2024-01-10

Genetics and genomics of bovine mastitis resistance

Gunasegaram Narayana, Saranya

UNIVERSITY OF CALGARY

Genetics and genomics of bovine mastitis resistance

by

Saranya Gunasegaram Narayana

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN VETERINARY MEDICAL SCIENCES

CALGARY, ALBERTA

JANUARY, 2024

ABSTRACT

Mastitis, a multifactorial infectious disease primarily caused by bacteria, remains a major economic challenge for the dairy cattle industry, despite decades of extensive research. To combat mastitis effectively, breeding for enhanced mastitis resistance through genetic selection, in combination with improved management strategies, offers a promising alternative. This thesis delves into the genetic and genomic aspects of mastitis, exploring the following key areas: **1) Genetic analysis of subclinical mastitis (SCM)**: By employing both linear and threshold models, the research presented in **Chapter 2** investigated the genetic basis of SCM in early lactating heifers. Despite the relatively low heritability for six different SCM traits, this analysis revealed a notable genetic variation among sires, suggesting the possibility of long-term genetic selection to enhance heifer mastitis resistance. **2) Genetics of pathogen-specific intramammary infections (IMI)**: In **Chapter 3**, the genetic variation of overall and pathogen-specific IMI in nonclinical primiparous and multiparous cows was explored. The results demonstrated substantial exploitable genetic variability among daughters of sires for overall IMI and six IMI pathogen group traits. Furthermore, results indicated that the selection for lower somatic cell score (SCS) may not uniformly improve resistance to IMI from every pathogen group. **3) Underlying genetic architecture of mastitis resistance**: Through a systematic review of genome-wide association studies and gene prioritization analysis, the research presented in **Chapter 4** aimed to decipher the genetic architecture of mastitis resistance in dairy cattle. This comprehensive study identified numerous unique genes and quantitative trait loci, emphasizing the polygenic nature of mastitis, with many of the prioritized genes linked to immune response. **4) Genomic Copy Number Variants (CNVs) and SCS**: The investigation in **Chapter 5** uncovered three significant CNVs

associated with SCS in Holstein cows, including four novel findings. Annotation analysis again identified numerous genes linked to immune response. **5) Genomic predictions of mastitis-related traits**: **Chapter 6** explored genomic prediction of mastitis-related traits in early first lactation of dairy cows, utilizing a single-step genomic approach. This approach demonstrated a substantial increase in validation and theoretical reliabilities for genomic prediction of genomic estimated breeding value of young animals compared to traditional pedigree-based prediction. In summary, this thesis highlights the genetic potential for enhancing mastitis resistance in dairy cattle, shedding light on various aspects of genetic and genomic influences on mastitis-associated traits. These findings provide valuable insights for the development of improved breeding strategies in the dairy industry.

PREFACE

This thesis consists of 5 manuscripts – 3 have been published, and 2 are currently prepared for submission. The following manuscripts are included in this thesis:

Chapter 2

Narayana, S. G., F. Miglior, S. A. Naqvi, F. Malchiodi, P. Martin, and H. W. Barkema. 2018. Genetic analysis of subclinical mastitis in early lactation of heifers using both linear and threshold models. J. Dairy Sci. 101:11120-11131. https://doi.org/10.3168/jds.2018-15126

Chapter 3

Narayana, S. G., F. Schenkel, F. Miglior, T.Chud, E. A. Abdalla, S. A. Naqvi, F. Malchiodi, and H. W. Barkema. 2021. Genetic analysis of pathogen-specific intramammary infections in dairy cows. J. Dairy Sci. 104:1982-1992. https://doi.org/10.3168/jds.2020-19062

Chapter 4

Narayana S. G, E. de Jong, F. S. Schenkel, P. A. S. Fonseca, T. C. S. Chud, D. Powel, G. Wachoski-Dark, P. E. Ronksley, F. Miglior, K. Orsel, and H. W. Barkema. 2023. Underlying genetic architecture of resistance to mastitis in dairy cattle: A systematic review and gene prioritization analysis of genome-wide association studies. J. Dairy Sci. 106:323–351. https://doi.org/10.3168/jds.2022-21923

Chapter 5

Narayana. S. G, H. W. Barkema, T. C. S. Chud, H. R. de Oliveira, P. A. S. Fonseca, G. de Oliveira Jr, A. Butty, C. Baes, F. Miglior, F. Malchiodi, and F. S. Schenkel. Association between genomic copy number variants and somatic cell score in Holsteins cows (manuscript prepared for submission to J. Dairy Sci.)

