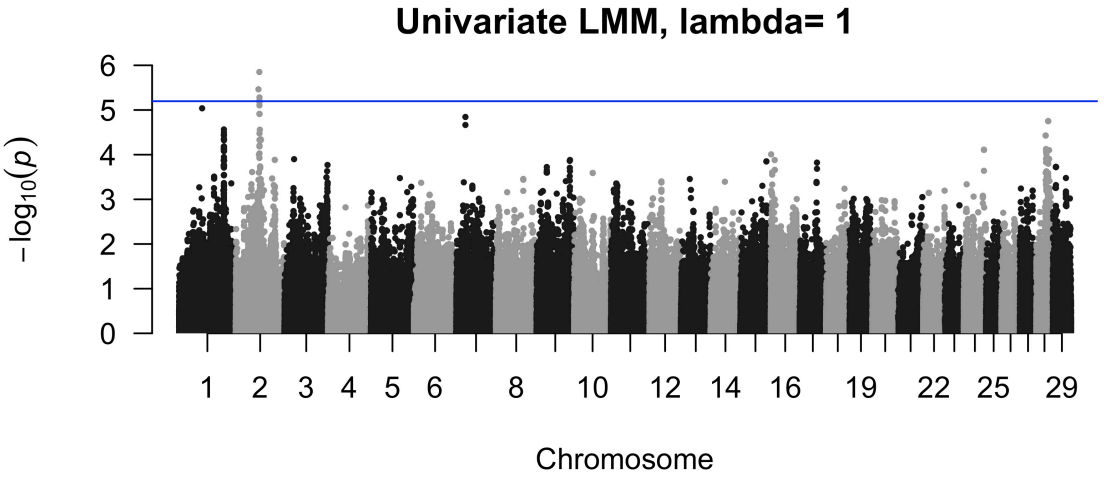


Figure II-S2. Manhattan plot for the linear mixed model genome-wide association analysis using binary phenotypes from recurrent cases vs. controls.

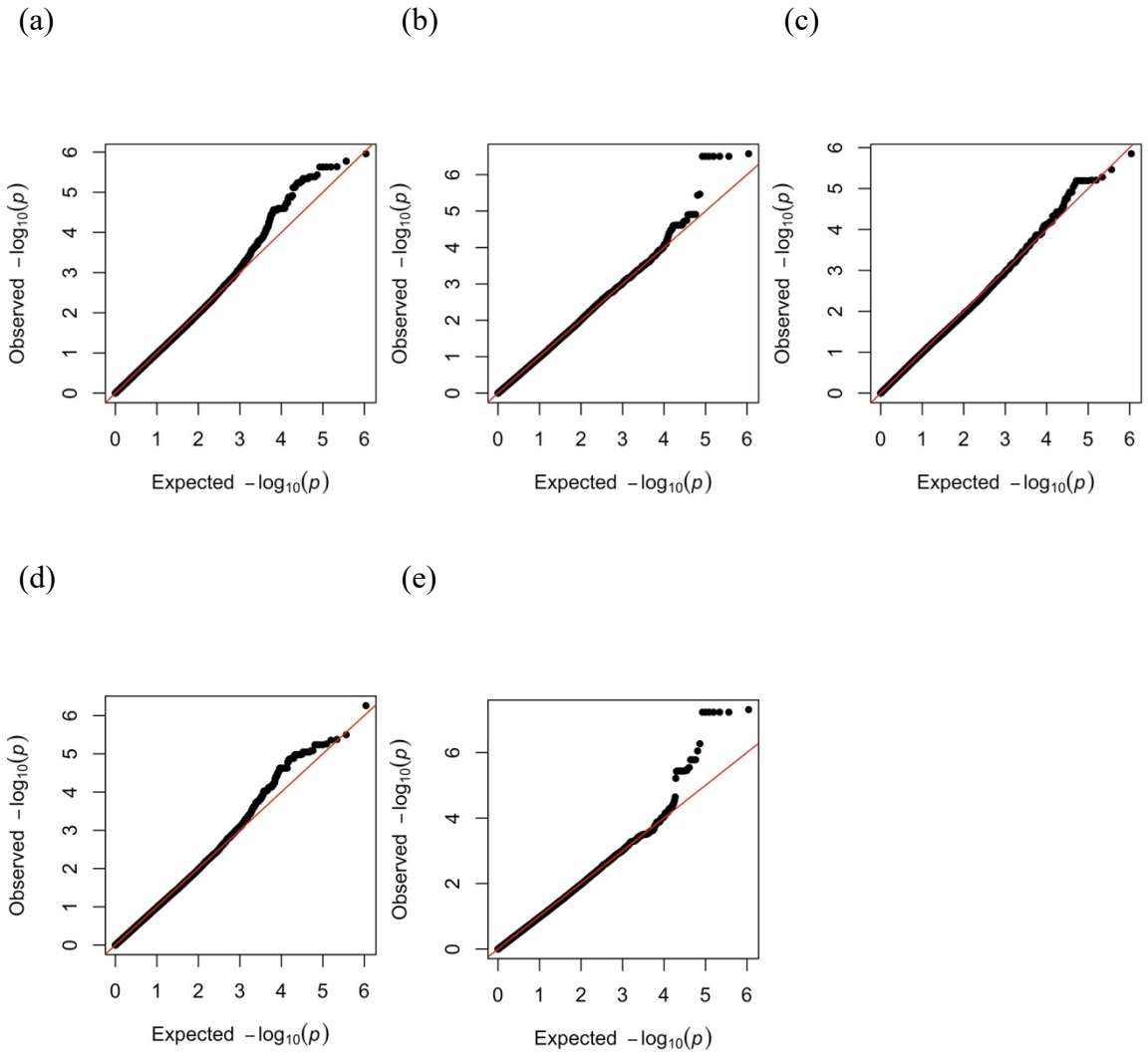
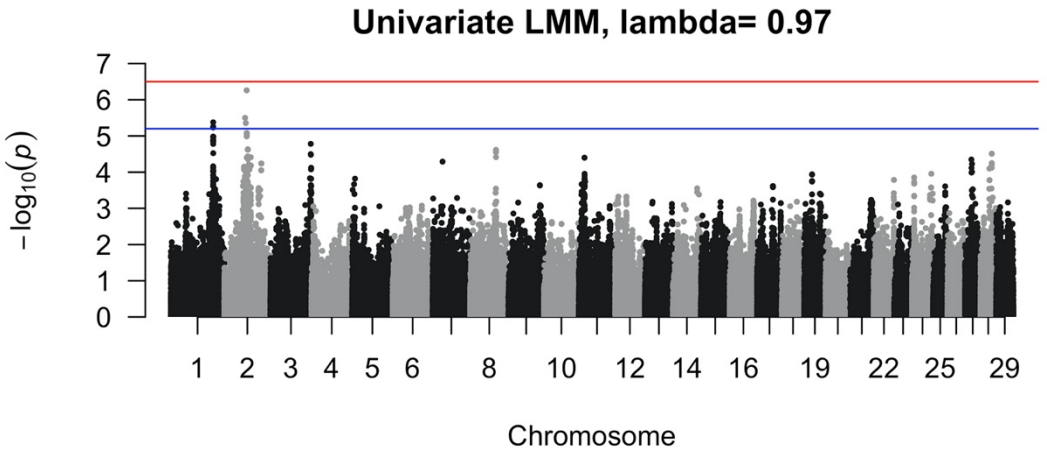


Figure II-S3. Quantile-quantile plots depicting observed and expected p-values from linear mixed model genome-wide association analyses using (a) binary, (b) quantitative, and (c) binary recurrent phenotypes in the full dataset of 261 cows; and (d) binary and (e) quantitative phenotypes in the subset of 188 cows after removing outlier cows. The red line indicates when observed and expected *p*-values are equivalent.

(a)



(b)

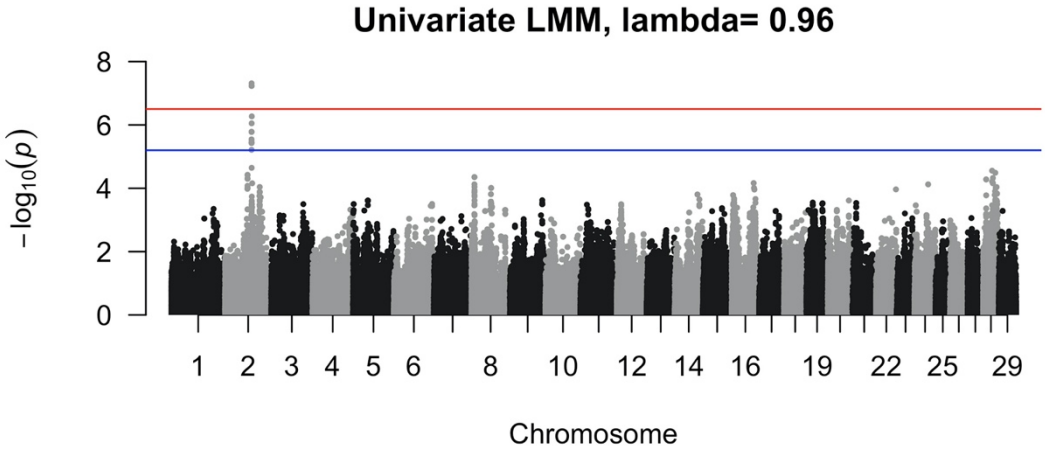


Figure II-S4. Manhattan plots from linear mixed model genome-wide association analyses excluding the outlier control cows using (a) binary phenotypes designating the presence of digital dermatitis (DD) lesions or absence of any lameness issues and (b) quantitative phenotypes calculated by dividing the number of DD episodes by the total number of hoof trimming records. The red line indicates the threshold for genome-wide significance (Bonferroni-corrected using the number of independent SNPs at $p < 0.05$), and the blue line indicates the threshold for suggestive significance (Bonferroni-corrected using the number of independent SNPs at $p < 1$). Genomic inflation factors (lambda) are indicated in figure titles.

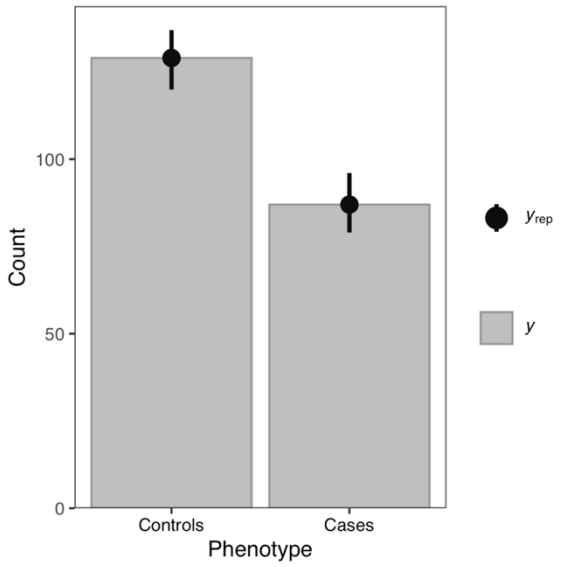


Figure II-S5. Posterior predictive check bar plot for Bayesian regression estimating effects of suggestive SNPs detected in the linear mixed model genome-wide association study using binary phenotypes. Gray bars represent the actual phenotypes and black dots with intervals represent the median and uncertainty intervals of the phenotypes of replicates (y_{rep}), which were simulated from estimated effects of predictors.

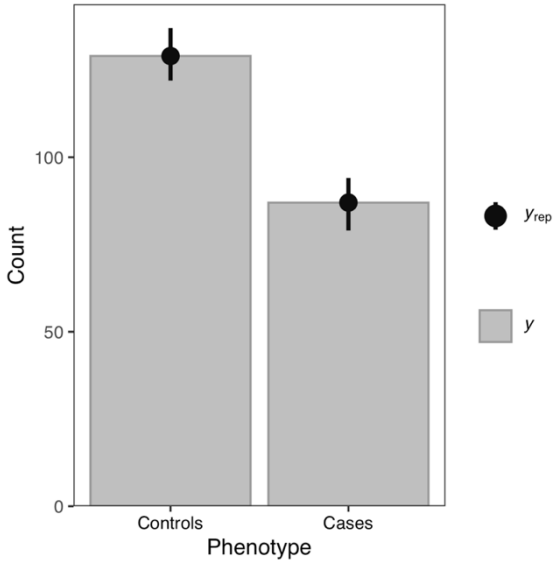


Figure II-S6. Posterior predictive check bar plot for Bayesian regression estimating effects of suggestive SNPs detected in the random forest using binary phenotypes. Gray bars represent the actual phenotypes and black dots with intervals represent the median and uncertainty intervals of the phenotypes of replicates (y_{rep}), which were simulated from estimated effects of predictors.

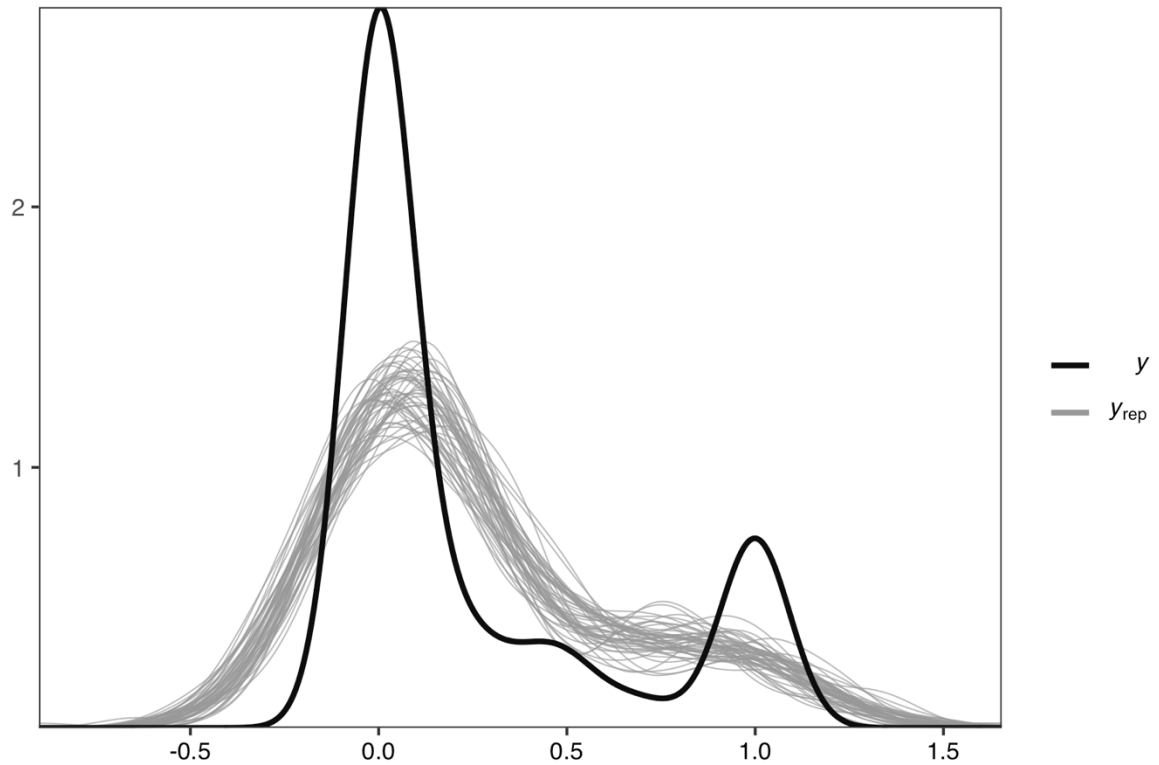


Figure II-S7. Posterior predictive check distribution plot for Bayesian regression estimating effects of suggestive SNPs detected in the random forest using quantitative phenotypes. The black line represents the actual phenotypic distribution and grey lines dots represent the phenotypic distribution of replicates (y_{rep}), which were simulated from estimated effects of predictors.

Chapter III. Genome-wide association studies reveal susceptibility loci for noninfectious claw lesions in Holstein dairy cattle

This chapter has been published in Frontiers in Genetics under the Specialty Section Livestock Genomics.

Citation: Lai, E., A.L. Danner, T.R. Famula, and A. Oberbauer. 2021. Genome-wide association studies reveal susceptibility loci for noninfectious claw lesions in Holstein dairy cattle. Front. Genet. 12:728. doi:10.3389/FGENE.2021.657375.

Abstract

Sole ulcers (SU) and white line disease (WLD) are two common noninfectious claw lesions that arise due to compromised horn production and are frequent causes of lameness in dairy cattle, imposing welfare and profitability concerns. Low to moderate heritability estimates of SU and WLD susceptibility indicate that genetic selection could reduce their prevalence. To identify susceptibility loci for SU, WLD, SU and/or WLD, and any type of noninfectious claw lesion, genome-wide association studies (GWAS) were performed using generalized linear mixed model (GLMM) regression, chunk-based association testing (CBAT), and a random forest (RF) approach. Cows from five commercial dairies in California were classified as controls having no lameness records and ≥ 6 years old (n = 102) or cases having SU (n = 152), WLD (n = 117), SU and/or WLD (SU+WLD, n = 198), or any type of noninfectious claw lesion (n = 217). Top SNPs were defined as those passing Bonferroni-corrected suggestive and significance thresholds in the GLMM analysis or those that a validated RF model considered important. Effects of top SNPs were quantified using Bayesian estimation. Linkage disequilibrium (LD) blocks defined by top SNPs were explored for candidate genes and previously identified, functionally relevant quantitative trait loci. The GLMM and CBAT approaches revealed the same regions of association on BTA8 for SU and BTA13 common to WLD, SU+WLD, and noninfectious claw lesions. These SNPs had effects significantly different from zero, and the LD blocks they defined explained a significant amount of phenotypic variance for each dataset (6.1 to 8.1%, $p < 0.05$), indicating the small but notable contribution of these regions to susceptibility. These regions contained candidate genes involved in wound healing, skin lesions, bone growth and mineralization, adipose tissue, and keratinization. The LD block defined by the most significant SNP on BTA8 for SU included a SNP previously associated with SU. The RF models

were overfitted, indicating that SNP effects were very small, thereby preventing meaningful interpretation of SNPs and any downstream analyses. These findings suggested that variants associated with various physiological systems may contribute to susceptibility for noninfectious claw lesions, demonstrating the complexity of genetic predisposition.

Keywords: dairy cattle, sole ulcer, pododermatitis circumscripta, white line disease, lameness, genome-wide association study, random forest, Bayesian regression

1 INTRODUCTION

Lameness, or abnormal gait and/or posture, is a pathognomonic sign that the affected cow is in pain and frequently reflects claw damage. Many claw conditions can cause lameness including injury, infectious foot lesions, and noninfectious claw lesions. The two most common noninfectious claw lesions causing lameness in dairy cattle are sole ulcers (SU), also known as pododermatitis circumscripta, and white line disease (WLD) (Green et al., 2002; Shearer and van Amstel, 2017). These lesions are not only a welfare issue, but are also associated with reduced milk production and decreased fertility (Green et al., 2002, 2010; Hernandez et al., 2005; Charfeddine and Pérez-Cabal, 2017). Consequently, SU and WLD represent a considerable financial burden with average costs associated with prevention, treatment, and losses from reduced productivity ranging from $181 (Dolecheck et al., 2019) to $258 (Cha et al., 2010) per case of SU and $155 for WLD (Dolecheck et al., 2019) (adjusted to 2020 US dollars). Production losses from extended calving interval, increased culling, and decreased milk production increase greenhouse gas emissions by 33 (3.6%) and 39 (4.3%) kg CO_2 equivalents per ton of fat-and-protein-corrected milk per case of SU and WLD respectively (Mostert et al., 2018). Reducing the prevalence of SU and WLD would alleviate these welfare, economic, and environmental concerns and thereby improve the sustainability of dairy production.

Both genetic and non-genetic factors contribute to susceptibility to SU and WLD, and prevention can be achieved through genetic means and herd management. Current prevention methods focus on management control primarily through regular claw trimming (Shearer and van Amstel, 2001) and providing rubber flooring in stalls and alleys (Vanegas et al., 2006; Fjeldaas et al., 2011; Eicher et al., 2013). Although dairies have implemented these prevention methods, SU and WLD remain prevalent worldwide with estimates ranging from 4.1 to 27.8% for SU and 2.0 to 11% for WLD in Holstein cattle depending on parity and housing style (Cramer et al., 2008; Bicalho et al., 2009; van der Linde et al., 2010; Oberbauer et al., 2013). Heritability estimates of susceptibility range from 0.01 to 0.3 for SU and 0.017 to 0.26 for WLD (Van der Waaij et al., 2005; van der Linde et al., 2010; Häggman and Juga, 2013; Oberbauer et al., 2013; van der Spek et al., 2013, 2015a; Malchiodi et al., 2015a), implying that these non-genetic means to reduce prevalence could be bolstered by genetic selection against susceptibility to these claw lesions. Although many genome-wide association studies (GWAS) have been performed to identify susceptibility loci, loci previously associated with SU and WLD are discordant (Malchiodi et al., 2015b; van der Spek et al., 2015b; Sánchez-Molano et al., 2019), and susceptibility to these claw lesions are believed to be complex traits governed by loci of small effect (van der Spek et al., 2015b). Some have postulated that selection against susceptibility to SU, WLD, and other noninfectious claw lesions could be achieved through indirect selection on body conformation traits or feet and leg traits (Van der Waaij et al., 2005; Häggman et al., 2013). However, the genetic correlation between conformation traits and susceptibility to noninfectious claw lesions appears to be low (Häggman and Juga, 2013; Malchiodi et al., 2015b; Ring et al., 2018), further accentuating the need to identify loci associated directly with susceptibility to noninfectious claw lesions. Thus, the objective of this

study was to identify genomic regions associated with susceptibility to SU, WLD, SU and/or WLD, and noninfectious claw lesions using well characterized herds under similar management practices: we hypothesized that we would identify small effect loci associated with predisposition to noninfectious claw lesions in addition to those already identified.

2 MATERIALS AND METHODS

All procedures were conducted in accordance with ethical standards set by the University of California, Davis and approved by the Institutional Animal Care and Use Committee (protocol #22099).

2.1 Phenotypic data

Dairies were selected to minimize environmental variation by including dairies in Central and Northern California using freestall housing, a flush system for waste removal, and diets balanced to meet the nutrition requirements from the National Research Council (NRC, 2001). Case/control phenotypes were defined using hoof trimming records. Hoof trimming records were generated by three hoof trimmers: one serviced Dairies A, B, and C; one serviced Dairy D; and the last trimmer serviced Dairy E. Hoof trimmer qualifications were described in a previous paper (Lai et al., 2020) and the three trimmers employed common criteria in defining the lesions. Hoof trimming regimens varied among dairies: cows were trimmed at the beginning of and mid-lactation, at dry off, and when lame (Dairy A); at dry off and when lame (Dairy B and C); only when lame (Dairy D); and at mid-lactation, at dry off, and when lame (Dairy E). The following foot disorders were documented in hoof trimming records: sole ulcer, hemorrhage, sole fracture, sole abscess, wall abscess, white line abscess (WLD), heel abscess, laminitis, foot wart, and foot rot. Cows were phenotyped as cases or controls based on whether they had or lacked records of claw lesions, respectively. Four case/control datasets were generated based on the type(s) of claw

lesions the cases had. For datasets 1 (SU) and 2 (WLD), cases were defined as cows with at least one record of SU or WLD, respectively. For dataset 3 (SU+WLD), cases included cows with either one or both of the claw lesions. Cases for dataset 4 (NICL) included cows with at least one of the following noninfectious claw lesions: SU, hemorrhage, sole fracture, sole abscess, wall abscess, WLD, heel abscess, and/or laminitis. Cows with no claw lesions and that were at least 6.0 years old were considered sound controls. The age restriction was imposed to avoid misphenotyping younger cows who had insufficient time to develop claw lesions. The same sound controls were used to compare against the cases in each of the four datasets.

2.2 Genotypes

Whole blood was collected from cows phenotyped as cases and controls. DNA was extracted from whole blood samples using the QIAGEN QIAamp DNA Blood Mini Kit (QIAGEN Inc., Valencia, CA) and quantified using the NanoDrop (ND-2000 v3.2.1) spectrophotometer (Thermo Scientific, Wilmington, DE). DNA samples were genotyped on the BovineHD BeadChip (777K SNPs, Illumina Inc., San Diego, CA) by GeneSeek (Lincoln, NE), and Illumina's GenCall algorithm was used to call genotypes. A portion of the controls used in this study were the same controls used in our previous study (Lai et al., 2020) for which raw and processed genotype data are publicly available at the NCBI Gene Expression Omnibus database (GEO series record GSE159157). Additional cows genotyped in this study are available in the GEO database (GEO series record GSE165945).

Genotypes were updated to the ARS-UCD1.2 assembly positions (Rosen et al., 2020) and quality filtered using PLINK 1.9 (Chang et al., 2015; Purcell and Chang, 2015) to remove from further analyses SNPs and cows with genotyping rates $< 95\%$, SNPs with significant deviation from Hardy-Weinberg equilibrium ($p < 1E\text{-}6$) to exclude systematic genotyping errors, and SNPs

with minor allele frequencies (MAF) < 5% to exclude rare variants. To visualize genetic similarity among the remaining cows, multidimensional scaling (MDS) analysis was performed, and the first two dimensions were plotted. Because the downstream programs for GWAS analysis [the generalized linear mixed model (GLMM) and random forest (RF)] require genotypes at each SNP, missing genotypes remaining after quality filtering were imputed using BEAGLE 5.1 (Browning et al., 2018) using the default parameters and an effective population size of 58 previously estimated for North American Holstein cattle (Makanjuola et al., 2020).

2.3 Generalized linear mixed model GWAS

Because disease phenotype was binary (cases and controls), the model used for association testing needed to reflect this binary outcome. Accordingly, logistic regression was used to model the binary outcome for the power analysis and for association testing. Power analysis was conducted using the genpwr R package (Moore et al., 2019) assuming an additive genetic effect and a sample size and case rate similar to the sample population (sample size = 275, case rate = 0.6). Given these parameters, the smallest effect SNP that the GWAS was expected to detect would have an odds ratio of at least 1.7 and a MAF of at least 0.34. For association testing, a genetic relatedness matrix (GRM) and farms were included as covariates in the model to account for population stratification and relatedness as well as the effect of farm, respectively. The probability of disease was defined as p_{ijk} for the *k*-th cow on the *i*-th farm identified in the *j*-th SNP genotype class and the logit of this probability, as $\theta_{ijk} = log\left[p_{ijk}/\left(1 - p_{ijk}\right)\right]$. The logit of the probability of disease was modeled as a function of recorded explanatory variables (e.g., farm, SNP genotype) along with a presumed quantitative genetic contribution for each SNP:

$$\theta_{ijk} = \mu + F_i + S_j + a_k$$

where μ was an unknown constant common to all cows, F_i contribution of *i*-th farm to the risk of disease, and S_j was the contribution of the *j*-th SNP genotype to the risk of disease. The additive genetic effects a_k were assumed to be drawn from the multivariate normal density N(0, $A\ \sigma_a^2$), with A as the standardized GRM among the animals in the dataset calculated in GEMMA (Zhou and Stephens, 2012), and σ_a^2 is the unknown additive genetic variance of disease risk. Model fitting and association testing via the score test (i.e. the Legrange multiplier test) were implemented with the generalized linear mixed model association test (GMMAT) R package (Chen et al., 2016).

The effective number of independent markers (M_e) was calculated as the number of SNPs remaining after linkage disequilibrium (LD) pruning using the Genetic Type I error calculator and used as the denominator for Bonferroni correction of association p-values (Li et al., 2012). Significant SNPs were defined as those with $p \leq 0.05/M_e$ and suggestive SNPs were defined as those with $p \leq 1/M_e$ (Lander and Kruglyak, 1995). Genomic inflation factors were calculated as the ratio of the median of observed and expected p values. Quantile-quantile plots (qqplots) and Manhattan plots were plotted using the R package *qqman* (R Development Core Team, 2010; Turner, 2014).

2.4 Chunk-based association testing

Chunk-based association testing (CBAT), also called set-based association testing, was performed to decrease multiple testing and in turn improve power of detecting associated regions in the small sample size. In contrast to gene-based association testing, which jointly tests variants within genes for association with phenotype [e.g. Xia et al. (2017)], chunk-based association testing analyzes consecutive windows of variants (i.e. chunks) across each chromosome without prior filtering. Accordingly, CBAT includes variants in noncoding regions containing regulatory

elements that could contribute to phenotypic variation in complex traits (Koufariotis et al., 2014, 2018). Quality-filtered SNPs were split into 100 kb chunks overlapping by 50 kb. Each chunk was LD-pruned to remove SNPs that were in strong LD ($R^2 > 0.98$) and then tested for association with phenotype by determining whether the phenotypic variance explained (PVE) by the chunk was significantly greater than zero. Specifically, association testing for each chunk was performed by calculating a GRM using the SNPs in the chunk and regressing the phenotype on the GRM. In addition to the chunk-based GRM, a thinned GRM (from genome-wide SNPs) and farms were included as covariates in the model to adjust for population stratification and differences among farms. The thinned GRM was calculated using genome-wide LD-pruned SNPs: SNPs within a window of 1 Mb and a $R^2 > 0.5$ were pruned out such that only SNPs in linkage equilibrium were used in the GRM calculation. For each chunk of SNPs, the following linear model was used to define the disease phenotype *y* for the *k*-th cow as a function of phenotypic contribution from the *j*-th chunk comprised of *m* SNPs and the *i*-th farm:

$$y_{ijk} = \mu + F_i + C_j + a_k + \varepsilon_{ijk}$$

where μ, F_i, and a_k were the same components outlined in the previous equation contributing to phenotype (coded as 0 for controls, 1 for cases), $C_j = \sum_{l=1}^{m} S_l$ was the contribution of the chunk to the phenotype in which S_l was the contribution of the *l*-th SNP in the chunk, and ε_{ijk} was the residual term. Estimates of PVE for each chunk were transformed to the underlying liability scale to adjust for ascertainment of cases using prevalence estimates from the literature: 4.08% for SU, 7.89% for WLD, and 0.10 for SU+WLD and 0.10 for NICL (DeFrain et al., 2013; Oberbauer et al., 2013). Calculating the thinned GRM, estimating PVE by each chunk, association testing with the likelihood ratio test, and p-value estimation via ten permutations for each chunk (Listgarten et al., 2013) were performed using the linkage

disequilibrium adjusted kinships (LDAK) program (Speed et al., 2012). For each dataset, significance thresholds were adjusted using Bonferroni correction: chunks with $p \leq 0.05/(\text{number of chunks})$ were defined as significant and chunks with $p \leq 1/(\text{number of chunks})$ were defined as suggestive (Lander and Kruglyak, 1995). Manhattan plots and qqplots were plotted using the R package *qqman* (R Development Core Team, 2010; Turner, 2014).

2.5 Random forest GWAS

A random forest fits a model that includes all SNPs and does not require an assumption about the mode of inheritance (e.g. additive, dominant, recessive), making RFs an appealing approach for complex traits such as susceptibility to claw lesions, in which the trait is highly polygenic and epistasis is present (Goldstein et al., 2010). Furthermore, RFs are insensitive to uneven sampling of cases and controls across different dairies, as RFs first build decision trees, then quantify importance values afterwards with data available in the trees.

Linkage-disequilibrium pruning and RF analyses were performed as previously detailed (Lai et al., 2020) for each of the four datasets. Briefly, LD-pruned genotypes and farms were used as predictors for the RF analyses performed using the *caret* R package (Kuhn, 2008; R Development Core Team, 2010). For each dataset, the population was randomly divided into a training (2/3 of the cows) and test (1/3 of the cows) population. Using the training population, the number of predictors considered at each node of each decision tree, *mtry*, was tuned using five values, 0.1*p*, 0.2*p*, 0.5*p*, 0.8*p*, and *p*, where *p* is the total number of predictors (Goldstein et al., 2010; Brieuc et al., 2018). The *mtry* resulting in the most accurate RF model was used for downstream analyses. The most important predictor was assigned a value of 100, and any other predictor's importance values was scaled accordingly (e.g., a predictor with an importance value of 50 is 50% as important as the most important predictor). Model validation was performed by

using the predictors and their importance values to predict case/control phenotype in the test population. To determine which SNPs were important and worthy of further investigation, a scree plot was plotted and the second order point of inflection was identified using the inflection R package (Christopoulos, 2016, 2017a) (i.e. the "elbow method"). Predictors with importance values equal to or greater than the second order point of inflection were defined as important SNPs and explored in downstream analyses if and only if the RF model was significantly more accurate at predicting phenotype in the test population than the noninformation rate (i.e., the frequency of the more common phenotype).

2.6 Defining associated regions

For each of the four datasets, top SNPs were defined as significant and suggestive SNPs from GLMM regression or important SNPs from a significantly predictive RF model. Boundaries of genomic regions of association were defined using SNPs in LD with top SNPs. Similar to the methodology of Richardson et al. (2016) and Twomey et al. (2019), positions of SNPs within 5 Mb and with $R^2 \geq 0.5$ of each top SNP were determined using non-pruned imputed genotypes, and the furthest SNP upstream and downstream in LD with the significant or suggestive SNP defined LD block boundaries. Overlapping LD blocks were combined. Using the same procedure outlined for CBAT, the PVE by the LD blocks defined from the GLMM and RF analyses was estimated and compared against the PVE by chunks of SNPs of the same size that overlapped by 50 kb from all chromosomes.

2.7 Bayesian estimation of SNP effects and assessing model fit

A Bayesian approach was used to test association of top SNPs identified in the GLMM and the RF with case/control phenotype for the four datasets. Bayesian methodology was selected because it allows multiple SNPs to be fitted jointly, recognizes that some SNPs are

correlated and most likely have small effects on susceptibility (van der Spek et al., 2015b), and can account for the uneven sampling of cases and controls from dairies. Additionally, the effect size estimates obtained from Bayesian estimation are directly interpretable, and Bayesian model evaluation is extremely thorough. Because highly correlated predictors complicate Bayesian regression, the significant and suggestive SNPs detected in the GLMM GWASs were LD-pruned ($R^2 > 0.9$) using PLINK 1.9 (Chang et al., 2015; Purcell and Chang, 2015) prior to estimating effects to keep the most significant SNP in each LD block for inclusion in the Bayesian model. Estimation of SNP effects was performed using a Bayesian logistic regression model as described in Lai et al. (2020). The important SNPs from the RF did not need to be LD-pruned, as SNPs were LD-pruned prior to RF analyses. Briefly, each set of top SNPs (i.e. LD-pruned suggestive/significant SNPs from the GLMM analyses and important SNPs from RF analyses) was used as predictors along with farm as a covariable in a Bayesian logistic regression model, and the model was fitted via sampling the posterior using the Hamiltonian Monte Carlo algorithm in the R package *rstanarm* (Gelman et al., 2020; Goodrich et al., 2020). The same population was used in the GLMM and RF GWAS as for SNP effect estimation, which could lead to inclusion of false positive associations in the Bayesian model. Thus, to discern whether the included SNPs were false positives, the fit of the Bayesian model using the estimated parameters was evaluated using leave-one-out (LOO) cross validation and posterior predictive checking (PPC) using the *loo* and *bayesplot* R packages (Vehtari et al., 2017, 2020; Gabry et al., 2019a). Bayesian estimation of SNP effects generated a distribution of where the true value of the SNP effect was, and this range was quantified in the 95% uncertainty intervals (UI), as opposed to a point estimate in frequentist methods. SNPs with 95% UIs that did not overlap zero were considered significantly associated with susceptibility to the respective claw lesion(s).

2.8 Functional annotation of associated regions

Genes and previously defined quantitative trait loci (QTL) falling within or overlapping with the associated LD blocks and chunks were obtained using FAANGMine using the genomic regions search function (FAANG, 2019) and the CattleQTLdb (Hu et al., 2019). RefSeq genes were extracted from the resulting gene list and used in pathway and gene ontology enrichment analysis in FAANGMine. Genes were searched in the Mouse Genome Informatics batch query database to find associated mammalian phenotypes (Smith and Eppig, 2009). Genes were also queried in the Cattle Gene Atlas (Fang et al., 2020) to determine in which tissues they were expressed.

3 RESULTS

3.1 Descriptive data

The percentage and count of cows with records of each claw lesion from each dairy are presented in Table III-2. Of the cows that had hoof trimming records from the five dairies, 5.6% and 12.0% had records of SU and WLD, respectively, similar to previous prevalence estimates (Cramer et al., 2008; Bicalho et al., 2009; van der Linde et al., 2010; Oberbauer et al., 2013). For cows that were genotyped, cases were sampled from all five dairies, whereas controls were sampled from Dairies A and D which had cows that met our strict soundness and age criteria for controls. The data set included 156 SU cases, 119 WLD cases, 203 SU+WLD cases (72 cows had both SU and WLD), 222 NICL cases, and 104 sound controls for a total of 287 cows (Table III-1). The average age of controls sampled was 8.7 years old (SD = 1.4), and when compared to the average age of onset of 4.2 (SD = 1.7) for SU and 4.5 (SD = 2.6) years for WLD, indicated that our age cutoff of 6.0 years old was sufficient to avoid misphenotyping control cows.

After quality filtering, ~556,000 SNPs for 152 SU cases, 117 WLD cases, 198 SU+WLD

cases (71 cases had both SU and WLD), 217 NICL cases, and 102 sound controls remained for MDS, GLMM, Genetic Type I error calculation, CBAT, and RF analyses. The MDS plot showed some population stratification, with a prominent center cluster and two other sparse clusters, though clustering was not by farm or case/control phenotype (

(A)

(B)

Figure III-1). Pairwise relationship coefficients calculated for the GRM ranged from -0.094 to 0.50, with negative values indicating the two cows were less related to each other than other random pairs of individuals. The distribution of pairwise relationship coefficients did not differ greatly between pairs of cows from the same farm and pairs from different farms (Figure III-S2). The Genetic Type I error calculator determined that the effective number of markers on autosomal chromosomes for Bonferroni correction was ~156,000 SNPs for the four datasets, yielding a significance threshold of $p = 3.2 \times 10^{-7}$ [6.5 on $-\log_{10}(p)$ scale] and a suggestive threshold of $p = 6.4 \times 10^{-6}$ [5.2 on -log10(p) scale]. The total number of 100 kb chunks used in CBAT was ~51,730 for the four datasets, yielding a significance threshold of $p = 9.7 \times 10^{-7}$ [6.0 on $-\log_{10}(p)$ scale] and a suggestive threshold of $p = 1.9 \times 10^{-5}$ [4.7 on $-\log_{10}(p)$ scale]. Linkage disequilibrium pruning at $R^2 > 0.90$ left 215343 to 218185 SNPs for RF analysis, depending on the dataset.

3.2 Generalized linear mixed model GWAS and chunk-based association testing

The GLMM analyses detected a region of association on BTA8 for SU and BTA13 for

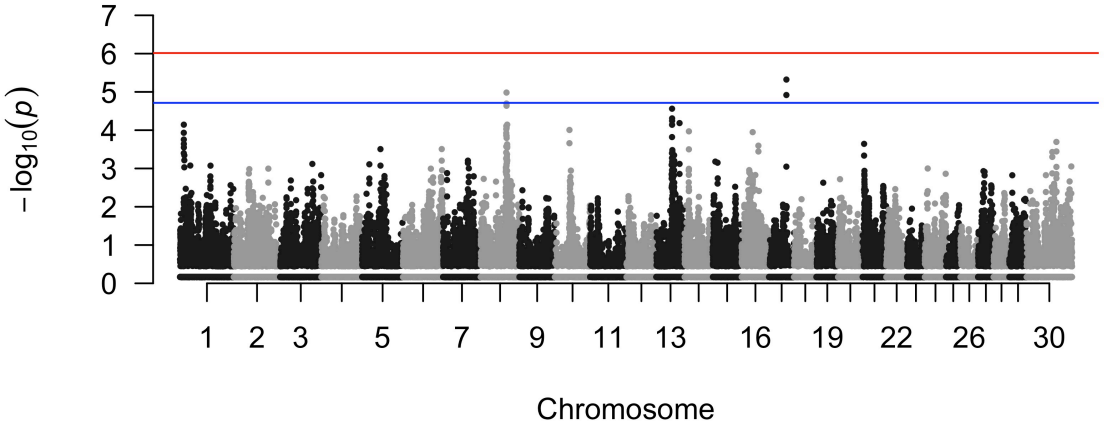
WLD, SU+WLD, and NICL while sufficiently accounting for population stratification and relatedness as indicated by the qqplots and the genomic inflation factors of 1.01, 1.02, 1.01, and 0.99 for SU, WLD, SU+WLD, and NICL, respectively (Figure III-S3). The CBAT using 100 kb overlapping chunks across the genome also properly accounted for population stratification and relatedness (qqplots in

(A) (B)

(C) (D)

Figure III-S5) and identified the same regions as the single-marker GLMM GWAS for each of the four datasets, providing further support for these regions (Table III-1, Manhattan plots in

(A)



(B)

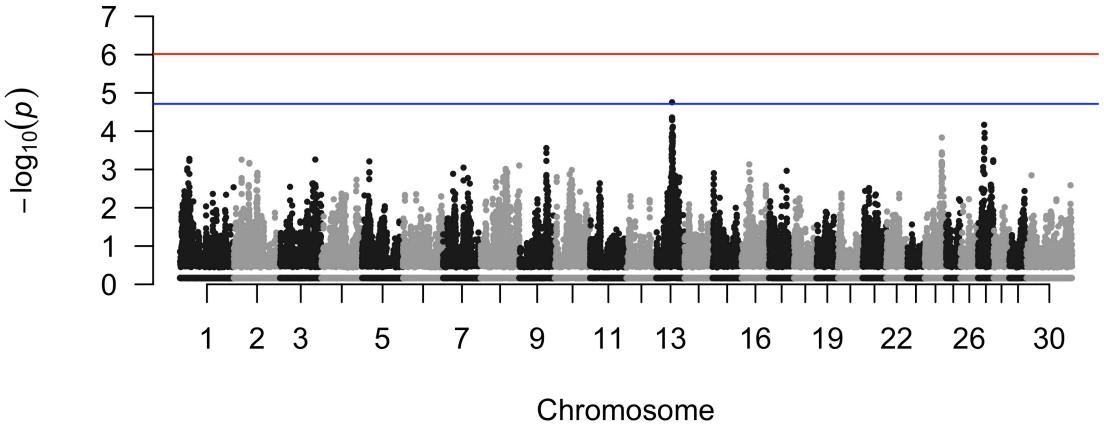
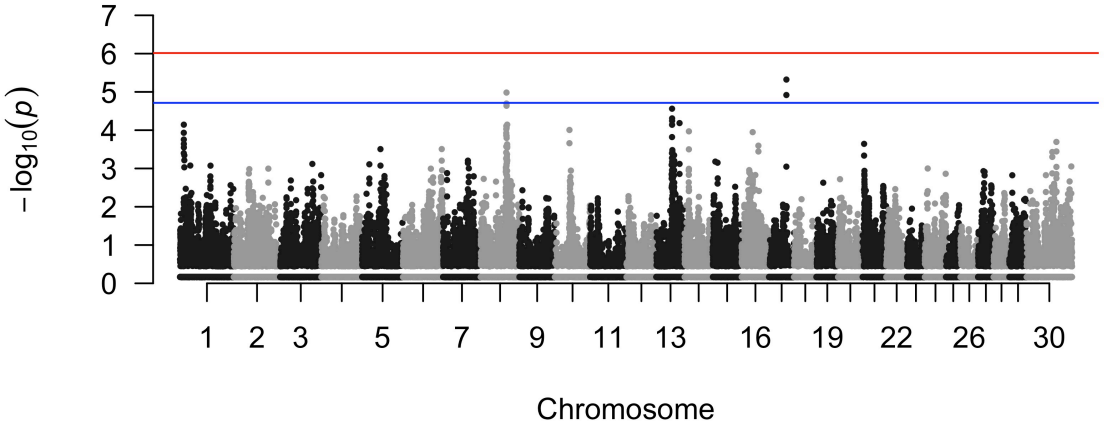


Figure III-S6). The SU CBAT also identified two suggestive chunks on BTA17 (Table III-1**Error! Reference source not found.**,

(A)



(B)

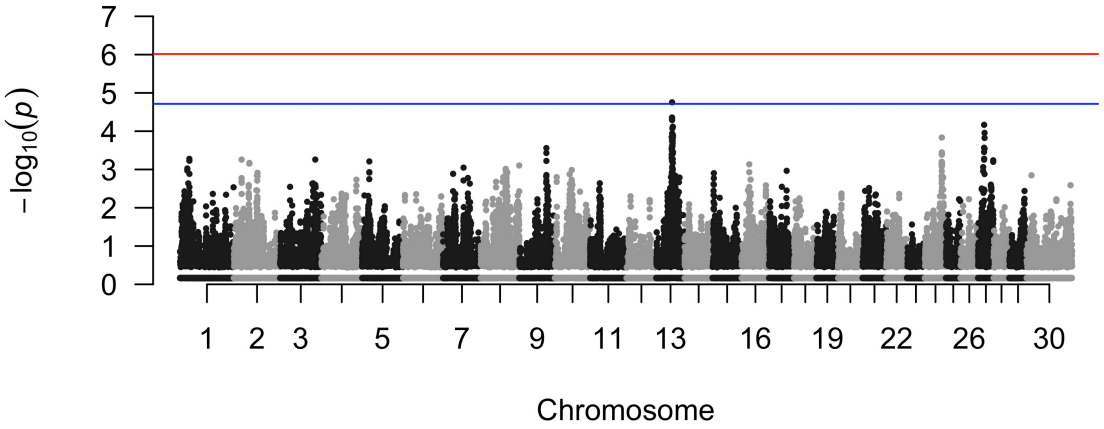
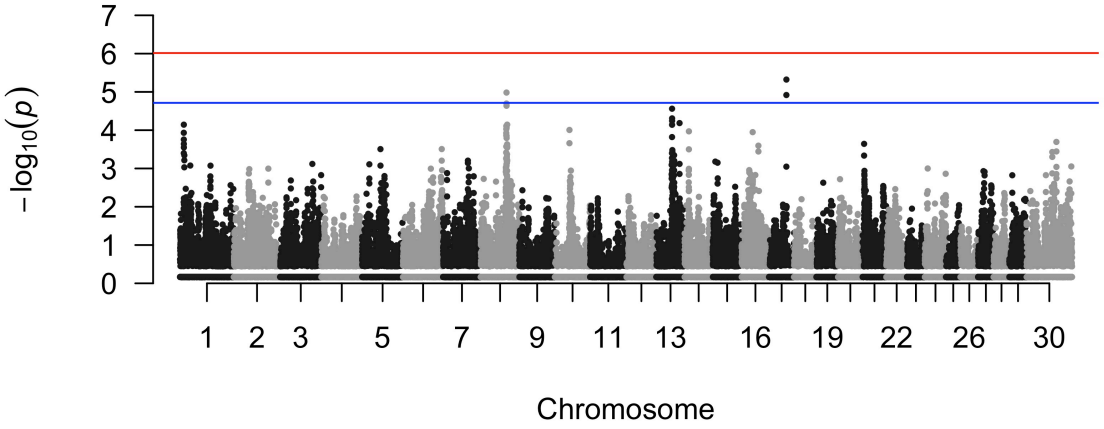


Figure III-S6A). For the NICL CBAT, the reduction in the number of tests performed allowed the chunk at BTA13:46,450,001-46,550,001 to reach genome-wide significance ($p = 6.9 \times 10^{-7}$, Table III-1,

(A)



(B)

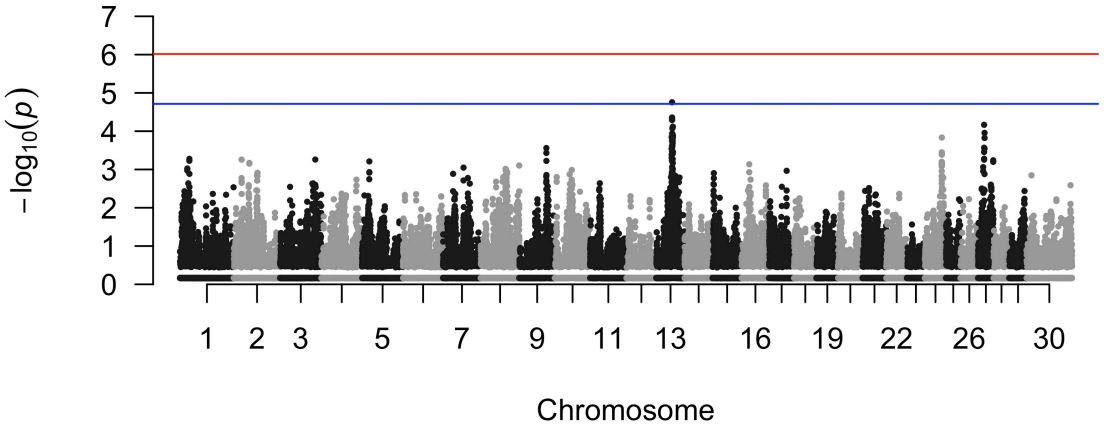


Figure III-S6D). This significant chunk contained the most significant SNP from the single-marker GLMM GWAS and three suggestive SNPs downstream.

Generalized linear mixed model association testing for SU susceptibility identified 12 suggestive SNPs on BTA8 falling in or directly upstream of the gene *DCAF12* (also known as DDB1 and CUL4 associated factor 12) (Table III-3, Table III-4). The 12 suggestive SNPs collectively defined a 3.2 Mb LD block at BTA8:74345807-77546693 (Table III-3 and

(A)

(B)

Figure III-1A) encompassing or overlapping with 60 genes: 52 protein-coding genes, four lncRNA genes, a tRNA gene, a miRNA gene, a snRNA gene, and a snoRNA gene. Because the 12 suggestive SNPs from the SU GLMM were in strong LD ($R^2 > 0.9$), the most significant SNP, BovineHD0800023021, was selected to represent this LD block in the Bayesian logistic regression model. The minor allele at BovineHD0800023021 (T) had an effect that was significantly less than zero at 95% UI (Table III-3and

(A)

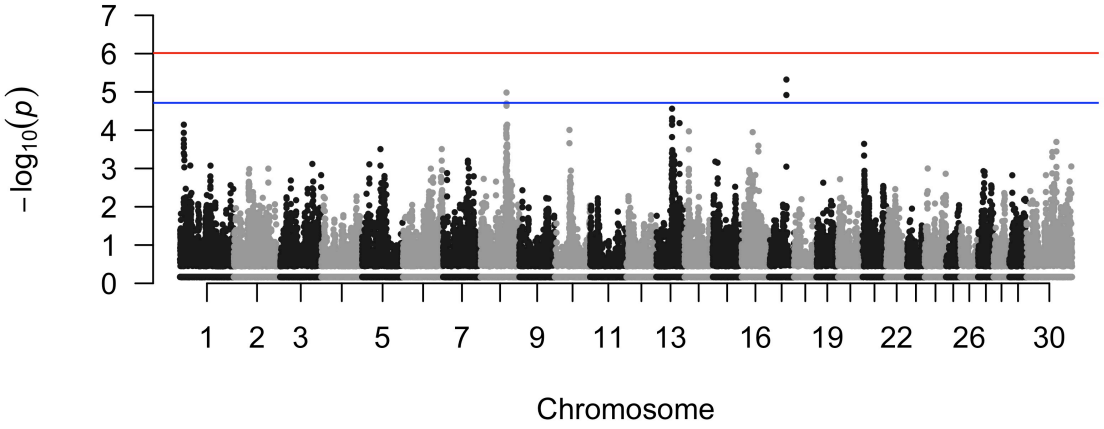
(B)

Figure III-2A), indicating that it was associated with reduced susceptibility to SU. The LOO analysis yielded acceptable Pareto k values ($k < 0.5$) for all cows, which indicated that the model was able to predict the phenotype of each cow with similar accuracy using genotypes at BovineHD0800023021 from all other cows. Goodness-of-fit assessment via PPC also showed that the distribution of phenotypes simulated using the estimated SNP effect closely aligned with that of the observed data (

(A) (B)

Figure III-S7A), further validating the fit of the model. In addition to identifying suggestive chunks in the same regions on BTA8, CBAT for SU detected two significant chunks on BTA17 (Figure III-S1,

(A)



(B)

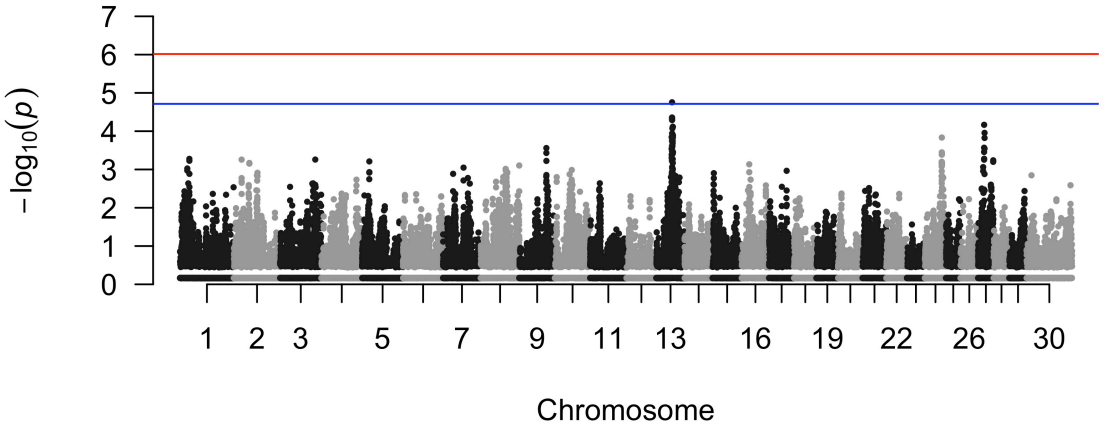


Figure III-S6A) that both fell within *TMEM12* (transmembrane protein 132B).

For WLD, the GLMM association testing found a single suggestive intergenic SNP at BTA13:46491619 (BovineHD1300013725,

(A)

(B)

Figure III-S4A), which was also the most significant SNP identified by the GLMM analysis for SU+WLD and NICL (Table III-3,

(A)

(B)

Figure III-S4B,

(A)

(B)

Figure III-1B). In addition to detecting BovineHD1300013725, the GLMM analyses for the SU+WLD and NICL datasets detected eight other suggestive SNPs in the same LD block as BovineHD1300013725 (Table III-3,

(A)

(B)

Figure III-S4). These nine suggestive SNPs detected in the SU+WLD GWAS were slightly more significant in the NICL GWAS and defined a 2.4 Mb LD block at BTA13:45283136-47676681 containing 27 genes: 16 protein coding genes, six lncRNA genes, two snRNA genes, two snoRNA genes, and one miRNA gene. For all four GLMM GWASs, the limited number of genes in LD blocks defined from suggestive SNPs precluded pathway and gene ontology analyses.

Given that the GLMM GWAS for SU+WLD and NICL identified nine suggestive SNPs in the same LD block ($R^2 > 0.5$) on BTA13 (

(A)

(B)

Figure III-1B,

(A)

(B)

Figure III-S4) and the top SNP is the same as that in the WLD GWAS, only the NICL Bayesian SNP effect estimation results are presented. Eight of these suggestive SNPs were in strong LD ($R^2 > 0.9$) whereas the remaining suggestive SNP (BTB-00525539) was in weaker LD with the others ($R^2 = 0.7$). Consequently, the most significant SNP in the LD block of eight SNPs (BovineHD1300013725) and BTB-00525539 were included in the Bayesian logistic regression model. The minor allele at BovineHD1300013725 representing the eight SNPs in strong LD had an effect that was significantly greater than zero at 95% UI (

(A)

(B)

Figure III-2B), indicating that the minor allele (C) was associated with increased susceptibility to NICL (Table III-3). In contrast, the effect of the minor allele at BTB-00525539

was not significantly different from zero (

(A)

(B)

Figure III-2B). Although the score variances of the suggestive SNPs were large (Table III-3), possibly due to the sample cohort, Bayesian estimation was less affected than GLMM regression by these limitations and indicated that the SNP effects were significant for SU and NICL (

(A)

(B)

Figure III-2). For the LOO analysis of the model, the acceptable Pareto k values ($k < 0.5$) from all cows demonstrated that the model including BovineHD1300013725 and BTB-00525539 was able to predict NICL phenotype of each cow based on genotypes at these two SNPs from the other cows with similar accuracy. The PPC-simulated data based on estimated SNP effects of these two SNPs that was similar to the observed data, indicating good model fit (

(A) (B)

Figure III-S7B).

To draw attention to the impactful SNPs shown in Table III-3 and the LD blocks they defined in Table III-4, the minor allele frequency at the most significant SNP for SU (BovineHD0800023021) in cases and controls was 0.253 and 0.476, respectively. The GLMM output score was negative and Bayesian estimation indicated a significant negative effect on susceptibility; that is, the minor allele was associated with reduced susceptibility. In contrast, the minor allele frequency at the most significant SNP for NICL (BovineHD1300013725) was higher in cases (0.459) than in controls (0.235), indicating that the minor allele was associated with higher susceptibility. Likewise, the GLMM score was positive, and Bayesian estimation of effect size resulted in a significant positive effect. Similar minor allele frequencies, scores, and significantly positive effect size estimates were observed at BovineHD1300013725 for WLD and SU+WLD. As seen in Table III-4, the LD blocks defined by suggestive SNPs had PVE between 0.06 and 0.08, depending on the dataset (SU, WLD, SU+WLD, or NICL), all of which were significantly greater than zero (permuted $p < 0.05$). In contrast, the genome-wide chunks with the same length as the LD blocks had an average PVE ~0.008, with PVE increasingly slightly with increasing chunk size, and average permuted p-values ~0.5.

3.3 Random forest GWAS

The RF models for all four datasets were not significantly more accurate at predicting phenotype in the test population compared to the non-information rate (i.e. the frequency of the more common phenotype), indicating that the RF models were overfitted (Brieuc et al., 2018) such that the SNPs that passed the significance threshold were likely random noise. Because importance values are assigned and the importance threshold defined after fitting the RF model,

some SNPs will always pass the importance threshold. Consequently, the value of these important SNPs and the likelihood that the important SNPs are truly trait linked must be gauged using model validation. In this case, the models were invalidated because of its poor phenotype prediction in the test population, indicating that the SNPs classified/categorized/determined to be important were unlikely associated with phenotype.

Additionally, the genomic regions identified by SNPs that passed the importance threshold did not overlap across the four datasets despite their shared etiology, nor with the genomic regions on BTA8 and BTA13 detected in the GLMM association analyses. Model overfitting combined with the lack of common genomic regions across the four datasets indicated that the RFs were unable to overcome the complex genetic architecture of noninfectious claw lesions and identify genomic regions of biological importance. Thus, downstream analyses to estimate SNP effects and conduct pathway and gene ontology analyses were not pursued.

4 DISCUSSION

Using GLMM regression, CBAT, and a RF approach to compare SNP genotypes of sound controls and various types of noninfectious claw lesion cases, we identified genomic regions associated with susceptibility to these claw lesions. Given the overlapping etiology of the noninfectious claw lesion in this study, we expected that association testing would detect genomic regions shared across some or all four datasets. Common genomic regions were identified from the GLMM and CBAT approaches, but not for the RF approach. Although RFs are a promising tool to identify loci associated with complex traits, the RF models in this study were overfitted, precluding meaningful interpretation of SNPs that passed the importance threshold. For GLMM testing and CBAT, the associated region detected on BTA8 for SU

appeared to be specific for SU because the analyses for the other claw lesions did not detect this region; a SNP in this region (ARS-BFGL-NGS-108587) has previously been associated with SU (van der Spek et al., 2015b). The SNP detected on BTA13 for WLD increased in significance as cows with SU and other noninfectious lesions were added to the GLMM GWAS and CBAT analysis, implying that these lesions shared a genetic component that was less prevalent in SU cases. LD blocks defined by top SNPs from the GLMM GWAS with nonzero effects from Bayesian estimation were explored further for candidate genes and previously defined QTL that were also functionally relevant to NICL etiology. Identification of promising candidate genes within associated regions may lend more confidence to those regions; however, genetic selection does not require candidate gene identification and instead uses markers that are associated with, but not necessarily causal for, the trait. Thus, candidate genes are presented below to postulate their contribution to etiology rather than to inform genetic selection.

Sole ulcers and WLD are thought to result from increased laxity of the suspensory system from collagen breakdown and a thinner digital cushion, allowing the distal phalanx to rotate and sink within the claw (Lischer et al., 2002; Bicalho et al., 2009; Newsome et al., 2017a, 2017b; Shearer and van Amstel, 2017; Stambuk et al., 2019). As the distal phalanx crushes the underlying corium, a hemorrhage develops at the pressure site and horn production through keratinization in the corium is disrupted, leading to horn thinning and eventually a hole in the horn through which the corium protrudes and develops into a SU (Greenough, 2007; Shearer et al., 2015). Similarly, WLD is thought to develop as a result of improper weight bearing and/or flooring causing defective horn production along the white line that is more prone to debris and bacteria infiltration, and when the bacteria reach the corium, an abscess forms (Shearer and van Amstel, 2017). It has been theorized that subclinical laminitis weakens the suspensory system

and thereby predisposes the cow to SU and WLD (Thoefner et al., 2004), though evidence supporting this theory is limited (Danscher et al., 2010). New bone development on the third phalanx (Rusterholz, 1920; Blowey et al., 2000; Lischer et al., 2002) is associated with increasing age (Tsuka et al., 2012; Newsome et al., 2016) and is thought to contribute to higher incidence of ulceration (Rusterholz, 1920; Tsuka et al., 2012). Because feet and leg conformation influence weight distribution within and between claws, feet and leg conformation traits are thought to be correlated with SU+WLD susceptibility, though stronger evidence is needed to support the low to moderate phenotypic (Capion et al., 2008; Pérez-Cabal and Charfeddine, 2016) and genetic (Chapinal et al., 2013) correlations that were previously observed. Based on the etiology of noninfectious claw lesions and the possible genetic correlation of susceptibility of these claw lesions with conformation traits, genes and QTL related to collagen, keratinization, bone growth, adipose, and feet and leg conformation were considered functionally relevant.

For SU, the suggestive SNPs fell in or near *DCAF12* (DDB1 and CUL4 associated factor 12), an evolutionarily conserved apoptosis regulation gene involved in DNA repair and protein degradation that is required for tissue homeostasis under stress conditions as demonstrated in *Drosophila* (Hwangbo et al., 2016). The metabolic stress associated with NICL could potentially disrupt regulation of *DCAF12* and contribute to aberrant tissue homeostasis within the claw. Within the LD block, *APTX, AQP7, B4GALT1, ENHO, GALT, GULO,* and *UBAP2* had functions involved in wound healing, skin lesions, bone growth and mineralization, adipose tissue, and keratin summarized in Table III-5. Notably, the LD block included a SNP that van der Spek et al. (2015b) had previously associated with SU susceptibility, ARS-BFGL-NGS-108587, supporting this SNP as a susceptibility locus for SU and the investigation into the region. No other previously defined QTL, physiologically relevant, or feet and leg conformation QTL were

identified in the LD block. The two suggestive chunks on BTA17 both fell in *TMEM132B* (transmembrane protein 132B, Table III-5), which in humans encodes a member of the TMEM132 family of evolutionarily ancient cell adhesion molecules that connect the extracellular medium with the intracellular skeleton (Sanchez-Pulido and Ponting, 2018).

For NICL, all nine suggestive SNPs fell directly upstream or within introns of *DIP2C* (disco-interacting protein 2 homolog C), which is hypothesized to play a role in transcription and methylation regulation. *DIP2C* has been shown to regulate DNA methylation and the epithelial-mesenchymal transition in human cell lines (Larsson et al., 2017), and mutations in *DIP2C* have been associated with skeletal dysplasia affecting bone and cartilage development in humans (Maddirevula et al., 2018). The LD block contained three additional candidate genes with functions related to adipose tissue, bone growth, and bone mineralization (Table III-5). The LD block on BTA13 did not overlap with previously defined QTL that were apparently related to NICL or feet and leg conformation traits. According to the Cattle Gene Atlas (Fang et al., 2020), some candidate genes were expressed ubiquitously *(DCAF12, APTX, GALT, UBAP2, DIP2C, PCNA*, and *WDR37*), and others were expressed more highly in specific tissues such as adipose, cardiovascular, bone marrow, central nervous system, mammary, liver, or immune tissues (*AQP7, B4GALT1, ENHO, GULO,* and *RASSF2*; Table III-5).

Prior GWAS studies of NICL, while having larger sample sizes, were sampled from larger geographical regions and used lower density SNP panels. An acknowledged limitation of this study is the small sample size. However, previous GWAS with smaller sample sizes using the high-density SNP array were able to detect associated loci in Holstein populations for digital cushion thickness (n = 502) (Stambuk et al., 2020) and left displaced abomasum (n = 406) (Lehner et al., 2018), implying that loci detection is possible despite smaller sample sizes. By

maintaining stringent phenotyping for sound controls, minimizing environmental and housing variability, and increasing SNP density, we aimed to optimize the ability to detect genomic variants at the expense of larger sample sizes. Additionally, the CBAT approach reduced the number of tests performed to increase power and found the same regions of association, providing further support for these regions. Because SU susceptibility is also affected by environmental management, including housing and nutrition, we sought to minimize environmental variability by sampling cows at dairies with similar nutrition and flooring, as the diets fed at the five dairies was similar and all dairies used a freestall flush barn system and rubber flooring in alleys.

Whereas previous published studies of noninfectious claw lesions have not used the high-density panel, our study with the 777K SNP panel allowed for higher resolution when defining LD blocks. Furthermore, RF analysis and Bayesian regression methods were implemented to perform joint association testing of multiple top SNPs while working around the uneven sampling of controls. The two published GWAS for SU susceptibility found associated SNPs on different chromosomes than those identified in this study, specifically on BTA 8, 10, 11, 18, and 22 using a linear animal model (van der Spek et al., 2015b) and on BTA12 and 25 using a linear mixed model (Sánchez-Molano et al., 2019). Other GWASs for traits related to SU+WLD included digital cushion thickness (Sánchez-Molano et al., 2019; Stambuk et al., 2020), sole hemorrhage susceptibility (van der Spek et al., 2015b; Sánchez-Molano et al., 2019), and laminitis susceptibility (Naderi et al., 2018), though SNPs detected in these studies were also on different chromosomes than those from this study.

Because noninfectious claw lesions have similar etiology, it has been postulated that pleiotropy may exist across the different noninfectious claw lesions and related traits. For

instance, estimates of genetic correlation between SU and WLD are significant, ranging from 0.41 to 0.60 depending on parity (van der Linde et al., 2010). However, past GWAS have not found associations on the same chromosomes among SU, WLD, digital cushion thickness, sole hemorrhage, or laminitis (van der Spek et al., 2015b; Naderi et al., 2018; Sánchez-Molano et al., 2019), or if SNPs from the same chromosome were detected, they were in different regions. Specifically, the only common chromosome among these three GWAS was BTA11: van der Spek et al. (2015b) found Hapmap38795-BTA-97039 for SU at BTA11:23302850 and Naderi et al. (2018) found BTB-00466773 for laminitis at BTA11:48309332 (SNP positions were updated to the ARS-UCD1.2 map). The QTL identified on BTA13 may thus represent a portion of the common genetic contribution to different types of noninfectious claw lesions.

5 CONCLUSIONS

Using logistic mixed model single-marker regression and CBAT, genomic regions associated with susceptibility were identified on BTA8 for SU and BTA13 for WLD, SU+WLD, and NICL. The associated regions on BTA8 and BTA13 contained candidate genes related to wound healing, skin lesions, bone growth and mineralization, adipose tissue, and keratin. The RF approach was unable to overcome the complexity of these lesion traits and reliably identify potential candidate QTL. Although these findings must be validated in larger populations in other geographical regions, the detection of a region associated with SU susceptibility that included a previously reported locus suggested that the study cohort was adequate to identify regions of susceptibility for NICL. Further exploration of these regions through targeted sequencing or RNA-seq in claw tissues with and without noninfectious claw lesions may uncover variants in genes or regulatory elements contributing to lameness. The multiplicity of associations detected in this and other studies demonstrated the complexity of the genetic

architecture underlying noninfectious claw lesion susceptibility.

6 TABLES

Table III-1.Distribution of cases for sole ulcers (SU), white line disease (WLD), SU+WLD, and noninfectious claw lesions (NICL) and sound controls after quality filtering across the five dairies

		Cases			
Farm	**Controls**	**SU**	**WLD**	**SU+WLD**	**NICL**
A	81	44	48	75	87
B	0	8	13	17	23
C	0	4	7	9	10
D	21	71	33	72	72
E	0	25	16	25	25
Total	102	152	117	198	217

Table III-2. Percent (and number) of cows with records of foot lesions across the five dairies

Farm	Percentage of cows with lesion, (no.)										Total no. cows with records
	Sole ulcer	White line disease	Foot wart	Wall abscess	Sole abscess	Sole fracture	Foot rot	Bruise	Heel abscess	Severe laminitis	
A	6.9 (467)	15.6 (1050)	5.6 (376)	2.2 (146)	4.7 (319)	1.9 (125)	0.9 (62)	NR	NR	NR	6734
B	1.2 (44)	2.6 (91)	8.5 (301)	0.2 (6)	0.6 (23)	1 (37)	0 (1)	NR	NR	NR	3549
C	1.3 (35)	2.1 (57)	8.8 (236)	0.5 (13)	0.3 (9)	0.4 (10)	0 (1)	NR	NR	NR	2676
D	5.5 (254)	12.8 (596)	5.8 (268)	NR	NR	NR	0.9 (42)	1.6 (73)	NR	NR	4658
E	11.1 (380)	21.4 (733)	17.9 (614)	NR	NR	NR	1.1 (39)	NR	16.3 (559)	8.3 (284)	3427
Total[1]	5.6 (1180)	12.0 (2527)	8.5 (1795)	1.3 (165)	2.7 (351)	1.3 (172)	0.7 (145)	1.6 (73)	16.3 (559)	8.3 (284)	21044

NR: not recorded

[1] Totals were calculated across dairies that had records of the lesion; dairies that did not record the lesion were excluded from calculation of totals

Table III-3. SNPs that were suggestive in the generalized linear mixed model association analysis and the linkage disequilibrium (LD) blocks they defined for sole ulcers (SU), white line disease (WLD), sole ulcers and/or white line disease (SU+WLD), and noninfectious claw lesions (NICL)

					Minor allele count		Minor allele frequency							
Dataset	BTA	SNP	SNP position (bp)	Minor/ Major allele	Cases	Controls	Cases	Controls	Score[1] (variance)	P	SNP significance in Bayesian estimation[2]	LD block start (bp)	LD block end (bp)	LD block length (kb)
SU	8	BovineHD0800023014	75,489,164	T/C	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023015	75,490,011	T/G	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	ARS-BFGL-NGS-112795	75,490,692	A/G	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023016	75,491,531	C/T	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023017	75,492,307	G/A	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023018	75,493,464	T/C	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023019	75,494,163	C/T	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023021	75,496,244	T/C	77	97	0.253	0.476	-20.4 (18.8)	2.66E-06	*	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023022	75,496,918	A/G	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023023	75,497,471	C/T	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023024	75,498,118	A/G	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
	8	BovineHD0800023025	75,501,482	T/C	75	94	0.247	0.461	-20 (18.1)	2.71E-06	-	74,345,807	77,546,693	3,200.9
WLD	13	BovineHD1300013725	46,491,619	C/T	106	48	0.453	0.235	19.9 (19.4)	6.13E-06	*	46,307,416	47,584,595	1,277.2
SU+WLD	13	BovineHD1300013725	46,491,619	C/T	183	48	0.462	0.235	25 (25.5)	7.03E-07	*	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013733	46,526,509	C/T	188	52	0.475	0.255	24.8 (25.6)	9.86E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013739	46,540,186	G/T	188	52	0.475	0.255	24.8 (25.6)	9.86E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013740	46,541,925	C/T	188	52	0.475	0.255	24.8 (25.6)	9.86E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013750	46,561,964	C/T	188	52	0.475	0.255	24.8 (25.6)	9.86E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013759	46,582,769	G/A	188	52	0.475	0.255	24.8 (25.6)	9.86E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013765	46,596,264	A/G	188	52	0.475	0.255	24.8 (25.6)	9.86E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013774	46,637,235	A/G	188	52	0.475	0.255	24.8 (25.6)	9.86E-07	-	45,283,136	47,676,681	2,393.5
	13	BTB-00525539	47,420,271	C/A	195	59	0.492	0.289	24.6 (27.8)	3.03E-06	ns	45,283,136	47,676,681	2,393.5
NICL	13	BovineHD1300013725	46,491,619	C/T	199	48	0.459	0.235	26.4 (27.2)	3.96E-07	*	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013733	46,526,509	C/T	204	52	0.470	0.255	26 (27.3)	6.68E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013739	46,540,186	G/T	204	52	0.470	0.255	26 (27.3)	6.68E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013740	46,541,925	C/T	204	52	0.470	0.255	26 (27.3)	6.68E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013750	46,561,964	C/T	204	52	0.470	0.255	26 (27.3)	6.68E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013759	46,582,769	G/A	204	52	0.470	0.255	26 (27.3)	6.68E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013765	46,596,264	A/G	204	52	0.470	0.255	26 (27.3)	6.68E-07	-	45,283,136	47,676,681	2,393.5
	13	BovineHD1300013774	46,637,235	A/G	204	52	0.470	0.255	26 (27.3)	6.68E-07	-	45,283,136	47,676,681	2,393.5
	13	BTB-00525539	47,420,271	C/A	213	59	0.491	0.289	25.8 (29.3)	1.79E-06	ns	45,283,136	47,676,681	2,393.5

[1] Scores are for the minor allele generated from generalized linear mixed model analysis. Negative scores indicate the minor allele is associated with reduced susceptibility and positive scores indicate the minor allele is associated with increased susceptibility.
[2] SNPs used in Bayesian model were either significantly different from zero at 95% uncertainty interval (*) or not significant (ns). Other SNPs were in LD with SNPs that were used in the model and were excluded from the model (-).

Table III-4. Proportion of phenotypic variance explained (PVE) by each linkage disequilibrium (LD) block defined from the generalized linear mixed model association analysis compared to the mean PVE of all chunks of genomic regions with the same length for sole ulcers (SU), white line disease (WLD), sole ulcers and/or white line disease (SU+WLD), and noninfectious claw lesions (NICL)

Dataset	BTA	LD block start (bp)	LD block end (bp)	LD block length (kb)	PVE (SD)	PVE P	Genome-wide mean chunk PVE (SE)	Genome-wide mean chunk PVE P (SE)
SU	8	74,345,807	77,546,693	3,200.9	0.081 (0.054)	3.93E-04	0.00809 (0.0004)	0.478 (0.006)
WLD	13	46,307,416	47,584,595	1,277.2	0.061 (0.047)	2.93E-05	0.00794 (0.0002)	0.485 (0.004)
SU+WLD	13	45,283,136	47,676,681	2,393.5	0.071 (0.050)	1.05E-06	0.00873 (0.0004)	0.482 (0.006)
NICL	13	45,283,136	47,676,681	2,393.5	0.074 (0.051)	5.79E-09	0.00828 (0.0003)	0.484 (0.005)

Table III-5. Candidate genes in linkage disequilibrium blocks defined by suggestive SNPs from generalized linear mixed model and chunk-based association testing for sole ulcers (SU), white line disease (WLD), sole ulcers and/or white line disease (SU+WLD), and noninfectious claw lesions (NICL) and the tissues in which they were expressed

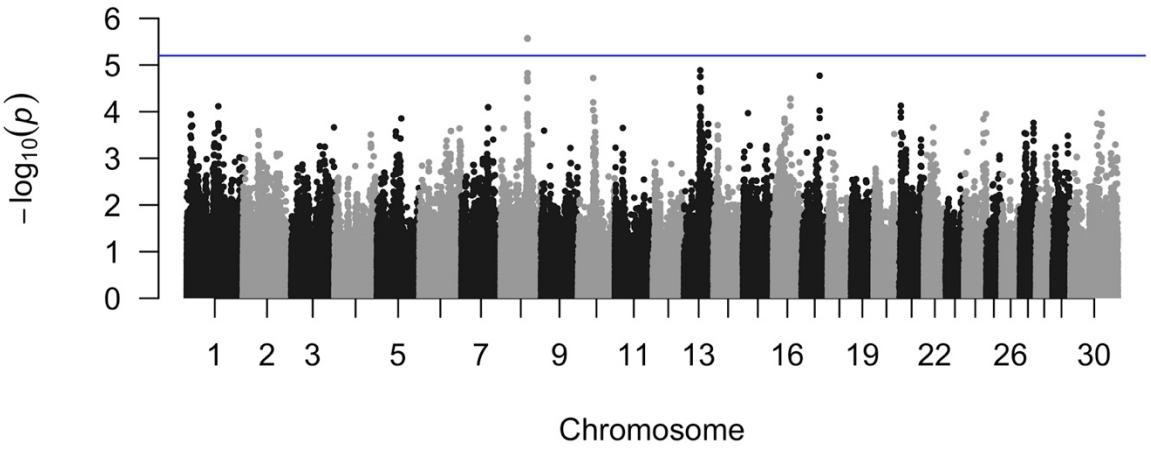
Claw lesion	Gene symbol	Gene description	Functional relevance	RNA tissue specificity
SU	*DCAF12*	DDB1 (damage specific binding protein) and CUL4 (cullin 4) associated factor 12	Regulates apoptosis required for tissue homeostasis under stress conditions (Hwangbo et al., 2016)	Ubiquitous
	APTX	Aprataxin	Decreased bone mineral content (MGI) Increased total body fat amount (MGI)	Ubiquitous
	AQP7	Aquaporin 7	Abnormal white adipose tissue physiology (MGI) Increased fat cell size (MGI)	Adipose, cardiovascular, and bone marrow
	B4GALT1	Beta-1,4-Galactosyltransferase 1	Decreased subcutaneous adipose tissue amount (MGI) Delayed wound healing (MGI) Hyperkeratosis (MGI) Skin lesions (MGI) Thin skin (MGI)	Mammary gland
	ENHO	Energy homeostasis associated	Increased body fat mass (MGI) Increased percent body fat/body weight (MGI)	Central nervous system
	GALT	Galactose-1-Phosphate Uridylyltransferase	Decreased subcutaneous adipose tissue amount (MGI) Delayed wound healing (MGI) Hyperkeratosis (MGI) Skin lesions (MGI)	Ubiquitous

Claw lesion	Gene symbol	Gene description	Functional relevance	RNA tissue specificity
			Thin skin (MGI)	
	GULO	Gulonolactone (L-) oxidase	Abnormal bone mineralization (MGI) Abnormal long bone epiphyseal plate morphology (MGI) Abnormal trabecular bone morphology (MGI) Decreased bone mineral density (MGI) Decreased compact bone thickness (MGI)	Liver
	TMEM132B	Transmembrane protein 132B	Cell adhesion molecule that connects the extracellular medium with the intracellular skeleton (Sanchez-Pulido and Ponting, 2018)	Central nervous system, testes
	UBAP2	Ubiquitin-associated protein 2	Abnormal adipose tissue amount (MGI)	Ubiquitous
WLD, SU+WLD, NICL	*DIP2C*	Disco-interacting protein 2 homolog C	Regulates DNA methylation and the epithelial-mesenchymal transition in human cell lines (Larsson et al., 2017) Mutations associated with skeletal dysplasia (Maddirevula et al., 2018)	Ubiquitous
	PCNA	Proliferating cell nuclear antigen	Abnormal adipose tissue development (MGI) Decreased percent body fat/body weight (MGI) Decreased white fat cell numberv	Ubiquitous
	RASSF2	Ras association (RalGDS/AF-6) domain family member 2	Abnormal bone mineralization (MGI) Abnormal trabecular bone morphology (MGI) Decreased bone marrow cell number (MGI) Decreased bone mass (MGI) Decreased bone mineral density (MGI) Decreased bone trabecula number (MGI) Decreased trabecular bone thickness (MGI) Decreased trabecular bone volume (MGI)	White blood cells and immune tissues
	WDR37	WD repeat domain 37	Increased bone mineral content (MGI)	Ubiquitous

MGI: Mammalian phenotype associated with gene from Mammalian Genome Informatics batch query

7 FIGURES

(A)



(B)

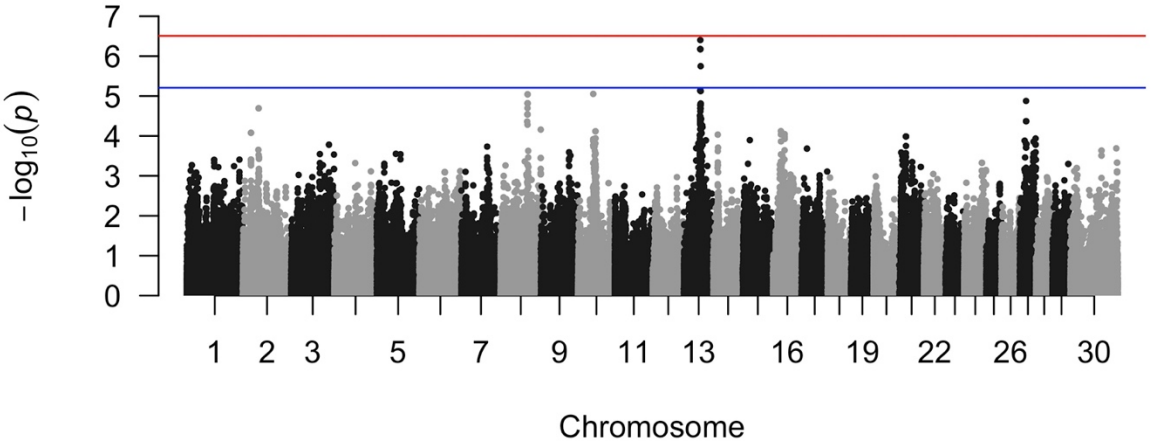
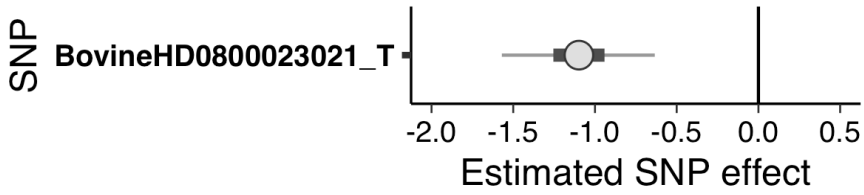


Figure III-1.Manhattan plots from the generalized linear mixed model regression association analyses for (A) sole ulcers, and (B) noninfectious claw lesion susceptibility. The blue line indicates the threshold of genome-wide suggestive significance, and the red line indicates the threshold of genome-wide significance.

(A)



(B)

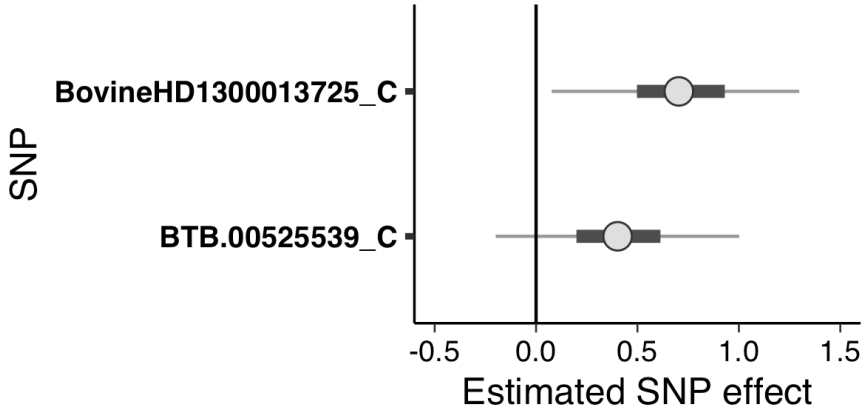


Figure III-2. Bayesian uncertainty interval (UI) plots depicting the estimated SNP effects of the suggestive SNPs detected in the generalized linear mixed model regression analysis for (A) sole ulcers, and (B) noninfectious claw lesion susceptibility. Dots indicate the median of the SNP effect, thick black bars indicate the 50% UI, and thin lines indicate the 95% UI of the effect size distribution. The letters following SNP names indicate the minor allele for which the effect was calculated. Positive values indicate the minor allele of the SNP increases susceptibility, and negative values indicate the minor allele of the SNP decreases susceptibility.

Abbreviations: *Bos taurus* autosome (BTA), chunk-based association testing (CBAT), generalized linear mixed model (GLMM), genetic relatedness matrix (GRM), genome-wide association study (GWAS), linkage disequilibrium (LD), minor allele frequency (MAF), noninfectious claw lesions (NICL), proportion of phenotypic variance explained (PVE), random forest (RF), sole ulcers (SU), white line disease (WLD)

8 SUPPLEMENTARY MATERIAL

The following supplementary materials are also available online at

https://www.frontiersin.org/articles/10.3389/fgene.2021.657375/full#supplementary-material

8.1 Supplementary tables

Table III-S1. Chunks of SNPs that were significant or suggestive in chunk-based association testing and the proportion of phenotypic variance they explained (PVE) for sole ulcers (SU), white line disease (WLD), sole ulcers and/or white line disease (SU+WLD), and noninfectious claw lesions (NICL)

Dataset	BTA	Chunk start (bp)	Chunk end (bp)	Number of SNPs in chunk	PVE (SD)	P	
SU	8	75450001	75550001	13	0.053 (0.044)	1.04E-05	*
	17	50500001	50600001	21	0.113 (0.063)	1.21E-05	*
	17	50550001	50650001	17	0.150 (0.077)	4.76E-06	*
WLD	13	46450001	46550001	5	0.061 (0.058)	1.76E-05	*
SU+WLD	13	46400001	46500001	5	0.093 (0.080)	3.86E-06	*
	13	46450001	46550001	5	0.059 (0.054)	9.73E-07	*
	13	46500001	46600001	7	0.059 (0.051)	2.51E-06	*
	13	46550001	46650001	8	0.061 (0.051)	2.99E-06	*
	13	46600001	46700001	6	0.071 (0.058)	3.18E-06	*
NICL	13	46400001	46500001	5	0.095 (0.081)	2.59E-06	*
	13	46450001	46550001	5	0.060 (0.054)	6.91E-07	**
	13	46500001	46600001	7	0.059 (0.051)	1.83E-06	*
	13	46550001	46650001	8	0.059 (0.049)	2.36E-06	*
	13	46600001	46700001	6	0.069 (0.056)	2.70E-06	*

*PVE by chunk reached genome-wide suggestive significance
**PVE by chunk reached genome-wide significance

8.2 Supplementary figures

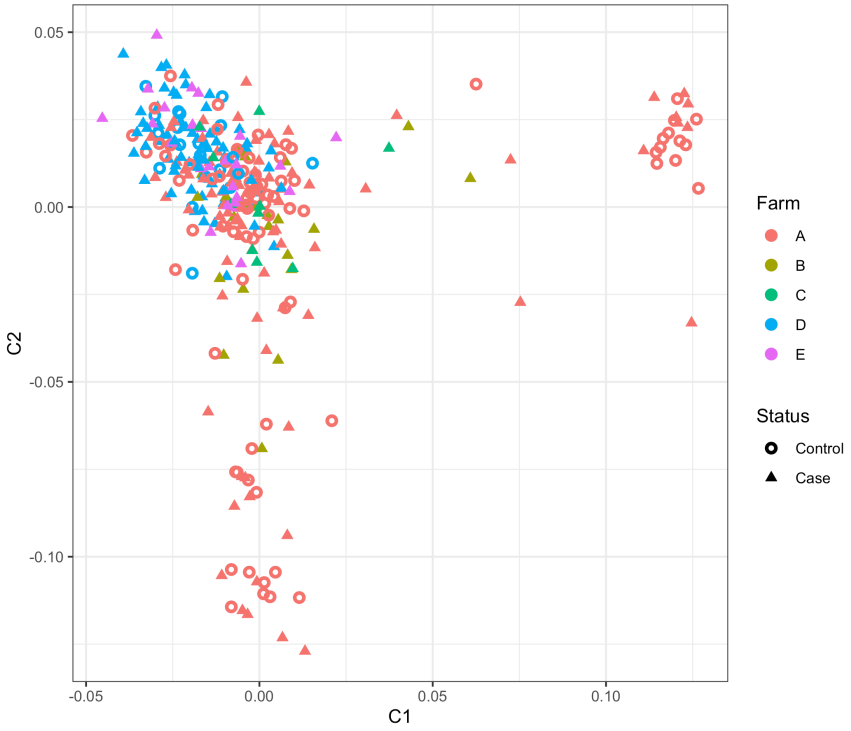


Figure III-S1. Multidimensional scaling plot showing the first two dimensions for the 217 noninfectious claw lesion cases and 102 sound controls from five dairies used in the genome-wide association analyses.

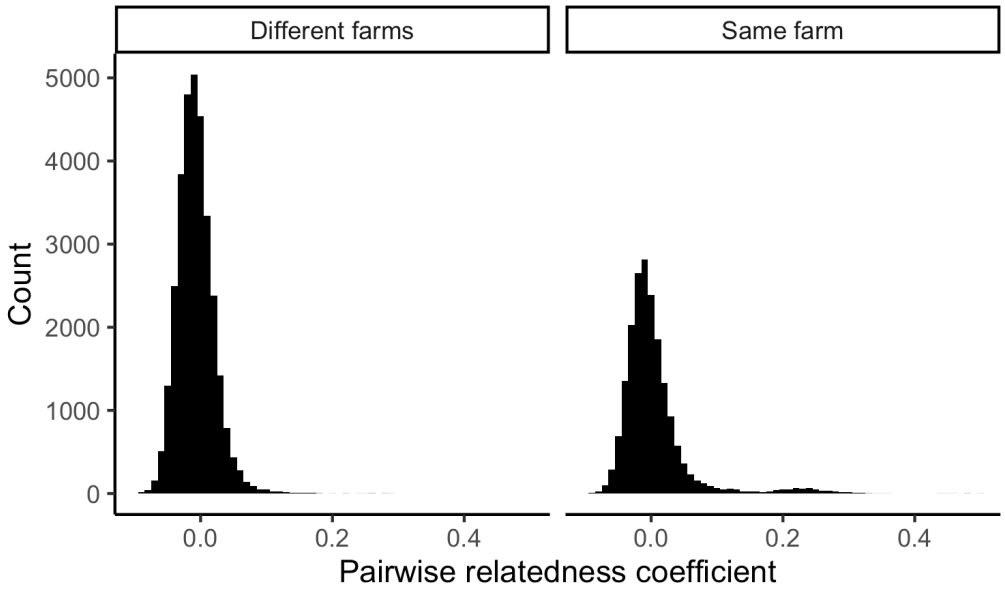
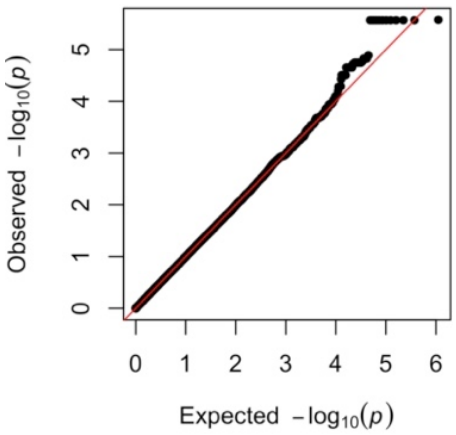
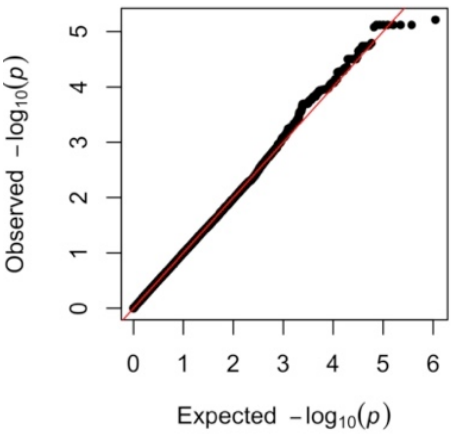


Figure III-S2. Histogram illustrating the distribution of pairwise relatedness coefficients in cows from the same farm and different farms.

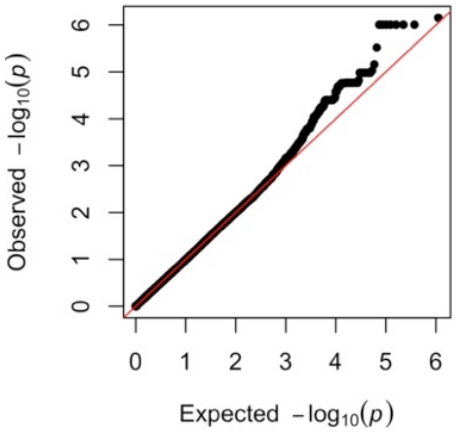
(A)



(B)



(C)



(D)

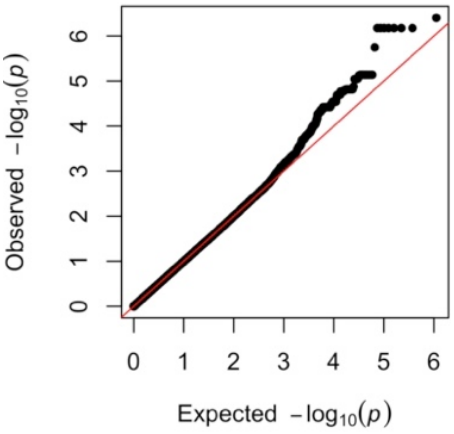
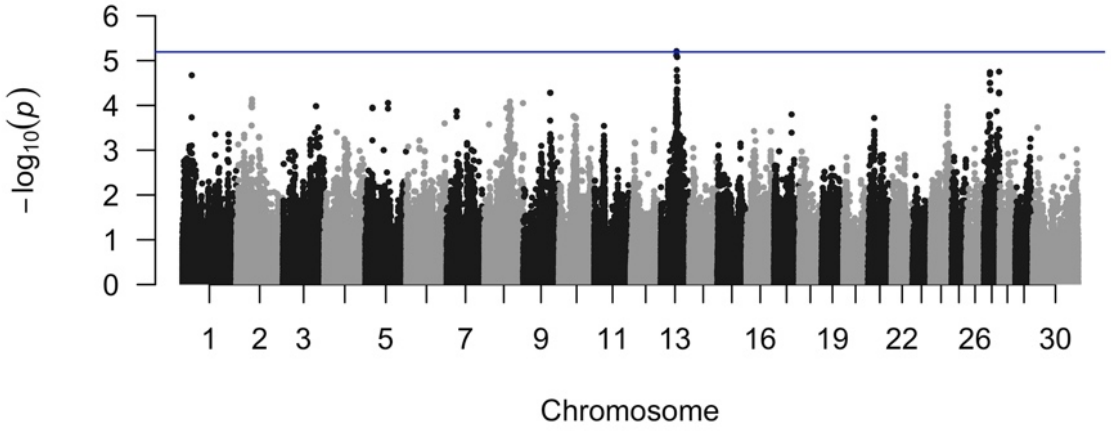


Figure III-S3. Quantile-quantile plots showing the observed vs. expected p-values outputted from the generalized linear mixed model regression analysis for (A) sole ulcers, (B) white line disease, (C) sole ulcers and white line disease, and (D) noninfectious claw lesions. The red line indicates where observed and expected p-values are equivalent.

(A)



(B)

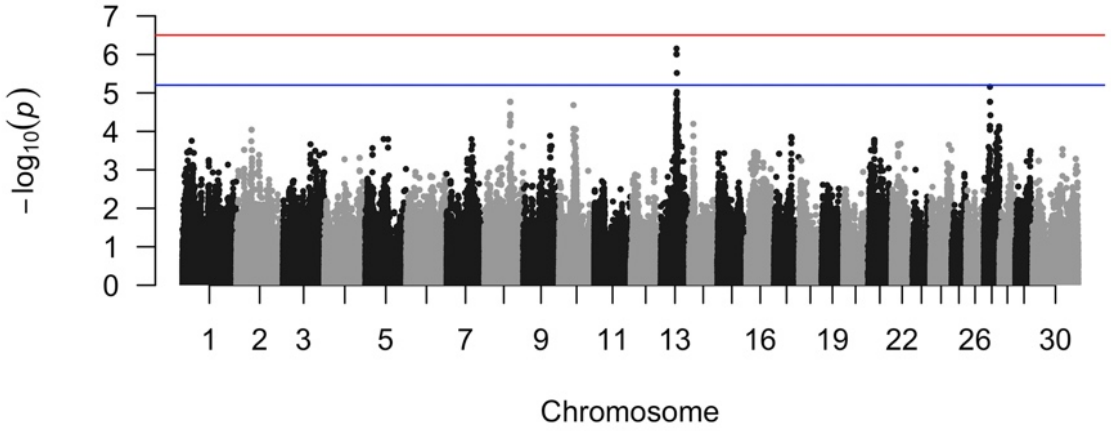
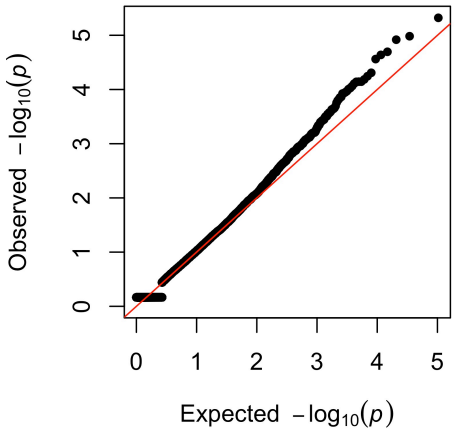
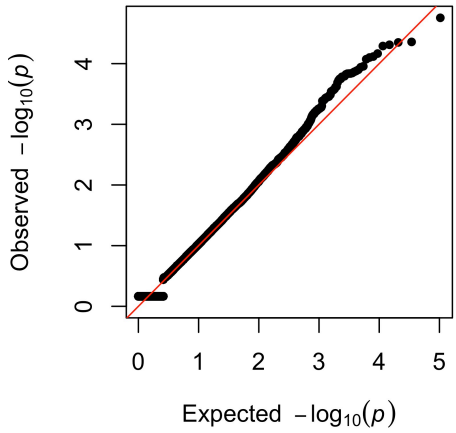


Figure III-S4. Manhattan plots from the generalized linear mixed model regression association analyses for (A) white line disease and (B) sole ulcers and white line disease. The blue line indicates the threshold of genome-wide suggestive significance and the red line indicates the threshold of genome-wide significance.

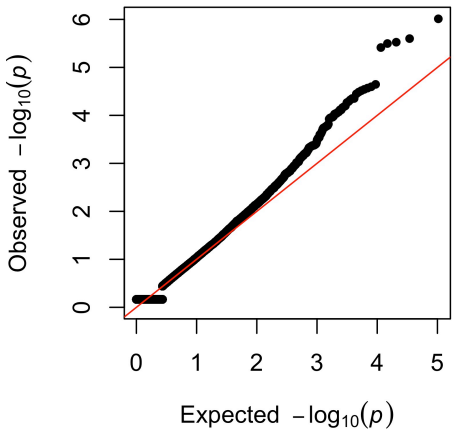
(A)



(B)



(C)



(D)

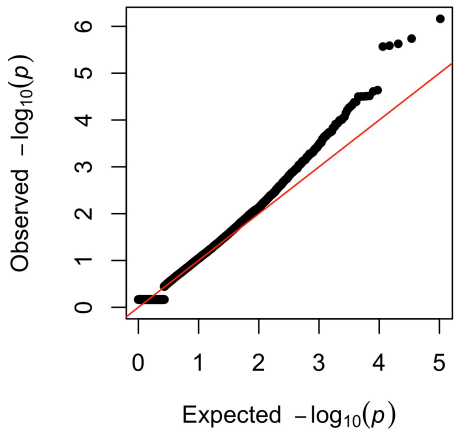
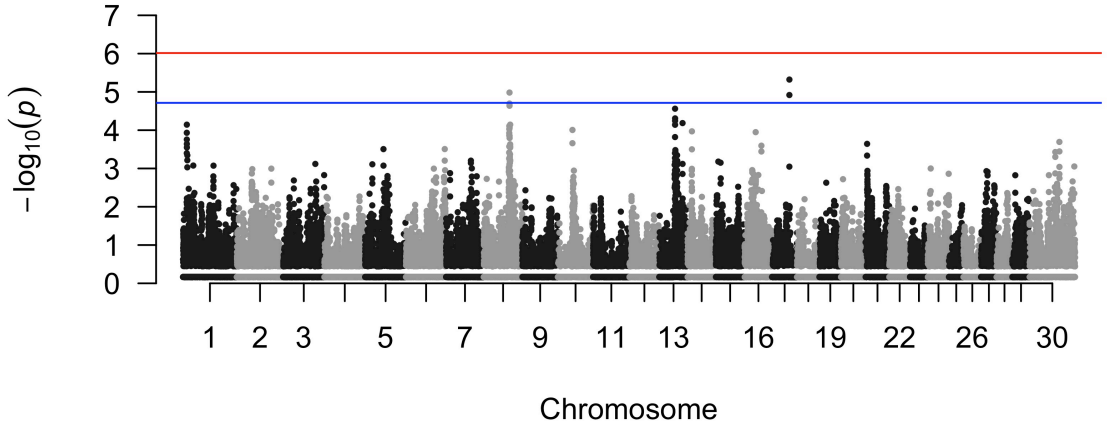


Figure III-S5. Quantile-quantile plots showing the observed vs. expected p-values outputted from chunk-based association testing for (A) sole ulcers, (B) white line disease, (C) sole ulcers and white line disease, and (D) noninfectious claw lesions. The red line indicates where observed and expected p-values are equivalent.

(A)



(B)

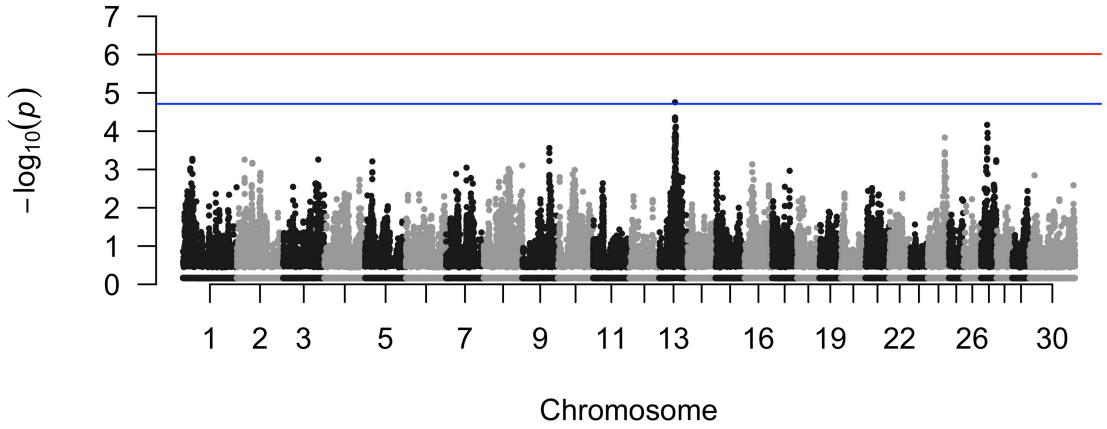


Figure III-S6. Manhattan plots from chunk-based association testing using 100 kb chunks spanning the genome for (A) sole ulcers, (B) white line disease, (C) sole ulcers and white line disease, and (D) noninfectious claw lesions. The blue line indicates the threshold of genome-wide suggestive significance and the red line indicates the threshold of genome-wide significance.

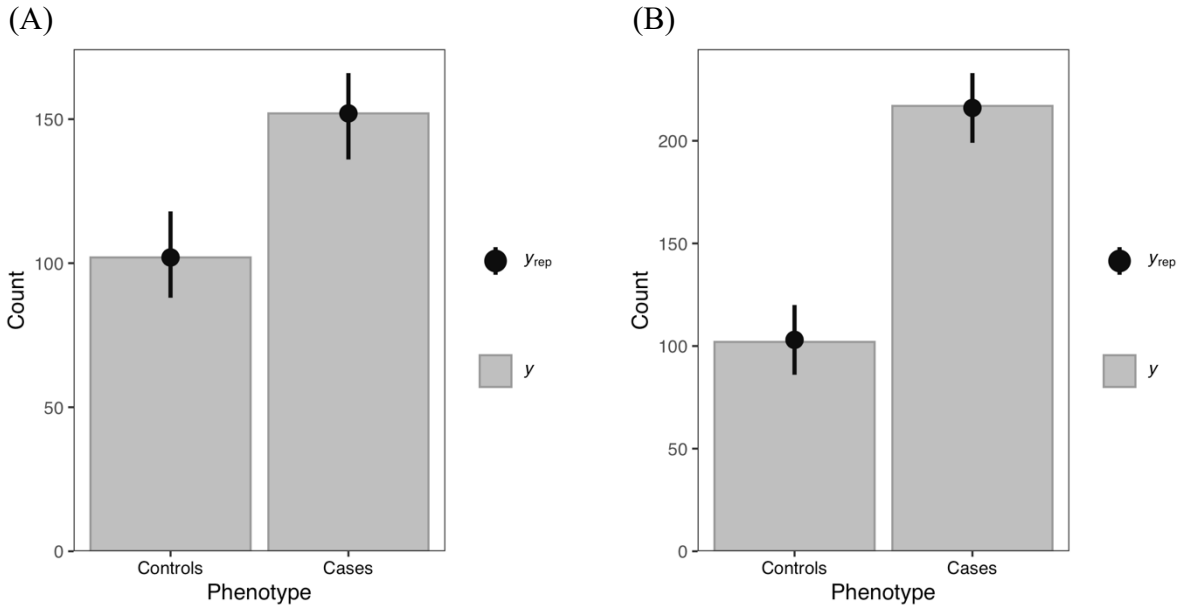


Figure III-S7. Posterior predictive check for Bayesian estimation of SNP effects of suggestive SNPs from the generalized linear mixed model regression for (A) sole ulcers and (B) noninfectious claw lesions. Gray bars represent the distribution of observed cases and controls, and black dots with intervals represent the median and uncertainty intervals of replicate phenotypes (y_{rep}) simulated using estimated SNP effects.

9 DATA AVAILABILITY STATEMENT

The microarray datasets generated for this study can be found in NCBI's Gene Expression Omnibus data repository (GEO series record GSE159157 and GSE165945).

Chapter IV. Pleiotropic loci associated with foot disorders and common periparturient diseases in Holstein cattle

This chapter has been submitted for publication in the Journal of Animal Science and Biotechnology.

Abstract

Background: Lameness is an animal welfare issue that incurs substantial financial and environmental costs. This condition is commonly caused by digital dermatitis (DD), sole ulcers (SU), and white line disease (WLD). Susceptibility to these three foot disorders is due in part to genetics, indicating that genomic selection against these foot lesions can be used to reduce lameness prevalence. It is unclear whether selection against foot lesions will lead to increased susceptibility to other common diseases such as mastitis and metritis. Thus, the aim of this study was to determine the genetic correlation between causes of lameness and other common health disorders to identify loci contributing to the correlation.

Results: Genetic correlation estimates between SU and DD and between SU and WLD were significantly different from zero ($p < 0.05$), whereas estimates between DD and mastitis, DD and milk fever, and SU and metritis were suggestive ($p < 0.1$). All five of these genetic correlation estimates were positive. Two-trait GWAS for each of these five pairs of traits revealed common regions of association on BTA1 and BTA8 for pairs that included DD or SU as one of the traits, respectively. Other regions of association were unique to the pair of traits and not observed in GWAS for other pairs of traits.

Conclusions: The positive genetic correlation estimates between foot disorders and other health disorders imply that selection against foot disorders may also decrease susceptibility to other health disorders. Linkage disequilibrium blocks defined around significant and suggestive SNPs from the two-trait GWASs included genes and QTL that were functionally relevant, supporting that these regions included pleiotropic loci.

Key words: Pleiotropy, multivariate, genome-wide association study, dairy cattle, lameness, disease, genetic correlation

1 BACKGROUND

Abnormal gait or posture in a cow are considered indicators of lameness and signifies pain and discomfort. Lameness is the second most prevalent disease after mastitis and the third most common reason for culling after mastitis and infertility (USDA, 2018). Lameness is commonly caused by foot lesions classified as infectious (e.g. digital dermatitis (DD), heel horn erosion, and foot rot) or noninfectious lesions (e.g. sole hemorrhage, sole ulcer (SU), white line disease (WLD), and laminitis). The etiology of heel horn erosion is not well known, though this foot disorder is commonly categorized as infectious because heel horn erosion is more prevalent in damp, unhygienic conditions like manure slurry (Bergsten and Herlin, 1996), often coincides with digital dermatitis (Knappe-Poindecker et al., 2013), and decreases with disinfecting (but not water) footbaths (Fjeldaas et al., 2014). Lameness not only raises welfare concerns, but also has economic and environmental consequences. Financial costs associated with lameness include direct costs for treatment and increased labor and indirect costs from reduced milk production and fertility; together these costs range from $64 per case of DD to $178 per case of SU (Cha et al., 2010; Dolecheck and Bewley, 2018; Dolecheck et al., 2019). Reduced fertility, premature culling, and reduced milk production associated with lameness reduces the efficiency of resource use, as resources used for the cow are invested over a less productive and shorter lifetime, inflating the environmental costs per unit of milk by 14 (1.5%) kg CO_2 equivalents per ton of fat-and-protein-corrected milk, on average (of DD, SU, and WLD combined) (Mostert et al., 2018).

Prevention of lameness is achieved through routine claw trimming, foot baths (for prevention of infectious causes), maintaining floor hygiene, and nutrition. Despite these prevention efforts, lameness remains highly prevalent in the United States, affecting 16.8% of cows and 3.2% of bred heifers (USDA, 2018). These non-genetic methods of prevention can be

aided by genetic selection, as implied by the low to moderate estimates of heritability for foot lesions, ranging from 0.01 to 0.4 for DD, 0.01 to 0.3 for SU, and 0.017 to 0.26 for WLD (Van der Waaij et al., 2005; Onyiro et al., 2008; van der Linde et al., 2010; Häggman and Juga, 2013; Oberbauer et al., 2013; van der Spek et al., 2013, 2015a; Malchiodi et al., 2015a; Biemans et al., 2018). Genetic selection uses prior knowledge about the contribution of certain genetic markers to traits of interest and creating a selection index reflecting a weighted average of multiple traits that is used to rank animals. Selective breeding programs utilize both genetic correlation among traits that are included in the selection index and specific susceptibility loci associated with the traits. Accordingly, selection against foot disorders would likely account for correlated lesion traits because some foot disorders are genetically correlated with each other, particularly within the infectious (strongest between DD and heel erosion) and noninfectious (strongest among sole hemorrhage, SU, and WLD) groupings of lesions (Koenig et al., 2005; Van der Waaij et al., 2005; van der Linde et al., 2010; Buch et al., 2011; Gernand et al., 2012; Häggman and Juga, 2013; van der Spek et al., 2013; Pérez-Cabal and Charfeddine, 2015; Malchiodi et al., 2017).

Additionally, certain foot lesions are genetically correlated with mastitis or indicator traits of mastitis. For example, the genetic correlations between clinical mastitis and sole hemorrhage or SU were estimated at 0.35 and 0.32, respectively, in Swedish Red cows (Buch et al., 2011). For Holstein cows, the genetic correlation between somatic cell score and individual foot lesions or lameness in general ranged from 0.15 to 0.24 (Koenig et al., 2005), and 0.23 (Gernand et al., 2012), respectively, although other studies failed to identify significant genetic correlations between DD or interdigital hyperplasia and clinical mastitis (Buch et al., 2011; Gernand et al., 2013). Nevertheless, the genetic correlation among foot disorders and between individual foot disorders and mastitis traits imply that common loci may coordinately influence

these traits (Koenig et al., 2005; Buch et al., 2011). Such pleotropic loci have not been identified as of yet. The values for genetic correlation between foot and other health disorders that have been reported were estimated using pedigree information. To our knowledge, no DNA-based studies have been performed to estimate the genetic correlation between foot disorders and disease traits other than mastitis. Using genomic data from individual cows to estimate relationships may be more accurate than using pedigree data (Goddard, 2009; Hayes et al., 2009) because using genomic data reduces the standard error of the genetic correlation estimate (Visscher et al., 2014). Therefore, the aim of this study was to identify loci associated with susceptibility to multiple foot disorders and other common diseases, which could be coordinately used to inform breeding programs.

2 MATERIALS AND METHODS

2.1 Phenotypes

Five large commercial dairies (Dairies A-E, each with > 1000 cows) in Northern and Central California participated in this study. Phenotypes were derived from hoof trimming and other health records provided by the dairies. Three hoof trimmers recorded the foot lesions used for phenotyping foot lesions, one who serviced Dairies A, B, and C; one who serviced Dairy D, and another who serviced Dairy E. Hoof trimmer experience and hoof trimming regimens were described previously (Lai et al., 2020, 2021). Foot disorders recorded included DD, foot rot, sole hemorrhage, SU, WLD, wall abscess, sole abscess, heel abscess, and laminitis. Other health events were also recorded by dairy personnel, which included diarrhea, displaced abomasum, ketosis, mastitis, metritis, milk fever, pneumonia, and retained placenta. For each foot or other health disorder, cases were defined as cows with at least one record of the disorder and controls were defined as cows that did not have records of the given foot or health disorder.

Consequently, for each trait, controls included cows with disorders other than the disorder the cases had.

2.2 Genotypes

Whole blood samples were obtained and the buffy coat was used to extract genomic DNA using the QIAGEN QIAamp DNA Blood Mini Kit (QIAGEN Inc., Valencia, CA). DNA samples were quantified using the NanoDrop (ND-2000 v3.2.1) spectrophotometer (Thermo Scientific, Wilmington, DE) and sent to GeneSeek (Lincoln, NE) for SNP genotyping on the high-density BovineHD BeadChip (777K SNPs, Illumina Inc., San Diego, CA). Genotype calls were made using Illumina's GenCall algorithm. SNP genotypes from a subset of the cows used in this study were used in our past studies (Lai et al., 2020, 2021) and are publicly available at the NCBI Gene Expression Omnibus database (GEO series record GSE159157 and GSE165945), along with the additional samples from this study (GSE to be added when received from GEO). SNP genotypes were updated to the ARS-UCD1.2 assembly positions (Rosen et al., 2020) and quality-filtered in PLINK 1.9 (Purcell and Chang, 2015) by removing from further analyses SNPs and cows with < 95% genotyping rate, SNPs with significant deviation from Hardy-Weinberg equilibrium ($p < 1E-6$) to exclude systematic genotyping errors, and SNPs with minor allele frequency < 5% to exclude rare variants. Missing genotypes for each cow were imputed using BEAGLE 5.1 (Browning et al., 2018) using the other cows in the sample population as the reference population, an effective sample size of 58 for the United States Holstein cattle population (Makanjuola et al., 2020), and default parameters. Genetic similarity among cows was visualized in a multidimensional scaling (MDS) plot depicting the first two dimensions.

2.3 Estimation of genetic correlation

Genetic correlation was estimated between each foot lesion and other health trait, including other foot lesions (e.g. genetic correlation was estimated between SU and WLD, SU and DD, SU and mastitis, SU and metritis, etc.) using cows that had phenotypes for both traits and at least 40 case cows for each disease. PLINK 2.0 was used to filter cows by requiring phenotypes for both traits (Chang et al., 2015; Purcell and Chang, 2021). The genome-wide complex trait analysis (GCTA) program was used to calculate the genetic relatedness matrix (GRM), which was used with farm as a covariate to estimate the additive genetic variance and covariance between the two traits using two-trait genome-based restricted maximum likelihood (GREML) (Yang et al., 2011; Lee et al., 2012). Specifically, the phenotype for trait 1 of the *k*-th cow was modeled as a function of the phenotypic contribution from the *j*-th SNP and the *i*-th farm:

$$y_{1_{ijk}} = \mu_1 + F_{1_i} + S_{1_j} + a_{1_k} + \varepsilon_{1_{ijk}}$$

where μ_1 was an unknown constant common to all cows for trait 1, F_{1_i} was contribution of *i*-th farm to the risk of disease, S_{1_j} was the contribution of the *j*-th SNP genotype to risk of the disorder, and a_{1_k} were the additive genetic effects assumed to be drawn from the multivariate normal density N(0, $A\ \sigma_a^2$), where A was the GRM and σ_a^2 is the variance of the additive genetic effects. $\varepsilon_{1_{ijk}}$ was the residual term for trait 1. Similarly, the phenotype for trait 2 of the *k*-th cow was modeled using the same components for trait 2 as

$$y_{2_{ijk}} = \mu_2 + F_{2_i} + S_{2_j} + a_{2_k} + \varepsilon_{2_{ijk}}$$

Additive genetic variance for each trait and covariance between the two traits were estimated and used to calculate genetic correlation. All genetic correlation estimates were transformed from the observed scale (0/1) to the underlying liability scale to account for case

ascertainment using the prevalence of each disorder obtained from the literature (Oberbauer et al., 2013; USDA, 2018). Genetic correlation estimates were considered significantly different from zero if the estimate had $p < 0.05$ from the likelihood ratio test, and suggestive genetic correlation estimates were those with $p < 0.1$.

2.4 Two-trait genome-wide association analyses

Pairs of traits that had significant or suggestive genetic correlation estimates using the frequentist approach were evaluated further in two-trait GWAS to identify regions potentially contributing to both traits. Multi-trait association testing can improve the power to detect associations while accounting for population stratification (Banerjee et al., 2008; Korte et al., 2012; Zhou and Stephens, 2012, 2014) because the additional information from the covariance of traits is still informative, even if only one of the traits is associated with the genotype (Stephens, 2013). Two-trait genome-wide association analysis was performed to test for association of each SNP with at least one of the traits. A standardized GRM was constructed and included in the linear mixed model to account for relatedness and population stratification, and farm was included as a covariate to adjust for differences among farms. The linear mixed model association testing was conducted using the multivariate association testing function in the genome-wide efficient mixed model association (GEMMA) program (Zhou and Stephens, 2012, 2014). Bonferroni correction for multiple testing assumes that each test for SNP association with phenotype(s) is independent. However, because SNPs are not independent due to linkage disequilibrium (LD) between SNPs, the Genetic Type I error calculator (GEC) was used to calculate the effective number of markers after accounting for linkage disequilibrium between SNPs for use as the denominator in Bonferroni-corrected thresholds of significance (Li et al., 2012). Genome-wide significant SNPs were thus defined as those with likelihood ratio test

(LRT) $p < 0.05/M_e$ and suggestive SNPs, as those with LRT $p < 1/M_e$ (Lander and Kruglyak, 1995). Manhattan and quantile-quantile plots were generated using the qqman package in R (R Development Core Team, 2010; Turner, 2014).

Because SNPs are likely not causal for the traits, but rather more likely in LD with causal variants, SNPs were used to define LD blocks that were then mined for overlap with genes and previously defined QTL. SNPs in LD with significant and suggestive SNPs were used to define the start and end of LD blocks using a method similar to Richardson et al. (2016) and Twomey et al. (2019). SNPs that were within 5 Mb (upstream or downstream) and in LD ($R^2 > 0.5$) with significant or suggestive SNPs were considered belonging to the same LD block. LD blocks were queried in the region search of FAANGMine (FAANG, 2019) to identify genes within or overlapping with the LD block. LD blocks were also queried for overlap with previously defined QTL and associations related to feet and legs conformation traits and disease traits in the Cattle QTLdb (Hu et al., 2019) (version 46, accessed 4/30/2021). Functions of genes that were considered relevant to the etiology of each disorder were defined for each trait (Table IV-S1**Error! Reference source not found.**) and included those with a role in immune function, hair follicle morphology, hair density, skin integrity, fibroblast proliferation, bone growth and mineralization, adipose morphology and amount, and glucose metabolism.

3 RESULTS

3.1 Descriptive data

Hoof trimming records were available for 21044 cows across the five dairies (distribution of records for each type of foot lesion is described in detail by Lai et al. (2021), of which 418 cows were selected for SNP genotyping as controls or cases for a certain foot lesion(s). Traits that were recorded at multiple dairies were used for genetic correlation estimation and two-trait

GWAS, and the distribution of case/control phenotypes for each trait are listed in Table IV-1. All five dairies recorded SU, WLD, and DD foot disorders. All dairies except Dairy C also had health records available for phenotyping other health traits. These four dairies (Dairies A, B, D, and E) recorded mastitis, metritis, and pneumonia. Dairies A, B, and E also recorded ketosis, retained placenta, diarrhea, milk fever, and displaced abomasum. After excluding traits that had ≤ 40 cases, genetic correlation was estimated between each pair of foot disorders (SU, WLD, and DD) as well as each foot disorder with another health disorder (mastitis, metritis, retained placenta, milk fever, and pneumonia).

Table IV-1. Count of genotyped cows after quality filtering, split by cases for each foot disorder or other health condition and controls across the five dairies

	Farm					
	A	B	C	D	E	Total
Datasets for foot disorders						
Sole ulcer						
Cases	44	8	4	71	25	152
Controls	138	70	26	23	0	257
White line disease						
Cases	48	13	7	33	16	117
Controls	134	65	23	61	9	292
Digital dermatitis						
Cases	19	22	30	30	5	106
Controls	163	56	0	64	20	303
Datasets for other disorders						
Mastitis						
Cases	89	66	NR	77	17	249
Controls	93	12	NR	17	8	130
Metritis						
Cases	57	51	NR	8	15	131
Controls	125	27	NR	86	10	248
Ketosis						
Cases	13	17	NR	NR	0	30
Controls	169	61	NR	NR	25	255
Retained placenta						
Cases	16	35	NR	NR	0	51
Controls	166	43	NR	NR	25	234
Diarrhea						
Cases	19	0	NR	NR	1	20
Controls	163	78	NR	NR	24	265
Milk fever						
Cases	61	9	NR	NR	0	70
Controls	121	69	NR	NR	25	215
Displaced abomasum						
Cases	1	17	NR	NR	2	20
Controls	181	61	NR	NR	23	265
Pneumonia						
Cases	2	4	NR	22	13	41
Controls	180	74	NR	72	12	338

NR: No records available; cows were excluded from analyses

Quality filtering removed nine cows and 218306 SNPs, leaving 409 cows with 559656 SNP genotypes for analyses with case/control phenotypes presented in Table IV-1. The MDS plot indicated slight population stratification with a prominent center cluster, though cows were not strongly stratified by farm (Figure IV-S1).

3.2 Genetic correlation estimates

Of the pairs of traits for which genetic correlations were estimated, genetic correlation estimates between SU and WLD and between SU and DD were significantly different from zero ($p < 0.05$), and estimates between DD and mastitis, DD and milk fever, and SU and metritis were suggestive ($p < 0.1$, Table IV-2). Consequently, each pair of these traits was analyzed in two-trait GWAS.

Table IV-2. Genetic correlation estimates (and standard error, SE) between sole ulcer (SU), white line disease (WLD), digital dermatitis (DD), and other health traits that were significantly or suggestively different from zero

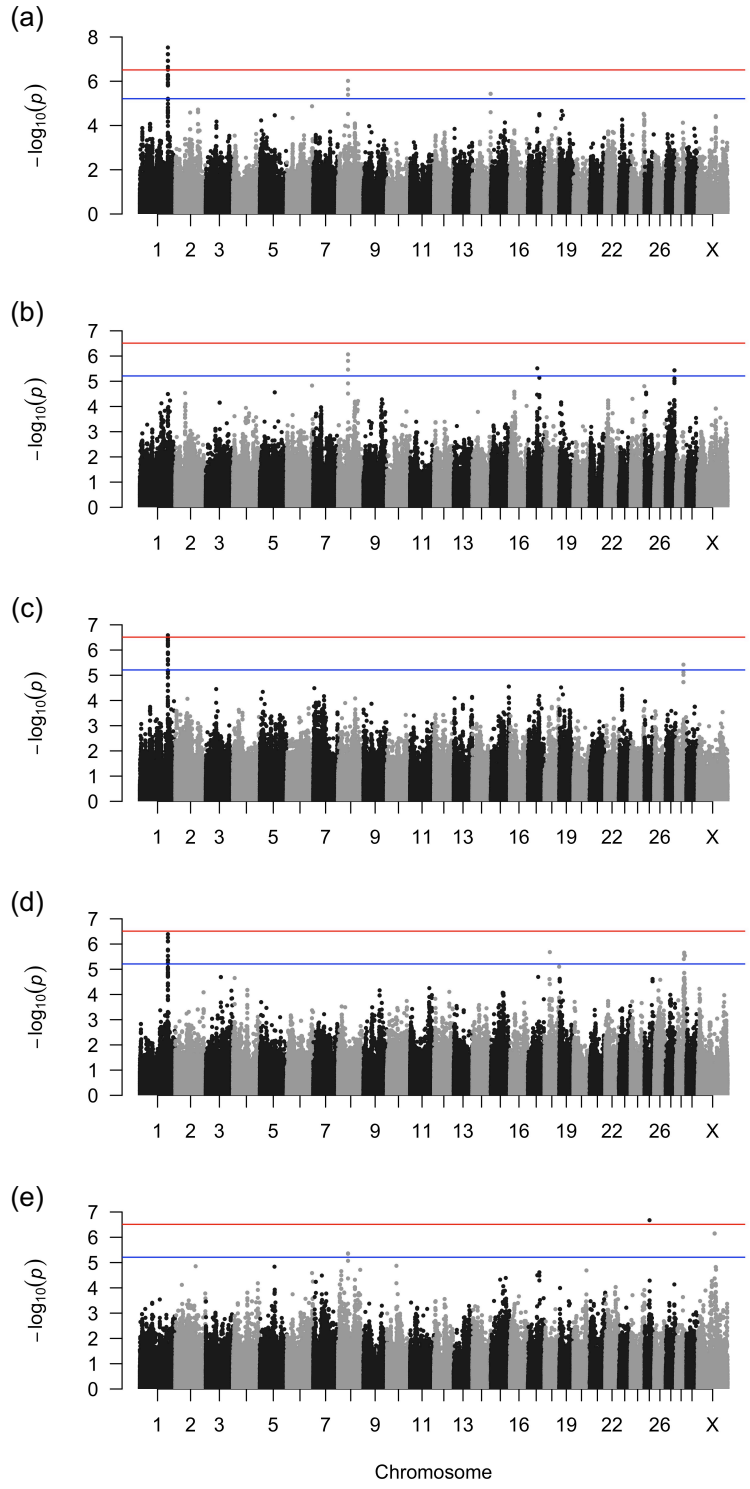
Trait 1	Trait 2	Genetic correlation (SE)	p	Significance
SU	DD	0.46 (0.25)	4.81E-02	*
SU	WLD	0.92 (0.46)	2.54E-02	*
DD	Mastitis	0.49 (0.36)	7.77E-02	†
DD	Milk fever	0.49 (0.39)	9.46E-02	†
SU	Metritis	0.70 (0.46)	5.22E-02	†

* genome-wide significant (Bonferroni-corrected $p < 0.05$)
† genome-wide suggestive significance (Bonferroni-corrected $p < 0.10$)

3.3 Two-trait genome-wide association analysis

The effective number of markers after accounting for LD was 162435 SNPs, corresponding to a suggestive threshold of 6.2×10^{-6} (5.2 on the $-\log_{10}(p)$ scale) for genome-wide suggestive significance and 3.1×10^{-7} (6.5 on the $-\log_{10}(p)$ scale) for genome-wide significance. Manhattan plots from the two-trait GWAS are shown in Figure IV-1. Genomic inflation factors

ranged from 1.02 to 1.06 and, combined with the qqplots (Figure IV-S2), indicated that



population stratification had been accounted for sufficiently.

Figure IV-1. Manhattan plot for two-trait genome-wide association analysis of (a) sole ulcer (SU) and digital dermatitis (DD), (b) SU and white line disease (WLD), (c) DD and mastitis, (d)

DD and milk fever, and (e) SU and metritis. The blue line indicates genome-wide suggestive significance, and the red line indicates genome-wide significance.

Significant and suggestive SNPs and the LD blocks they defined are shown in

Table **IV-S2**, and the most significant SNP (i.e., lowest p-value) within each LD block are also listed in Table IV-3 for brevity. Supplemental materials report the genes and QTL LD blocks (Figure IV-S1, Table IV-S3, and Table IV-S4). The GWAS that included DD as one of the traits (DD and mastitis, SU and DD, DD and milk fever) identified significant and suggestive SNPs belonging to the same LD block at BTA1:125550933-125822143. For the DD and mastitis and DD and milk fever GWAS, the peak on BTA1 reached or approached genome-wide significance despite the genetic correlation estimate only reaching suggestive significance (Table IV-1 and

Table **IV-S2**). GWA analyses that included SU as one of the traits (that is, between SU and WLD, SU and DD, and SU and metritis) all identified suggestive SNPs in an LD block at BTA8:42926603-44642925. Other SNP associations were unique to the pair of traits for which the GWAS was performed such that SNPs that were associated in a certain GWAS for a pair of traits were not associated in other GWASs for other pairs of traits. For instance, the LD block on BTA14 detected from the GWAS for SU and DD was only detected in the SU and DD GWAS and not detected in any other of the comparisons such as that for SU and WLD, DD and mastitis, DD and milk fever, and SU and metritis. Although SU and WLD were strongly genetically correlated (0.92), four of the seven suggestive SNPs identified in the two-trait GWAS had opposite effect signs between the two traits: the effects of four of the suggestive SNPs were negative for SU and positive for WLD (Table IV-3 and

Table **IV-S2**). The LD blocks defined from all the two-trait GWA analyses overlapped with 83 protein-coding genes, some functionally relevant to the etiology of the disorders (Table

IV-S3).

Table IV-3. Linkage disequilibrium (LD) blocks and the most significant SNP (i.e., lowest p-value) within the LD block defined from the two-trait genome-wide association analyses for pairs of traits including sole ulcer (SU), white line disease (WLD), digital dermatitis (DD) and other health disorders

										Variance matrix for beta effects				
Dataset (trait 1 and trait 2)	**BTA**	**LD block start (bp)**	**LD block end (bp)**	**LD block length (kb)**	**Most significant SNP**	**SNP position (bp)**	**Minor/ Major allele**	**Effect size for trait 1**	**Effect size for trait 2**	**Variance of effect size for trait 1**	**Covariance between effect sizes of trait 1 and 2**	**Variance of effect size for trait 2**	**P**	
SU and DD	1	125550933	125822143	271.21	BovineHD0100035768	125563251	A/G	1.85E-01	4.64E-03	1.09E-03	3.13E-05	1.46E-03	2.58E-07	*
	8	42926603	44642925	1716.32	BovineHD0800013406	44628587	T/C	2.05E-01	-8.65E-02	2.30E-03	6.91E-04	2.99E-03	3.06E-06	†
	14	81655298	81664096	8.80	BovineHD1400023802	81655298	G/T	-3.00E-02	1.58E-01	1.05E-03	3.17E-04	1.21E-03	3.68E-06	†
SU and WLD	8	42926603	44642925	1716.32	BovineHD0800013408	44632844	G/T	4.90E-02	1.91E-01	9.97E-04	-9.75E-06	1.14E-03	2.11E-07	*
	17	41328134	41328134	0	BovineHD1700011766	41328134	C/T	-1.21E-01	1.32E-01	1.12E-03	8.92E-05	1.23E-03	7.07E-07	†
	27	37518206	38922466	1404.26	BovineHD2700011209	38898651	T/C	-1.21E-01	1.32E-01	1.12E-03	8.92E-05	1.23E-03	7.07E-07	†
	27	37518206	38922466	1404.26	BovineHD2700011210	38901656	G/A	-1.21E-01	1.32E-01	1.12E-03	8.92E-05	1.23E-03	7.07E-07	†
DD and mastitis	1	125550933	125822143	271.21	BovineHD0100035835	125691064	A/G	1.02E-01	1.72E-01	1.06E-03	1.61E-04	9.47E-04	2.98E-08	*
	28	33357088	33385923	28.835	BovineHD2800009006	33385923	C/T	9.39E-02	1.62E-01	1.01E-03	1.59E-04	9.07E-04	1.20E-07	*
DD and milk fever	1	125550933	125822143	271.21	BovineHD0100035785	125585828	C/T	8.88E-02	1.64E-01	1.01E-03	1.61E-04	9.08E-04	1.19E-07	*
	1	125550933	125822143	271.21	BovineHD0100035800	125624770	A/C	8.88E-02	1.64E-01	1.01E-03	1.61E-04	9.08E-04	1.19E-07	*
	18	24087895	24329676	241.78	BovineHD1800007458	24087895	C/T	6.46E-02	1.50E-01	9.06E-04	1.51E-04	8.12E-04	8.13E-07	†
	28	34935232	35093950	158.72	BTB-00987935	35093950	G/T	6.46E-02	1.50E-01	9.06E-04	1.51E-04	8.12E-04	8.13E-07	†
	28	35837718	36740498	902.78	BovineHD2800010153	36916301	C/T	6.52E-02	1.49E-01	8.93E-04	1.47E-04	8.00E-04	7.86E-07	†
	28	35837718	36740498	902.78	BovineHD2800010156	36926419	C/T	6.52E-02	1.49E-01	8.93E-04	1.47E-04	8.00E-04	7.86E-07	†
	28	38776483	42482917	3706.43	BovineHD2800011177	40061500	G/A	8.15E-02	1.45E-01	9.42E-04	1.53E-04	8.54E-04	1.40E-06	†
SU and metritis	8	42926603	44642925	1716.32	BovineHD0800013412	44642925	A/C	6.16E-02	1.57E-01	9.06E-04	1.50E-04	8.03E-04	2.22E-07	*
	25	22127459	22966511	839.05	BovineHD2500006264	22127459	A/G	1.86E-01	-2.77E-02	1.52E-03	2.99E-04	1.49E-03	4.09E-06	†
	X	75319558	75610976	291.42	BovineHD3000022324	75393744	A/G	2.01E-01	-1.21E-02	1.49E-03	2.90E-04	1.46E-03	9.61E-07	†

BTA: *Bos taurus* autosome

* = genome-wide significance

† = genome-wide suggestive significance

4 DISCUSSION

We estimated the genetic correlation between common foot disorders (DD, SU, and WLD) and other health traits (mastitis, metritis, milk fever, retained placenta, and pneumonia). For pairs of traits having significant or suggestive genetic correlation, the loci which were contributing to the correlation were examined using two-trait GWAS. To our knowledge, this is the first study to estimate genetic correlation between foot disorders and diseases other than mastitis from individual-level genotype data rather than pedigree data and identify loci potentially contributing to the correlation. Genetic correlation estimates that were significant or suggestive included SU or DD as one of the traits (SU and DD, SU and WLD, DD and mastitis, DD and milk fever, and SU and metritis) and estimates were positive, indicating a favorable genetic correlation between pairs of disease traits such that genetic selection against one disease will lead to selection against the other disease. Genetic correlation estimates were positive despite some of the significantly and suggestively associated SNPs for each of these five pairs of traits having effects with opposite signs between the two traits. The opposite signs in the SNP effects for some of the top SNPs were likely overruled by the concordant signs in SNP effects for less significant SNPs, leading to an overall positive genetic correlation. Significant and suggestive SNPs were detected in the same regions on BTA1 and BTA8 for two-trait GWAS datasets that had DD and SU as one of the traits, respectively, suggesting DD and SU were driving the association in these genomic regions. Other significant and suggestive SNPs were specific to the dataset from which they were detected and not detected in GWA analyses for other pairs of traits.

Compared to previous estimates of genetic correlation between foot disorders and other health traits, estimates from this study were higher and had larger standard errors. Previous

estimates of genetic correlation between foot disorders and mastitis or somatic cell count were significantly different from zero (0.15 to 0.35) (Koenig et al., 2005; Buch et al., 2011) or close to zero (Gernand et al., 2012), whereas we estimated the genetic correlation between DD and mastitis at 0.49 (SE = 0.36). The genetic correlation between SU and WLD was 0.92 (SE = 0.46) and substantially higher than previous estimates, which ranged from 0.41 to 0.60 (van der Linde et al., 2010). The estimates of genetic correlation from this study were higher likely because controls were shared between the two traits and the proportion of cows with DD and/or SU was higher than for other disorders. Because case cows were sampled primarily for DD and SU and other disorders were phenotyped after sampling DD and SU cases, cases for other disorders frequently also had DD and/or SU. This overrepresentation of cases with DD and/or SU in addition to the disorder of interest likely inflated genetic correlation estimates, which the correction for case ascertainment was unable to overcome. The strong genetic correlation between SU and WLD in this study implied that whichever other traits SU is correlated with, WLD will also be correlated with and vice versa; however, SU was correlated with metritis and DD whereas WLD was not correlated with either disorder. This divergence would suggest that although SU and WLD share a genetic component, differences exist in the location or direction of the effect for susceptibility loci between SU and WLD, as indicated by the opposite signs of some of the suggestive SNP effects between the traits and the lack of association of WLD to metritis or DD in the two-trait GWA analyses.

Compared to our previous one-trait GWAS for DD and SU, the two-trait GWAS detected the same LD block on BTA1 for DD and a different LD block on BTA8 for SU. Specifically, the LD block at BTA1:125550933-125822143 common to all datasets that had DD as one of the traits (DD and mastitis, SU and DD, and DD and milk fever) was the same LD block detected in

our previous single trait DD GWAS (Lai et al., 2020). The increase in significance of association also suggests that this region may play a role in both infectious (mastitis) and metabolic (SU and milk fever) disorders. Infectious and metabolic disorders have been observed to coincide and happen most frequently during the early lactation period (USDA, 2018), potentially due to a common cause. Some have attributed the cause of higher incidence of infectious and noninfectious foot disorders during early lactation to the extreme negative energy balance during this period (Collard et al., 2000; Gernand et al., 2013). Accordingly, it is thought that cows that are better able to cope with the energy requirements during this period are consequently less susceptible to metabolic and infectious disorders, a hypothesis supported by the association of a more robust adaptive immune response with lower incidence of metabolic disease during the periparturient period (Thompson-Crispi et al., 2012). Another common LD block at BTA8:42926603-44642925 was detected from the two-trait GWAS with SU as one of the traits (SU and WLD, SU and DD, and SU and metritis). This LD block was 30 Mb upstream of the LD block on BTA8 observed in our previous one-trait SU GWAS (Lai et al., 2021). Our previous GWAS used the same SU cases but only sound, older (> 6.0 years old) cows as controls, whereas the present GWAS included controls with foot disorders other than the foot disorder the cases had. Consequently, the present GWAS controlled for other foot disorders that the cases had such that associated regions were more likely for SU specifically and not for other foot disorders correlated with SU, whereas the single-trait GWAS used the most phenotypically divergent cows as controls to maximize the power to detect genetic differences.

The LD blocks defined from each dataset overlapped with genes and/or QTL that were functionally relevant to both traits. The LD block at BTA1:125550933-125822143 from the GWAS that included DD as one of the traits contained *SLC9A9* (solute carrier family 9 member

A9) (Lai et al., 2020), which has been implicated in multiple sclerosis in humans through its role in regulating T-cell activation and differentiation to a induce a proinflammatory response (Esposito et al., 2015). Notably, the DD and mastitis LD block at 1:125839933-125852054 overlapped with a QTL associated with length of productive life (Cole et al., 2011), corroborating the shorter productive life associated with DD and mastitis susceptibility (Shabalina et al., 2020). Previous estimates of genetic correlation between foot lesion traits and productive life were close to zero (Dhakal et al., 2015), suggesting that uncorrelated traits may still share pleiotropic loci, as observed previously between various production, fertility, and conformation traits (Xiang et al., 2017). This LD block on BTA1 from the DD and mastitis GWAS and the LD block on BTA27 from the SU and WLD GWAS both overlapped with QTL for feet and legs conformation traits (Cole et al., 2011), and could be a pleiotropic locus contributing to the genetic correlation between feet and legs conformation and susceptibility to foot lesions (Häggman and Juga, 2013; Malchiodi et al., 2017; Ring et al., 2018), though this genetic correlation is too low to justify indirect selection on lameness using feet and legs conformation traits (Van Raden et al., 2021). The LD blocks from the GWAS for SU and DD, SU and WLD, and SU and metritis overlap with QTL for infectious disease traits (tuberculosis susceptibility, clinical mastitis, and somatic cell score/count) and blood cortisol, which may reflect the interplay of the stress from the negative energy balance during the periparturient period possibly potentiating metabolic and infectious foot disorders. Cows with SU tend to exhibit markers of chronic inflammation compared to cows without SU (O'Driscoll et al., 2015), though it is unclear if SU causes inflammation, vice versa, or both are the product of stress.

The main limitations of this study were the small sample size of genotyped cows and the variation in the number of case cows across the various disorders. At the expense of a larger

sample size, we minimized the environmental variation by constraining the sample population to cows to a small geographical region under similar management and nutrition practices and minimized the number of hoof trimmers to reduce variation in phenotyping foot lesions. Because health records were generated by different farm personnel within and among the farms, the other non-foot related disorders likely had more variation in diagnoses, which may have partially masked the genetic effect for these disorders and limited the power of the GWAS to detect significant associations. Minimizing environmental and consistent phenotyping improves the power to detect significant genetic correlation; however, the resulting small sample size limited the accuracy of genetic correlation estimates. For instance, one workaround for the inflation of genetic correlation estimates due to shared controls is to randomly partition the controls between the two traits before estimating genetic correlation; however, the small sample size prevented using this approach. The small sample size also limited the benefit of using genomic data instead of pedigree data to estimate genetic correlation. Although using genomic data to estimate relationships may be more accurate than using pedigree data (Goddard, 2009; Hayes et al., 2009) due to reduced standard error of the genetic correlation estimate (Visscher et al., 2014), the standard error of the genetic correlation estimates in this study were large, reflecting the limited sample size. The reduction in standard error from using genomic data would be more appreciable in larger sample sizes. Ascertainment bias for cows with DD and SU due to sample collection targeting DD- and SU-affected cows but not the other disorders likely led to an overrepresentation of cows with DD and/or SU in the dataset, resulting in inflated estimates between DD or SU and the other disorders. Despite the inflated and large standard errors of the genetic correlation estimates, some estimates were significantly or suggestively different from zero and provided grounds for further investigation of SNPs contributing to the correlation using

the two-trait GWAS. The sample size also provided sufficient power in the two-trait GWAS to detect significant and suggestive SNPs that defined LD blocks overlapping with functionally relevant genes and QTL, similar to previous GWA analyses using similar small sample sizes (~400 cows) and high-density SNP genotypes (Buzanskas et al., 2017; Lehner et al., 2018).

5 CONCLUSIONS

A genomic relatedness matrix calculated from SNP genotypes was used to estimate genetic correlation between individual foot disorders (DD, SU, and WLD) and other health disorders (mastitis, metritis, milk fever, retained placenta, and pneumonia). Genetic correlation between SU and WLD and between SU and DD were significantly greater than zero, and estimates between DD and mastitis, DD and milk fever, and SU and metritis were suggestively greater than zero. Although some of the significant and suggestive SNP effects had opposite signs between the two traits, other SNP effects had concordant directions that collectively outweighed the opposing SNP effects and led to positive genetic correlation estimates for each of these five trait pairs. The positive estimates of genetic correlation between individual foot disorders and other health disorders indicate that direct selection against foot disorders will not increase the incidence of other health disorders and may in fact reduce their prevalence. Genomic assessment for pairs of traits that were genetically correlated revealed multiple associated regions. Whereas some of these chromosomal regions were shared across multiple pairs of traits that included SU or DD as one of the traits, others were unique to the pair of traits, indicating the complexity of genetic contributions within and between traits. The LD blocks defined from associated SNPs included protein-coding genes and QTL that were functionally relevant to both traits, suggesting that selection for markers in these LD blocks would affect susceptibility to both traits.

6 SUPPLEMENTARY MATERIAL

6.1 Supplementary tables

Table IV-S1. Gene functions that were considered relevant to the etiology of digital dermatitis (DD), sole ulcers (SU), white line disease (WLD), and other health disorders.

	Infectious			Noninfectious		
Gene function	**DD**	**Mastitis**	**Metritis**	**SU**	**WLD**	**Milk fever**
Adipose/fat				x	x	
Bone				x	x	x
Cartilage				x	x	
Chondrocyte				x	x	
Immune	x	x	x			
Hair	x					
Skin	x					
Collagen				x	x	
Glucose metabolism				x	x	
Fibroblast proliferation				x	x	

Table IV-S2. Significant and suggestive SNPs detected in the two-trait linear mixed model association analysis and the linkage disequilibrium (LD) blocks they defined for sole ulcers (SU) and digital dermatitis (DD), SU and white line disease (WLD), DD and mastitis, DD and milk fever, and SU and metritis.

Table S2 is available in the Excel workbook containing all supplemental tables.

Table IV-S3. Genes that overlapped with linkage disequilibrium blocks defined from the genome-wide association analyses for sole ulcers (SU) and digital dermatitis (DD), SU and white line disease (WLD), DD and mastitis, DD and milk fever, and SU and metritis. Phenotypes associated with mouse knockout models of protein-coding genes and the functional relevance of the phenotype to the dataset, if applicable, are also listed. (See Table S4 for which phenotypes were considered functionally relevant for each disorder.)

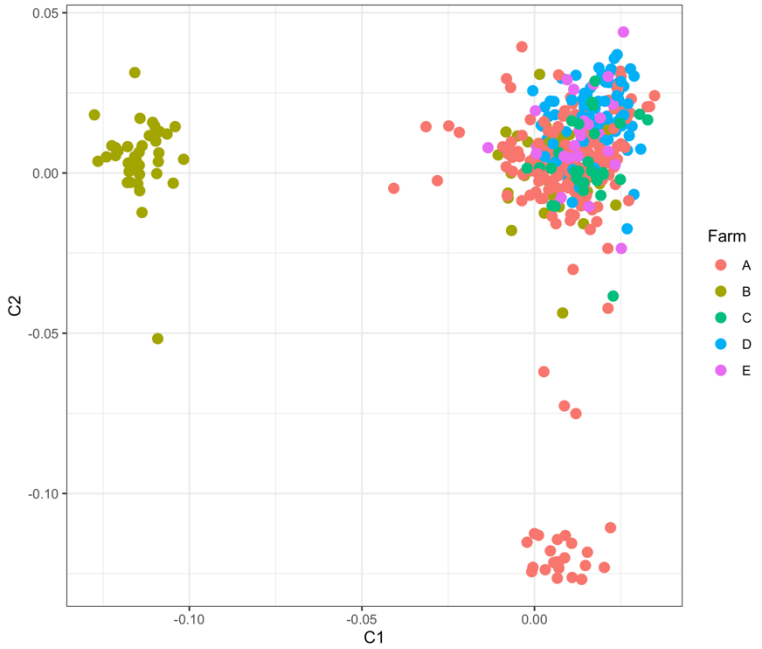
Table S3 is available in the Excel workbook containing all supplemental tables.

Table IV-S4. Previously defined quantitative trait loci (QTL) or associations overlapping with or in the linkage disequilibrium (LD) blocks defined from the two-trait genome-wide association analyses for sole ulcers (SU) and digital dermatitis (DD), SU and white line disease (WLD), DD and mastitis, DD and milk fever, and SU and metritis.

Table IV-S4 is also available in the Excel workbook containing all supplemental tables.

	LD block location					
Dataset	BTA	Start (bp)	End (bp)	Length (kb)	Overlapping QTL or association	Reference
DD and mastitis	1	125550933	125822143	271.2	Digital dermatitis	(Lai et al., 2020)
	1	125839933	125852054	12.1	Feet and legs conformation	(Cole et al., 2011)
					Length of productive life	(Cole et al., 2011)
	28	33357088	33385923	28.8	No overlap with functionally relevant QTL/associations	
SU and WLD	8	42926603	44642925	1716.3	Blood cortisol level	(Chen et al., 2020)
					Bovine tuberculosis susceptibility	(Richardson et al., 2016; González-Ruiz et al., 2019)
	17	41328134	41328134	0	No overlapping QTL/associations	
	27	37518206	38922466	1404.3	Foot angle	(Cole et al., 2011)
					Length of productive life	(Cole et al., 2011)
					Net merit	(Cole et al., 2011)
					Rear leg placement - rear view	(Cole et al., 2011)
					Rear leg placement - side view	(Cole et al., 2011)
					Somatic cell score	(Cole et al., 2011)
SU and DD	1	125550933	125822143	271.2	Digital dermatitis	(Lai et al., 2020)
	8	42926603	44642925	1716.3	Blood cortisol level	(Chen et al., 2020)
					Bovine tuberculosis susceptibility	(Richardson et al., 2016; González-Ruiz et al., 2019)
	14	81655298	81664096	8.8	No overlapping QTL/associations	
DD and milk fever	1	125550933	125822143	271.2	Digital dermatitis	(Lai et al., 2020)
	18	24087895	24329676	241.8	No overlapping QTL/associations	
	28	34935232	35093950	158.7	No overlapping QTL/associations	

Dataset	LD block location			Length (kb)	Overlapping QTL or association	Reference
	BTA	Start (bp)	End (bp)			
	28	35837718	36740498	902.8	No overlapping QTL/associations	
	28	38776483	42482917	3706.4	No overlapping QTL/associations	
SU and metritis	8	42926603	44642925	1716.3	Blood cortisol level	(Chen et al., 2020)
					Bovine tuberculosis susceptibility	(Richardson et al., 2016; González-Ruiz et al., 2019)
	25	22127459	22966511	839.1	Bovine respiratory disease susceptibility	(Neupane et al., 2018)
	X	75319558	75610976	291.4	No overlapping QTL/associations	



6.2 Supplemental figures

Figure IV-S1. Multidimensional scaling plot showing the first two dimensions for 409 cows from the five dairies used in the estimation of genetic correlation and genome-wide association analyses.

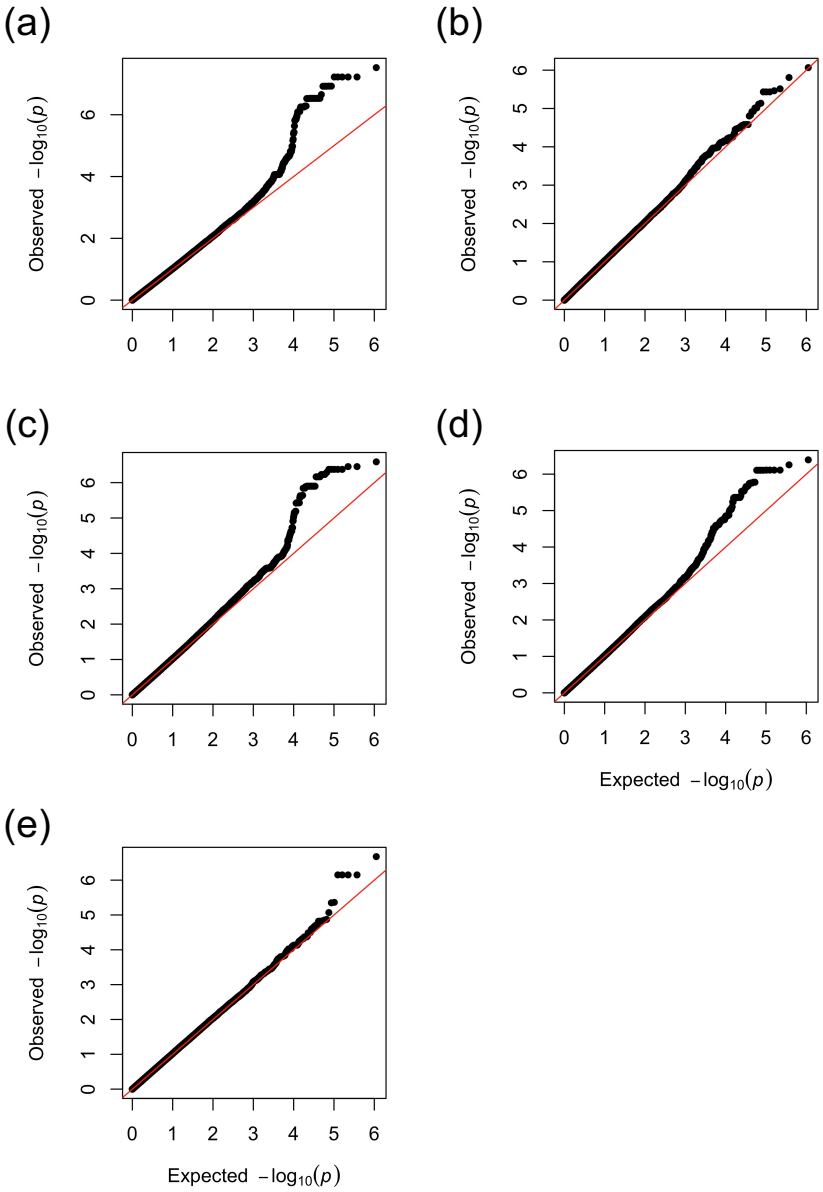


Figure IV-S2. Quantile-quantile plot for two-trait genome-wide association analysis of (a) sole ulcer (SU) and digital dermatitis (DD), (b) SU and white line disease (WLD), (c) DD and mastitis, (d) DD and milk fever, and (e) SU and metritis.

Abbreviations: *Bos taurus* autosome (BTA), digital dermatitis (DD), genetic relatedness matrix (GRM), Genetic Type I error calculator (GEC), genome-wide association study (GWAS), likelihood ratio test (LRT), linkage disequilibrium (LD), minor allele frequency (MAF), quantitative trait loci (QTL), single nucleotide polymorphism (SNP), sole ulcers (SU), white line disease (WLD)

7 DATA AVAILABILITY STATEMENT

The datasets generated and analyzed during the current study are available in the NCBI Gene Expression Omnibus database under the following GEO series records.

GSE159157: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159157

GSE165945: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM4820717

GSE[number and website to be added when GEO submission approved]

Conclusions and future directions

Lameness results in early culling of dairy cows from the herd and is the second most prevalent disease in dairy cattle after mastitis (USDA, 2018). Lameness is an animal welfare issue, incurs substantial financial losses for the producer, and inflates the environmental footprint per unit of milk due to losses in efficiency of resource use. Common causes of lameness are digital dermatitis (DD) colloquially referred to as foot warts, sole ulcers (SU), and white line disease (WLD). Through the work presented here, we have shown that there exist genetic regions having significant association with risk for these disorders. Although the actual genetic contribution of each region to risk is small, the existence of genetic variants indicates that concerted selection against these disorders can reduce the risk. Accordingly, the most effective method of reducing the prevalence of these foot disorders is through a combination of proactive/prophylactic management practices and genetically selecting for cows at lower risk of developing foot lesions.

Because non-genetic factors strongly influence lameness susceptibility, future research in larger populations, other geographical regions, and other types of production systems (e.g., pasture-based) is necessary to validate the genomic regions we identified. Additionally, future research should assess how well the associations hold in other ruminants because DD, SU, and WLD are also common causes of lameness in beef cattle, DD is also common in sheep and goats, and SU and WLD have also been observed in goats (Duncan et al., 2014; Groenevelt et al., 2015; Groenevelt, 2017). Because the etiology of DD is similar in beef cattle, sheep, and goats (Duncan et al., 2014; Nally et al., 2015; Sullivan et al., 2015a, 2015b; Magrin et al., 2018), it is plausible that similar genetic components play a role in susceptibility to DD in beef cattle and other ruminant species because the same treponemal phylotypes have been found in these

ruminants. How well the susceptibility loci are conserved across ruminant species remains unclear and is an area for future research. During the development of a selection index for lameness for Holstein cattle and eventually other breeds of dairy and beef cattle and other ruminants, the accuracy and potential to reduce lameness incidence should be compared between selection indices based on a few highly prevalent foot disorders and indices based on any lameness event. The accuracy and utility of selection indices based on a few vs. all types of lameness events will likely depend on the direction and magnitude of genetic correlation between types of lameness events. The continued use of individual-level genetic information (e.g., SNP genotypes) to supplement pedigree information (e.g., sire lines) in genomic evaluations for lameness will likely improve the accuracy of selection indices.

Funding

This research was funded by a Western Sustainable Agriculture Research and Education Graduate Student grant (project number GW18-126), U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service Animal Health Funding (project number 2250-AH), a Department of Animal Science Kellogg Endowment, and Jastro Shields awards from the College of Agricultural and Environmental Sciences at the University of California, Davis.

References

Banerjee, S., Yandell, B. S., and Yi, N. J. (2008). Bayesian quantitative trait loci mapping for multiple traits. *Genetics* 179, 2275–2289. doi:10.1534/genetics.108.088427.

Bergsten, C., and Herlin, A. H. (1996). Sole Haemorrhages and Heel Horn Erosion in Dairy Cows: The Influence of Housing System on their Prevalence and Severity. *Acta Vet. Scand.* 37, 395–408. doi:10.1186/BF03548079.

Berry, S. L., Read, D. H., Famula, T. R., Mongini, A., and Döpfer, D. (2012). Long-term observations on the dynamics of bovine digital dermatitis lesions on a California dairy after topical treatment with lincomycin HCl. *Vet. J.* 193, 654–658. doi:10.1016/j.tvjl.2012.06.048.

Berry, S. L., Read, D. H., Walker, R. L., and Famula, T. R. (2010). Clinical, histologic, and bacteriologic findings in dairy cows with digital dermatitis (footwarts) one month after topical treatment with lincomycin hydrochloride or oxytetracycline hydrochloride. *J. Am. Vet. Med. Assoc.* 237, 555–560. Available at: https://avmajournals.avma.org/doi/10.2460/javma.237.5.555?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub 0pubmed.

Bicalho, R. C., Machado, V. S., and Caixeta, L. S. (2009). Lameness in dairy cattle: A debilitating disease or a disease of debilitated cattle? A cross-sectional study of lameness prevalence and thickness of the digital cushion. *J. Dairy Sci.* 92, 3175–3184. doi:10.3168/jds.2008-1827.

Bicalho, R. C., and Oikonomou, G. (2013). Control and prevention of lameness associated with claw lesions in dairy cows. *Livest. Sci.* 156, 96–105. doi:10.1016/j.livsci.2013.06.007.

Biemans, F., Bijma, P., Boots, N. M., and de Jong, M. C. M. (2018). Digital Dermatitis in dairy cattle: The contribution of different disease classes to transmission. *Epidemics* 23, 76–84. doi:10.1016/j.epidem.2017.12.007.

Biemans, F., de Jong, M. C. M., and Bijma, P. (2019). A genome-wide association study for susceptibility and infectivity of Holstein Friesian dairy cattle to digital dermatitis. *J. Dairy Sci.* 102, 6248–6262. doi:10.3168/JDS.2018-15876.

Blowey, R. W., Ossent, P., Watson, C. L., Hedges, V., Green, L. E., and Packington, A. J. (2000). Possible distinction between sole ulcers and heel ulcers as a cause of bovine lameness. *Vet. Rec.* 147, 110–112. doi:10.1136/vr.147.4.110.

Brandt, S., Apprich, V., Hackl, V., Tober, R., Danzer, M., Kainzbauer, C., et al. (2011). Prevalence of bovine papillomavirus and Treponema DNA in bovine digital dermatitis lesions. *Vet. Microbiol. J.* 148, 161–167. doi:10.1016/j.vetmic.2010.08.031.

Brieuc, M. S. O., Waters, C. D., Drinan, D. P., and Naish, K. A. (2018). A practical introduction to random forest for genetic association studies in ecology and evolution. *Mol. Ecol. Resour.*

Britt, J. S., Carson, M. C., Von Bredow, J. D., and Condon, R. J. (1999). Antibiotic residues in milk samples obtained from cows after treatment for papillomatous digital dermatitis. *J. Am. Vet. Med. Assoc.* 215, 833–836. Available at: https://pubmed.ncbi.nlm.nih.gov/10496140/ [Accessed May 11, 2021].

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* 103, 338–348. doi:10.1016/j.ajhg.2018.07.015.

Bruijnis, M. R. N., Beerda, B., Hogeveen, H., and Stassen, E. N. (2012). Assessing the welfare

impact of foot disorders in dairy cattle by a modeling approach. *Animal* 6, 962–970. doi:10.1017/S1751731111002606.

Buch, L. H., Sørensen, A. C., Lassen, J., Berg, P., Eriksson, J. T., Jakobsen, J. H., et al. (2011). Hygiene-related and feed-related hoof diseases show different patterns of genetic correlations to clinical mastitis and female fertility. *J. Dairy Sci.* 94, 1540–1551. doi:10.3168/jds.2010-3137.

Buitenhuls, A. J., Lund, M. S., Thomasen, J. R., Thomsen, B., Nielsen, V. H., Bendixen, C., et al. (2007). Detection of quantitative trait loci affecting lameness and leg conformation traits in Danish Holstein cattle. *J. Dairy Sci.* 90, 472–481. doi:10.3168/jds.s0022-0302(07)72649-8.

Burchard, J., Gaddis, K. P., and Norman, D. (2021). CDCB leads efforts to reduce lameness with genetics. *Progress. Dairy*. Available at: https://www.progressivedairy.com/topics/a-i-breeding/cdcb-leads-efforts-to-reduce-lameness-with-genetics [Accessed May 7, 2021].

Buzanskas, M. E., Grossi, D. do A., Ventura, R. V., Schenkel, F. S., Chud, T. C. S., Stafuzza, N. B., et al. (2017). Candidate genes for male and female reproductive traits in Canchim beef cattle. *J. Anim. Sci. Biotechnol.* 8, 1–10. doi:10.1186/s40104-017-0199-8.

Capion, N., Thamsborg, S. M., and Enevoldsen, C. (2008). Conformation of hind legs and lameness in Danish Holstein heifers. *J. Dairy Sci.* 91, 2089–2097. doi:10.3168/jds.2006-457.

Cha, E., Hertl, J. A., Bar, D., and Gröhn, Y. T. (2010). The cost of different types of lameness in dairy cows calculated by dynamic programming. *Prev. Vet. Med. J.* 97, 1–8. doi:10.1016/j.prevetmed.2010.07.011.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8.

Chapinal, N., Koeck, A., Sewalem, A., Kelton, D. F., Mason, S., Cramer, G., et al. (2013). Genetic parameters for hoof lesions and their relationship with feet and leg traits in Canadian Holstein cows. *J. Dairy Sci.* 96, 2596–2604. doi:10.3168/jds.2012-6071.

Charfeddine, N., and Pérez-Cabal, M. A. (2017). Effect of claw disorders on milk production, fertility, and longevity, and their economic impact in Spanish Holstein cows. *J. Dairy Sci.* 100, 653–665. doi:10.3168/jds.2016-11434.

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., et al. (2016). Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am. J. Hum. Genet.* 98, 653–666. doi:10.1016/j.ajhg.2016.02.012.

Chen, Q., Qu, K., Ma, Z., Zhan, J., Zhang, F., Shen, J., et al. (2020). Genome-Wide Association Study Identifies Genomic Loci Associated With Neurotransmitter Concentration in Cattle. *Front. Genet.* 11, 1–10. doi:10.3389/fgene.2020.00139.

Christopoulos, D. (2017a). Introducing Unit Invariant Knee (UIK) As an Objective Choice for Elbow Point in Multivariate Data Analysis Techniques. *SSRN Electron. J.* doi:10.2139/ssrn.3043076.

Christopoulos, D. T. (2016). On the Efficient Identification of an Inflection Point On the Efficient Identification of an Inflection Point. *Int. J. Math. Sci. Comput.* 6, 13–20.

Christopoulos, D. T. (2017b). Introducing Unit Invariant Knee (UIK) As an Objective Choice for Elbow Point in Multivariate Data Analysis Techniques. *SSRN Electron. J.*, 1–7. doi:10.2139/ssrn.3043076.

Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, T. J., Crooker, B. A., et al.

(2011). Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* 12, 408. doi:10.1186/1471-2164-12-408.

Collard, B. L., Boettcher, P. J., Dekkers, J. C. M., Petitclerc, D., and Schaeffer, L. R. (2000). Relationships between energy balance and health traits of dairy cattle in early lactation. *J. Dairy Sci.* 83, 2683–2690. doi:10.3168/jds.S0022-0302(00)75162-9.

Cook, N. B. (2017). A Review of the Design and Management of Footbaths for Dairy Cattle. *Vet. Clin. North Am. - Food Anim. Pract.* 33, 195–225. doi:10.1016/j.cvfa.2017.02.004.

Cramer, G., Lissemore, K. D., Guard, C. L., Leslie, K. E., and Kelton, D. F. (2008). Herd- and Cow-Level Prevalence of Foot Lesions in Ontario Dairy Cattle. *J. Dairy Sci.* 91, 3888–3895. doi:10.3168/jds.2008-1135.

Danscher, A. M., Toelboell, T. H., and Wattle, O. (2010). Biomechanics and histology of bovine claw suspensory tissue in early acute laminitis. *J. Dairy Sci.* 93, 53–62. doi:10.3168/jds.2009-2038.

DeFrain, J. M., Socha, M. T., and Tomlinson, D. J. (2013). Analysis of foot health records from 17 confinement dairies. *J. Dairy Sci.* 96, 7329–7339. doi:10.3168/jds.2012-6017.

Dhakal, K., Tiezzi, F., Clay, J. S., and Maltecca, C. (2015). Short communication: Genomic selection for hoof lesions in first-parity US Holsteins. *J. Dairy Sci.* 98, 3502–3507. doi:10.3168/jds.2014-8830.

Dolecheck, K. A., and Bewley, J. M. (2018). Animal board invited review: Dairy cow lameness expenditures, losses and total cost. *Animal* 12, 1462–1474. doi:10.1017/S1751731118000575.

Dolecheck, K. A., Overton, M. W., Mark, T. B., and Bewley, J. M. (2019). Use of a stochastic simulation model to estimate the cost per case of digital dermatitis, sole ulcer, and white line disease by parity group and incidence timing. *J. Dairy Sci.* 102, 715–730. doi:10.3168/jds.2018-14901.

Duncan, J. S., Angell, J. W., Carter, S. D., Evans, N. J., Sullivan, L. E., and Grove-White, D. H. (2014). Contagious ovine digital dermatitis: An emerging disease. *Vet. J.* 201, 265–268. doi:10.1016/j.tvjl.2014.06.007.

Eicher, S. D., Lay, D. C., Arthington, J. D., and Schutz, M. M. (2013). Effects of rubber flooring during the first 2 lactations on production, locomotion, hoof health, immune functions, and stress1. *J. Dairy Sci.* 96, 3639–3651. doi:10.3168/jds.2012-6049.

El-Shafaey, E. S., Ateya, A., Ramadan, H., Saleh, R., Elseady, Y., Abo El Fadl, E., et al. (2017). Single Nucleotide Polymorphisms in IL8 and TLR4 Genes as Candidates for Digital Dermatitis Resistance/Susceptibility in Holstein Cattle. 28, 131–137. doi:10.1080/10495398.2016.1242489.

Epperson, B., and Midla, L. (2007). Copper sulfate for footbaths - issues and alternatives. in *Tri-State Dairy Nutrition Conference* (Fort Wayne, IN, USA), 51–54. Available at: email: Epperson.1@osu.edu%5Cnhttp://tristatedairy.osu.edu%5CnFull Text: http://www.cababstractsplus.org/fts/selectfulltextdisplay.asp?pan=20073121334%5Cnhttp://search.ebscohost.com/login.aspx?direct=true&db=lah&AN=20073121334&site=ehost-live.

Ertze, R. A., Read, D. H., Hird, D. W., and Berry, S. L. (2006). Field evaluation of prophylactic and therapeutic effects of a vaccine against (Papillomatous) digital dermatitis in dairy cattle on two California dairies. *Bov. Pract.* 40, 76–82.

Esposito, F., Sorosina, M., Ottoboni, L., Lim, E. T., Replogle, J. M., Raj, T., et al. (2015). A pharmacogenetic study implicates SLC9a9 in multiple sclerosis disease activity. *Ann.*

Neurol. 78, 115–127. doi:10.1002/ana.24429.
Evans, N. J., Brown, J. M., Demirkan, I., Singh, P., Getty, B., Timofte, D., et al. (2009). Association of unique, isolated treponemes with bovine digital dermatitis lesions. *J. Clin. Microbiol.* 47, 689–696. doi:10.1128/JCM.01914-08.
Evans, N. J., Brown, J. M., Hartley, C., Smith, R. F., and Carter, S. D. (2012). Antimicrobial susceptibility testing of bovine digital dermatitis treponemes identifies macrolides for in vivo efficacy testing. *Vet. Microbiol.* 160, 496–500. doi:10.1016/j.vetmic.2012.06.001.
Evans, N. J., Murray, R. D., and Carter, S. D. (2016). Bovine digital dermatitis: Current concepts from laboratory to farm. *Vet J* 211, 3–13. doi:10.1016/j.tvjl.2015.10.028.
Fang, L., Cai, W., Liu, S., Canela-Xandri, O., Gao, Y., Jiang, J., et al. (2020). Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *Genome Res.* 30, 790–801. doi:10.1101/gr.250704.119.
Fjeldaas, T., Knappe-Poindecker, M., Bøe, K. E., and Larssen, R. B. (2014). Water footbath, automatic flushing, and disinfection to improve the health of bovine feet. *J. Dairy Sci.* 97, 2835–2846. doi:10.3168/jds.2013-7531.
Fjeldaas, T., Sogstad, Å. M., and Østerås, O. (2011). Locomotion and claw disorders in Norwegian dairy cows housed in freestalls with slatted concrete, solid concrete, or solid rubber flooring in the alleys. *J. Dairy Sci.* 94, 1243–1255. doi:10.3168/jds.2010-3173.
Functional Annotation of Animal Genomes (FAANG) Consortium (2019). FAANGMine. Available at: http://128.206.116.18:8080/faangmine/begin.do [Accessed August 3, 2020].
Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019a). Visualization in Bayesian workflow. *J. R. Stat. Soc. Ser. A* 182, 389–402. doi:10.1111/rssa.12378.
Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019b). Visualization in Bayesian workflow Jonah. *J. R. Stat. Soc.* 182, 389–402. doi:https://doi.org/10.1111/rssa.12378.
Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., et al. (2020). R Package ARM: Data Analysis Using Regression and Multilevel/Hierarchical Models. Available at: https://cran.r-project.org/package=arm.
Gernand, E., Döhne, D. A., and König, S. (2013). Genetic background of claw disorders in the course of lactation and their relationships with type traits. *J. Anim. Breed. Genet.* 130, 435–444. doi:10.1111/jbg.12046.
Gernand, E., Rehbein, P., von Borstel, U. U., and König, S. (2012). Incidences of and genetic parameters for mastitis, claw disorders, and common health traits recorded in dairy cattle contract herds. *J. Dairy Sci.* 95, 2144–2156. doi:10.3168/jds.2011-4812.
Goddard, M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi:10.1007/s10709-008-9308-0.
Goldstein, B. A., Hubbard, A. E., Cutler, A., and Barcellos, L. F. (2010). An application of Random Forests to a genome-wide association dataset: Methodological considerations and new findings. *BMC Genet.* 11, 49. doi:10.1186/1471-2156-11-49.
González-Ruiz, S., Strillacci, M. G., Durán-Aguilar, M., Cantó-Alarcón, G. J., Herrera-Rodríguez, S. E., Bagnato, A., et al. (2019). Genome-wide association study in mexican holstein cattle reveals novel quantitative trait loci regions and confirms mapped loci for resistance to bovine tuberculosis. *Animals* 9, 1–15. doi:10.3390/ani9090636.
Goodrich, B., Gabry, J., Ali, I., and Brilleman, S. (2020). rstanarm: Bayesian applied regression modeling via Stan. Available at: https://mc-stan.org/rstanarm/.
Green, L. E., Borkert, J., Monti, G., and Tadich, N. (2010). Associations between lesion-specific

lameness and the milk yield of 1,635 dairy cows from seven herds in the Xth region of Chile and implications for management of lame dairy cows worldwide. *Anim. Welf.* 19, 419–427.
Green, L. E., Hedges, V. J., Schukken, Y. H., Blowey, R. W., and Packington, A. J. (2002). The impact of clinical lameness on the milk yield of dairy cows. *J. Dairy Sci.* 85, 2250–2256.
Green, L. E., Huxley, J. N., Banks, C., and Green, M. J. (2014). Temporal associations between low body condition, lameness and milk yield in a UK dairy herd. *Prev. Vet. Med.* 113, 63–71. doi:10.1016/j.prevetmed.2013.10.009.
Greenough, P. R. (2007). *Bovine Laminitis and Lameness*. 1st ed. , eds. C. Bergsten, A. Brizzir, and C. K. W. Mülling Saskatoon, Canada: Elsevier Ltd doi:10.1016/B978-0-7020-2780-2.X5001-0.
Groenevelt, M. (2017). Foot lameness in goats: like sheep, like cattle or completely different? *Livestock* 22, 98–101. doi:10.12968/live.2017.22.2.98.
Groenevelt, M., Anzuino, K., Langton, D. A., and Grogono-Thomas, R. (2015). Association of treponeme species with atypical foot lesions in goats. *Vet. Rec.* 176, 626. doi:10.1136/vr.102907.
Hadley, G. L., Wolf, C. A., and Harsh, S. B. (2006). Dairy cattle culling patterns, explanations, and implications. *J. Dairy Sci.* 89, 2286–2296.
Häggman, J., and Juga, J. (2013). Genetic parameters for hoof disorders and feet and leg conformation traits in Finnish Holstein cows. *J. Dairy Sci.* 96, 3319–3325. doi:10.3168/jds.2012-6334.
Häggman, J., Juga, J., Sillanpää, M. J., and Thompson, R. (2013). Genetic parameters for claw health and feet and leg conformation traits in Finnish Ayrshire cows. *J. Anim. Breed. Genet.* 130, 89–97. doi:10.1111/j.1439-0388.2012.01007.x.
Hassall, S. A., Ward, W. R., and Murray, R. D. (1993). Effects of lameness on the behaviour of cows during the summer. *Vet. Rec.* 132, 578–580. doi:10.1136/vr.132.23.578.
HAUSA (2017). Updates to the Total Performance Index ® (TPI ®) and Type Composites. Available at: www.holsteinusa.com.
Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res. (Camb).* 91, 47–60. doi:10.1017/S0016672308009981.
Hernandez, J. A., Garbarino, E. J., Shearer, J. K., Risco, C. A., and Thatcher, W. W. (2005). Comparison of milk yield in dairy cows with different degrees of lameness. *J. Am. Vet. Med. Assoc.* 227, 1292–1296. doi:10.2460/javma.2005.227.1292.
Holzhauer, M. (2017). Foot bathing dairy cows with antibiotics. *Vet. Rec.* 181, 270. doi:10.1136/vr.j4129.
Holzhauer, M., Bartels, C. J., Bergsten, C., van Riet, M. M. J., Frankena, K., and Lam, T. J. G. M. (2012). The effect of an acidified, ionized copper sulphate solution on digital dermatitis in dairy cows. *Vet. J.* 193, 659–663. doi:10.1016/j.tvjl.2012.06.049.
Holzhauer, M., Hardenberg, C., and Bartels, C. J. M. (2008). Herd and cow-level prevalence of sole ulcers in The Netherlands and associated-risk factors. *Prev. Vet. Med.* 85, 125–135.
Holzhauer, M., Hardenberg, C., Bartels, C. J. M., and Frankena, K. (2006). Herd-and cow-level prevalence of digital dermatitis in the Netherlands and associated risk factors. *J. Dairy Sci.* 89, 580–588. Available at: https://www.journalofdairyscience.org/article/S0022-0302(06)72121-X/fulltext.
Hu, Z. L., Park, C. A., and Reecy, J. M. (2019). Building a livestock genetic and genomic

information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 47, D701–D710. doi:10.1093/nar/gky1084.
Hwangbo, D. S., Biteau, B., Rath, S., Kim, J., and Jasper, H. (2016). Control of apoptosis by Drosophila DCAF12. *Dev. Biol.* 413, 50–59. doi:10.1016/j.ydbio.2016.03.003.
Klawitter, M., Braden, T. B., and Müller, K. E. (2019). Randomized clinical trial evaluating the effect of bandaging on the healing of sole ulcers in dairy cattle. *Vet. Anim. Sci.* 8, 100070. doi:10.1016/j.vas.2019.100070.
Kleinhenz, K. E., Plummer, P. J., Danielson, J., Burzette, R. G., Gorden, P. J., Coetzee, J., et al. (2014). Survey of veterinarians and hoof trimmers on methods applied to treat claw lesions in dairy cattle. *Bov. Pract.* 48, 47–52. Available at: http://www.aabp.org.
Knappe-Poindecker, M., Gilhuus, M., Jensen, T. K., Klitgaard, K., Larssen, R. B., and Fjeldaas, T. (2013). Interdigital dermatitis, heel horn erosion, and digital dermatitis in 14 Norwegian dairy herds. *J. Dairy Sci.* 96, 7617–7629. doi:10.3168/jds.2013-6717.
Koenig, S., Sharifi, A. R., Wentrot, H., Landmann, D., Eise, M., and Simianer, H. (2005). Genetic parameters of claw and foot disorders estimated with logistic models. *J. Dairy Sci.* 88, 3316–3325. doi:10.3168/jds.S0022-0302(05)73015-0.
Koenig, S., Wu, X. L., Gianola, D., Heringstad, B., and Simianer, H. (2008). Exploration of relationships between claw disorders and milk yield in Holstein cows via recursive linear and threshold models. *J. Dairy Sci.* 91, 395–406. doi:10.3168/jds.2007-0170.
Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., and Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* 44, 1066–1071. doi:10.1038/ng.2376.
Koufariotis, L., Chen, Y. P. P., Bolormaa, S., and Hayes, B. J. (2014). Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. *BMC Genomics* 15. doi:10.1186/1471-2164-15-436.
Koufariotis, L. T., Chen, Y. P. P., Stothard, P., and Hayes, B. J. (2018). Variance explained by whole genome sequence variants in coding and regulatory genome annotations for six dairy traits. *BMC Genomics* 19. doi:10.1186/s12864-018-4617-x.
Krull, A. C., Shearer, J. K., Gorden, P. J., Scott, H. M., and Plummer, P. J. (2016). Digital dermatitis: Natural lesion progression and regression in Holstein dairy cattle over 3 years. *J. Dairy Sci.* 99, 3718–3731. doi:10.3168/jds.2015-10535.
Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi:10.18637/jss.v028.i05.
Kurz, J. P., Yang, Z., Weiss, R. B., Wilson, D. J., Rood, K. A., Liu, G. E., et al. (2019). A genome-wide association study for mastitis resistance in phenotypically well-characterized Holstein dairy cattle using a selective genotyping approach. *Immunogenetics* 71, 35–47. doi:10.1007/s00251-018-1088-9.
Lai, E., Danner, A. L., Famula, T. R., and Oberbauer, A. (2021). Genome-wide association studies reveal susceptibility loci for noninfectious claw lesions in Holstein dairy cattle. *Front. Genet.* 12, 728. doi:10.3389/FGENE.2021.657375.
Lai, E., Danner, A. L., Famula, T. R., and Oberbauer, A. M. (2020). Genome-Wide Association Studies Reveal Susceptibility Loci for Digital Dermatitis in Holstein Cattle. *Animals* 10, 2009. doi:10.3390/ani10112009.
Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11, 241–247. doi:10.1038/ng1195-241.

Larsson, C., Ali, M. A., Pandzic, T., Lindroth, A. M., He, L., and Sjöblom, T. (2017). Loss of DIP2C in RKO cells stimulates changes in DNA methylation and epithelial-mesenchymal transition. *BMC Cancer* 17. doi:10.1186/s12885-017-3472-5.

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., and Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 28, 2540–2542. doi:10.1093/bioinformatics/bts474.

Lehner, S., Zerbin, I., Doll, K., Rehage, J., and Distl, O. (2018). A genome-wide association study for left-sided displacement of the abomasum using a high-density single nucleotide polymorphism array. *J. Dairy Sci.* 101, 1258–1266. doi:10.3168/jds.2017-13216.

Li, M.-X., Yeung, J. M. Y., Cherny, S. S., and Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131, 747–756. doi:10.1007/s00439-011-1118-2.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8, 833–835. doi:10.1038/nmeth.1681.

Lischer, C. J., Ossent, P., Räber, M., and Geyer, H. (2002). Suspensory structures and supporting tissues of the third phalanx of cows and their relevance to the development of typical sole ulcers (Rusterholz ulcers). *Vet. Rec.* 151, 694–698. doi:10.1136/vr.151.23.694.

Listgarten, J., Lippert, C., Kang, E. Y., Xiang, J., Kadie, C. M., and Heckerman, D. (2013). A powerful and efficient set test for genetic markers that handles confounders. *Bioinformatics* 29, 1526–1533. doi:10.1093/bioinformatics/btt177.

Maddirevula, S., Alsahli, S., Alhabeeb, L., Patel, N., Alzahrani, F., Shamseldin, H. E., et al. (2018). Expanding the phenome and variome of skeletal dysplasia. *Genet. Med.* 20, 1609–1616. doi:10.1038/gim.2018.50.

Magrin, L., Brscic, M., Armato, L., Contiero, B., Cozzi, G., and Gottardo, F. (2018). An overview of claw disorders at slaughter in finishing beef cattle reared in intensive indoor systems through a cross-sectional study. *Prev. Vet. Med.* 161, 83–89. doi:10.1016/j.prevetmed.2018.10.018.

Makanjuola, B. O., Miglior, F., Abdalla, E. A., Maltecca, C., Schenkel, F. S., and Baes, C. F. (2020). Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *J. Dairy Sci.* 103, 5183–5199. doi:10.3168/jds.2019-18013.

Malchiodi, F., Koeck, A., Chapinal, N., Sargolzaei, M., Fleming, A., Kelton, D. F., et al. (2015a). Genetic Analyses of Hoof Lesions in Canadian Holsteins Using an Alternative Contemporary Group.

Malchiodi, F., Koeck, A., Christen, A. M., Schenkel, F. S., Kelton, D. F., and Miglior, F. (2015b). Genetic parameters and genome wide association study of individual hoof lesions in Canadian Holsteins using different contemporary groups.

Malchiodi, F., Koeck, A., Mason, S., Christen, A. M., Kelton, D. F., Schenkel, F. S., et al. (2017). Genetic parameters for hoof health traits estimated with linear and threshold models using alternative cohorts. *J. Dairy Sci.* 100, 2828–2836. doi:10.3168/jds.2016-11558.

Moore, C. M., Jacobson, S. A., and Fingerlin, T. E. (2019). Power and Sample Size Calculations for Genetic Association Studies in the Presence of Genetic Model Misspecification. *Hum. Hered.* 84, 256–271. doi:10.1159/000508558.

Mostert, P. F., van Middelaar, C. E., de Boer, I. J. M., and Bokkers, E. A. M. (2018). The impact of foot lesions in dairy cows on greenhouse gas emissions of milk production. *Agric. Syst.* 167, 206–212. doi:10.1016/j.agsy.2018.09.006.

Murphy, K., and Weaver, C. (2016). *Janeway's Immunobiology*. 9th ed. New York, USA.

Murray, R. D., Downham, D. Y., Demirkan, I., and Carter, S. D. (2002). Some relationships between spirochaete infections and digital dermatitis in four UK dairy herds. *Res. Vet. Sci.* 73, 223–230. doi:10.1016/S0034-5288(02)00027-9.

Naderi, S., Bohlouli, M., Yin, T., and König, S. (2018). Genomic breeding values, SNP effects and gene identification for disease traits in cow training sets. *Anim. Genet.* 49, 178–192. doi:10.1111/age.12661.

Nally, J., Wilson-Welder, J., and Alt, D. (2015). The etiology of digital dermatitis in ruminants: recent perspectives. *Vet. Med. Res. Reports*, 155. doi:10.2147/vmrr.s62072.

National Research Council (2001). *Nutrient Requirements of Dairy Cattle*. 7th ed. Washington, D.C.: National Academies Press doi:10.17226/9825.

Network, C. D. (2016). Improving Existing Traits and Adding Exciting New Ones. Available at: https://www.cdn.ca/Articles/GEBAPR2016/10_Vision of GE Services - ENGLISH.pdf.

Neupane, M., Kiser, J. N., and Neibergs, H. L. (2018). Gene set enrichment analysis of SNP data in dairy and beef cattle with bovine respiratory disease. *Anim. Genet.* 49, 527–538. doi:10.1111/age.12718.

Newsome, R. F., Green, M. J., Bell, N. J., Bollard, N. J., Mason, C. S., Whay, H. R., et al. (2017a). A prospective cohort study of digital cushion and corium thickness. Part 1: Associations with body condition, lesion incidence, and proximity to calving. *J. Dairy Sci.* 100, 4745–4758. doi:10.3168/jds.2016-12012.

Newsome, R. F., Green, M. J., Bell, N. J., Bollard, N. J., Mason, C. S., Whay, H. R., et al. (2017b). A prospective cohort study of digital cushion and corium thickness. Part 2: Does thinning of the digital cushion and corium lead to lameness and claw horn disruption lesions? *J. Dairy Sci.* 100, 4759–4771. doi:10.3168/jds.2016-12013.

Newsome, R. F., Green, M. J., Bell, N. J., Chagunda, M. G. G., Mason, C. S., Rutland, C. S., et al. (2016). Linking bone development on the caudal aspect of the distal phalanx with lameness during life. *J. Dairy Sci.* 99, 4512–4525. doi:10.3168/jds.2015-10202.

Nocek, J. E. (1997). Bovine Acidosis: Implications on Laminitis. *J. Dairy Sci.* 80, 1005–1028. doi:10.3168/jds.S0022-0302(97)76026-0.

Nordlund, K. V., Cook, N. B., and Oetzel, G. R. (2004). Investigation strategies for laminitis problem herds. *J. Dairy Sci.* 87, E27–E35. doi:10.3168/jds.S0022-0302(04)70058-2.

NTP (2016). Report on Carcinogens. Research Triangle Park, NC doi:10.1163/9789004286801_013.

O'Driscoll, K., McCabe, M., and Earley, B. (2015). Differences in leukocyte profile, gene expression, and metabolite status of dairy cows with or without sole ulcers. *J. Dairy Sci.* 98, 1685–1695. doi:10.3168/jds.2014-8199.

Oberbauer, A. M., Berry, S. L., Belanger, J. M., McGoldrick, R. M., Pinos-Rodriquez, J. M., and Famula, T. R. (2013). Determining the heritable component of dairy cattle foot lesions. *J. Dairy Sci.* 96, 605–613. doi:10.3168/jds.2012-5485.

Onyiro, O. M., Andrews, L. J., and Brotherstone, S. (2008). Genetic Parameters for Digital Dermatitis and Correlations with Locomotion, Production, Fertility Traits, and Longevity in Holstein-Friesian Dairy Cows. *J. Dairy Sci.* 91, 4037–4046. doi:10.3168/jds.2008-1190.

Pérez-Cabal, M. A., and Charfeddine, N. (2015). Models for genetic evaluations of claw health

traits in Spanish dairy cattle. *J. Dairy Sci.* 98, 8186–8194. doi:10.3168/jds.2015-9562.
Pérez-Cabal, M. A., and Charfeddine, N. (2016). Short communication: Association of foot and leg conformation and body weight with claw disorders in Spanish Holstein cows. *J. Dairy Sci.* 99, 9104–9108. doi:10.3168/jds.2016-11331.
Pérez O'Brien, A. M., Mészáros, G., Utsunomiya, Y. T., Sonstegard, T. S., Garcia, J. F., Van Tassell, C. P., et al. (2014). Linkage disequilibrium levels in Bos indicus and Bos taurus cattle using medium and high density SNP chip data and different minor allele frequency distributions. *Livest. Sci.* 166, 121–132. doi:10.1016/j.livsci.2014.05.007.
Pryce, J. E., Veerkamp, R. F., Thompson, R., Hill, W. G., and Simm, G. (1997). Genetic aspects of common health disorders and measures of fertility in Holstein Friesian dairy cattle. *Anim. Sci.* 65, 353–360. doi:10.1017/S1357729800008559.
Purcell, S., and Chang, C. (2021). PLINK 2.0. Available at: https://www.cog-genomics.org/plink/2.0/.
Purcell, S. M., and Chang, C. C. (2015). PLINK 1.9. Available at: https://www.cog-genomics.org/plink2.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
Pyman, M. F. S. (1997). Comparison of bandaging and elevation of the claw for the treatment of foot lameness in dairy cows. *Aust. Vet. J.* 75, 132–135. doi:10.1111/j.1751-0813.1997.tb14173.x.
R Development Core Team (2010). R: A language and environment for statistical computing. Available at: https://www.r-project.org/.
Ratwan, P., Mandal, A., Kumar, M., Karunakaran, M., and Dutta, T. K. (2015). Carbon Footprints of Dairy Animals: Causes and Its Mitigation. *Res. Rev. J. Dairy Sci. Technol.* 4, 6–11.
Read, D. H., and Walker, R. L. (1998). Papillomatous digital dermatitis (footwarts) in California dairy cattle: clinical and gross pathologic findings. *J. Vet. Diagnostic Investig.* 10, 67–76.
Refaai, W., Ducatelle, R., Geldhof, P., Mihi, B., El-shair, M., and Opsomer, G. (2013). Digital dermatitis in cattle is associated with an excessive innate immune response triggered by the keratinocytes. *BMC Vet. Res.* 9, 193. doi:10.1186/1746-6148-9-193.
Richardson, I. W., Berry, D. P., Wiencko, H. L., Higgins, I. M., More, S. J., McClure, J., et al. (2016). A genome-wide association study for genetic susceptibility to Mycobacterium bovis infection in dairy cattle identifies a susceptibility QTL on chromosome 23. *Genet. Sel. Evol.* 48, 19. doi:10.1186/s12711-016-0197-x.
Ring, S. C., Twomey, A. J., Byrne, N., Kelleher, M. M., Pabiou, T., Doherty, M. L., et al. (2018). Genetic selection for hoof health traits and cow mobility scores can accelerate the rate of genetic gain in producer-scored lameness in dairy cows. *J. Dairy Sci.* 101, 10034–10047. doi:10.3168/jds.2018-15009.
Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., et al. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* 9, 1–9. doi:10.1093/gigascience/giaa021.
Rusterholz, A. (1920). Das spezifisch-traumatische Klauensohlengeschwür des Rindes (The specific traumatic sole ulcer of claws in cattle). *Schweiz. Arch. Tierheilkd.* 62, 505–525.
Sadiq, M. B., Ramanoon, S. Z., Mossadeq, W. M. S., Mansor, R., and Syed-Hussain, S. S. (2017). Association between lameness and indicators of dairy cow welfare based on

locomotion scoring, body and hock condition, leg hygiene and lying behavior. *Animals* 7. doi:10.3390/ani7110079.
Sánchez-Molano, E., Bay, V., Smith, R. F., Oikonomou, G., and Banos, G. (2019). Quantitative Trait Loci Mapping for Lameness Associated Phenotypes in Holstein–Friesian Dairy Cattle. *Front. Genet.* 10. doi:10.3389/fgene.2019.00926.
Sanchez-Pulido, L., and Ponting, C. P. (2018). TMEM132: An ancient architecture of cohesin and immunoglobulin domains define a new family of neural adhesion molecules. *Bioinformatics* 34, 721–724. doi:10.1093/bioinformatics/btx689.
Scholey, R. A., Blowey, R. W., Murray, R. D., Smith, R. F., Cameron, J., Massey, J. P., et al. (2012). Investigating host genetic factors in bovine digital dermatitis. *Vet. Rec.* 171, 624. doi:10.1136/vr.101251.
Scholey, R. A., Evans, N. J., Blowey, R. W., Massey, J. P., Murray, R. D., Smith, R. F., et al. (2013). Identifying host pathogenic pathways in bovine digital dermatitis by RNA-Seq analysis. *Vet. J.* 197, 699–706. doi:10.1016/j.tvjl.2013.03.008.
Shabalina, T., Yin, T., and König, S. (2020). Influence of common health disorders on the length of productive life and stayability in German Holstein cows. *J. Dairy Sci.* 103, 583–596. doi:10.3168/jds.2019-16985.
Shearer, J. K., Plummer, P., and Schleining, J. (2015). Perspectives on the treatment of claw lesions in cattle. *Vet. Med. Res. Reports*, 273. doi:10.2147/vmrr.s62071.
Shearer, J. K., and van Amstel, S. R. (2001). Functional and corrective claw trimming. *Vet. Clin. North Am. - Food Anim. Pract.* 17, 53–72. doi:10.1016/S0749-0720(15)30054-2.
Shearer, J. K., and van Amstel, S. R. (2017). Pathogenesis and Treatment of Sole Ulcers and White Line Disease. *Vet. Clin. North Am. - Food Anim. Pract.* 33, 283–300. doi:10.1016/j.cvfa.2017.03.001.
Smith, C. L., and Eppig, J. T. (2009). The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 1, 390–399. doi:10.1002/wsbm.44.
Speed, D., Hemani, G., Johnson, M. R., and Balding, D. J. (2012). Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* 91, 1011–1021. doi:10.1016/j.ajhg.2012.10.010.
Stambuk, C. R., McArt, J. A. A., Bicalho, R. C., Miles, A. M., and Huson, H. J. (2019). A longitudinal study of digital cushion thickness and its function as a predictor for compromised locomotion and hoof lesions in Holstein cows. *Transl. Anim. Sci.* 3, 74–83. doi:10.1093/tas/txy107.
Stambuk, C. R., Staiger, E. A., Nazari-Ghadikolaei, A., Heins, B. J., and Huson, H. J. (2020). Phenotypic characterization and genome-wide association studies of digital cushion thickness in Holstein cows. *J. Dairy Sci.* 103, 3289–3303. doi:10.3168/jds.2019-17409.
Stephens, M. (2013). A Unified Framework for Association Analysis with Multiple Related Phenotypes. *PLoS One* 8. doi:10.1371/journal.pone.0065245.
Stone, W. C. (2004). Nutritional approaches to minimize subacute ruminal acidosis and laminitis in dairy cattle. *J. Dairy Sci.* 87, E13–E26. doi:10.3168/jds.S0022-0302(04)70057-0.
Suchocki, T., Egger-Danner, C., Schwarzenbacher, H., and Szyda, J. (2020). Two-stage genome-wide association study for the identification of causal variants underlying hoof disorders in cattle. *J. Dairy Sci.* 103, 4483–4494. doi:10.3168/jds.2019-17542.
Sullivan, L. E., Evans, N. J., Blowey, R. W., Grove-White, D. H., Clegg, S. R., Duncan, J. S., et al. (2015a). A molecular epidemiology of treponemes in beef cattle digital dermatitis lesions

and comparative analyses with sheep contagious ovine digital dermatitis and dairy cattle digital dermatitis lesions. *Vet. Microbiol.* 178, 77–87. doi:10.1016/j.vetmic.2015.04.011.
Sullivan, L. E., Evans, N. J., Clegg, S. R., Carter, S. D., Horsfield, J. E., Grove-White, D., et al. (2015b). Digital dermatitis treponemes associated with a severe foot disease in dairy goats. *Vet. Rec.* 176, 283. doi:10.1136/vr.102858.
Thoefner, M. B., Pollitt, C. C., Van Eps, A. W., Milinovich, G. J., Trott, D. J., Wattle, O., et al. (2004). Acute bovine laminitis: A new induction model using alimentary oligofructose overload. *J. Dairy Sci.* 87, 2932–2940. doi:10.3168/jds.S0022-0302(04)73424-4.
Thompson-Crispi, K. A., Hine, B., Quinton, M., Miglior, F., and Mallard, B. A. (2012). Short communication: Association of disease incidence and adaptive immune response in Holstein dairy cows. *J. Dairy Sci.* 95, 3888–3893. doi:10.3168/jds.2011-5201.
Trott, D. J., Moeller, M. R., Zuerner, R. L., Goff, J. P., Waters, W. R., Alt, D. P., et al. (2003). Characterization of Treponema phagedenis-Like Spirochetes Isolated from Papillomatous Digital Dermatitis Lesions in Dairy Cattle. *J. Clin. Microbiol.* 41, 2522–2529.
Tsuka, T., Ooshita, K., Sugiyama, A., Osaki, T., Okamoto, Y., Minami, S., et al. (2012). Quantitative evaluation of bone development of the distal phalanx of the cow hind limb using computed tomography. *J. Dairy Sci.* 95, 127–138. doi:10.3168/jds.2011-4316.
Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *bioRxiv*, 5165. doi:https://doi.org/10.1101/005165.
Twomey, A. J., Berry, D. P., Evans, R. D., Doherty, M. L., Graham, D. A., and Purfield, D. C. (2019). Genome-wide association study of endo-parasite phenotypes using imputed whole-genome sequence data in dairy and beef cattle. *Genet. Sel. Evol.* 51, 1–17. doi:10.1186/s12711-019-0457-7.
USDA (2016). Dairy 2014 - Dairy Cattle Management Practices in the United States, 2014. Fort Collins, CO Available at: http://www.aphis.usda.gov/nahms.
USDA (2018). Dairy 2014, - Health and Management Practices on U.S. Dairy Operations, 2014. Fort Collins, CO: USDA–APHIS–VS–CEAH–NAHMS.
van der Linde, C., de Jong, G., Koenen, E. P. C., and Eding, H. (2010). Claw health index for Dutch dairy cattle based on claw trimming and conformation data. *J. Dairy Sci.* 93, 4883–4891. doi:10.3168/jds.2010-3183.
van der Spek, D., van Arendonk, J. A. M., and Bovenhuis, H. (2015a). Genetic relationships between claw health traits of dairy cows in different parities, lactation stages, and herds with different claw disorder frequencies. *J. Dairy Sci.* 98, 6564–6571. doi:10.3168/jds.2015-9561.
van der Spek, D., van Arendonk, J. A. M., and Bovenhuis, H. (2015b). Genome-wide association study for claw disorders and trimming status in dairy cattle. *J. Dairy Sci.* 98, 1286–1295. doi:10.3168/jds.2014-8302.
van der Spek, D., van Arendonk, J. A. M., Vallée, A. A. A., and Bovenhuis, H. (2013). Genetic parameters for claw disorders and the effect of preselecting cows for trimming. *J. Dairy Sci.* 96, 6070–6078. doi:10.3168/jds.2013-6833.
Van der Waaij, E. H., Holzhauer, M., Ellen, E., Kamphuis, C., and De Jong, G. (2005). Genetic parameters for claw disorders in Dutch dairy cattle and correlations with conformation traits. *J. Dairy Sci.* 88, 3672–3678.
Van Nuffel, A., Zwertvaegher, I., Pluym, L., Van Weyenberg, S., Thorup, V. M., Pastell, M., et al. (2015). Lameness Detection in Dairy Cows: Part 1. How to Distinguish between Non-Lame and Lame Cows Based on Differences in Locomotion or Behavior. *Animals* 5, 838–

860. doi:10.3390/ani5030387.
Van Raden, P. M., Cole, J. ., and Parker Gaddis, K. L. (2021). Net merit as a measure of lifetime profit: 2021 revision. Available at: https://www.ars.usda.gov/research/publications/publication/?seqNo115=353707.
Vanegas, J., Overton, M., Berry, S. L., and Sischo, W. M. (2006). Effect of rubber flooring on claw health in lactating dairy cows housed in free-stall barns. *J. Dairy Sci.* 89, 4251–4258. doi:10.3168/jds.S0022-0302(06)72471-7.
Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Andrew, Y., Bürkner, P.-C., et al. (2020). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. Available at: https://mc-stan.org/loo/.
Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27, 1413–1432. doi:10.1007/s11222-016-9696-4.
Visscher, P. M., Hemani, G., Vinkhuyzen, A. A. E., Chen, G. B., Lee, S. H., Wray, N. R., et al. (2014). Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *PLoS Genet.* 10. doi:10.1371/journal.pgen.1004269.
Vukasinovic, N., Bacciu, N., Przybyla, C. A. A., Boddhireddy, P., and DeNise, S. K. K. (2017). Development of genetic and genomic evaluation for wellness traits in US Holstein cows. *J. Dairy Sci.* 100, 428–438. doi:10.3168/jds.2016-11520.
Watts, K. M., Fodor, C., Beninger, C., Lahiri, P., Arrazuria, R., Buck, J. De, et al. (2018). A Differential Innate Immune Response in Active and Chronic Stages of Bovine Infectious Digital Dermatitis. *Front. Microbiol.* 9, 1586. doi:10.3389/fmicb.2018.01586.
White, M. E., Glickman, L. T., Embree, I. C., Powers, P. M., and Pearson, E. G. (1981). A randomized trial for evaluation of bandaging sole abscesses in cattle. *J. Am. Vet. Med. Assoc.* 178, 375–377. Available at: https://europepmc.org/article/med/7016819 [Accessed May 10, 2021].
Winters, M. (2018). Lameness advantage fact sheet. Available at: https://projectblue.blob.core.windows.net/media/Default/Imported Publication Docs/Lameness advantage genetic index factsheet.pdf.
Wu, X., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2016). Association analysis for feet and legs disorders with whole-genome sequence variants in 3 dairy cattle breeds. *J Dairy Sci* 99, 7221–7231. doi:10.3168/jds.2015-10705.
Xia, J., Fan, H., Chang, T., Xu, L., Zhang, W., Song, Y., et al. (2017). Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle. *Sci. Rep.* 7, 1–9. doi:10.1038/srep42048.
Xiang, R., MacLeod, I. M., Bolormaa, S., and Goddard, M. E. (2017). Genome-wide comparative analyses of correlated and uncorrelated phenotypes identify major pleiotropic variants in dairy cattle. *Sci. Rep.* 7, 1–12. doi:10.1038/s41598-017-09788-9.
Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82.
Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824. Available at: http://www.nature.com/ng/journal/v44/n7/pdf/ng.2310.pdf.
Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome- wide association studies. *Nat. Methods* 11, 407–409. doi:10.1038/nmeth.2848.
Zuerner, R. L., Heidari, M., Elliott, M. K., Alt, D. P., and Neill, J. D. (2007). Papillomatous

digital dermatitis spirochetes suppress the bovine macrophage innate immune response. *Vet. Microbiol.* 125, 256–264. doi:10.1016/j.vetmic.2007.06.001.