Chapter 6

Narayana. S. G, F. S. Schenkel, H. R. de Oliveira, J. Hidalgo, G. de Oliveira Jr, F. Miglior, E. Massender, and H. W. Barkema. Genomic predictions of mastitis-related traits in early lactating dairy heifers using single-step genomic approach (manuscript prepared for submission to J. Dairy Sci.).

Statement of work done

Saranya Narayana prepared and analysed the data and wrote the manuscript for Chapters 2-6. Dr. Ali Naqvi provided the data used for Chapter 2 and 3. Drs. Ali Naqvi and Diego Nobrega assisted with understanding the data for Chapter 2 and 3. Saranya Narayana performed the systematic review together with Ellen de Jong for Chapter 4. Ellen de Jong also contributed to writing the systematic review section of Chapter 4 and assisted with preparing Tables 4-1 to 4-3 and Figure 4-1. Dr. Pablo Fonseca helped with enrichment and gene prioritization analysis of Chapter 4. Dr. Adrian Butty provided the data for Chapter 5. Drs. Hinayah de Oliveira and Pablo Fonseca helped with data analysis of Chapter 5. For chapter 6, Dr. Hinayah de Oliveira assisted with data analysis. Drs. Herman Barkema and Flavio Schenkel oversaw the design and execution of all chapters. Dr. Filippo Miglior oversaw the design and implementation of Chapters 2 and 3. All authors approved the final versions to be published and agreed to be accountable for all aspects of the manuscripts. Permission has been obtained from all co-authors and the publishing journals to reprint the manuscripts in this thesis.

ACKNOWLEDGEMENTS

I owe a profound debt of gratitude to several remarkable individuals who played pivotal roles in the culmination of this thesis.

First and foremost, I extend my sincerest appreciation to Dr. Herman Barkema, my supervisor. His unwavering guidance, trust, and encouragement have been instrumental. Dr. Barkema not only kept me aligned during my PhD journey, but also provided the freedom and motivation to push beyond my perceived limits. Every interaction with Dr. Barkema was infused with a depth of understanding and trust that is rare to find. Perhaps most cherished has been Dr. Barkema's unfaltering faith in my potential. In times when I grappled with uncertainty and doubt, his reassuring presence served as a beacon, illuminating the path ahead and instilling in me the confidence to continue. Reflecting upon my journey, it is evident that Dr. Barkema's influence has been transformative, not just in the realm of academics, but in the way I approach challenges, decisions, and growth. My heartfelt gratitude to you, Dr. Barkema, for being not only an exemplary mentor, but also an integral force shaping my academic and personal evolution.

Equally paramount in this journey has been the steadfast influence and unwavering support of my co-advisor, Dr. Flavio Schenkel. I cannot overstate the significance of Dr. Schenkel's mentorship, which stands as a paragon of excellence. Through countless interactions and discussions, he has meticulously sculpted my analytical abilities, thereby amplifying my capabilities as a scientist. His availability and wisdom in crucial moments have left an indelible mark on both my academic journey and career trajectory. In gratitude, I reflect upon the countless hours Dr. Schenkel invested, the wisdom he imparted, and the confidence he instilled in me. His imprints on my academic and professional path are deep and transformative. My heartfelt thanks to you, Dr. Schenkel, for being such a monumental pillar in my academic odyssey.

I want to express my special thanks of gratitude to Dr. Filippo Miglior, my former co-advisor. From the very outset of my academic journey, Dr. Miglior was instrumental in charting the course of my professional trajectory. His vision and guidance became the bedrock upon which my research and studies were built. Dr. Miglior's genuine interest in my growth, both as a scholar and an individual, was evident in our every interaction. He was not just an advisor, but a mentor in the truest sense, continually pushing me to elevate my standards, think critically, and approach challenges with a nuanced perspective. His imprint on my work and life is profound, and I will always remain indebted for the knowledge, wisdom, and mentorship he provided. Thank you, Dr. Miglior for being such a pivotal force in my academic journey!

Having the privilege to learn from and be mentored by this triumvirate has been one of the most noteworthy advantages of my academic journey. Their collective guidance has not only honed my scientific acumen but also fostered personal growth. They have equipped me with the ability to recognize my strengths, understand my weaknesses, and learn from my mistakes. To all three, I extend my heartfelt thanks. Your influence has been transformative, and the lessons imbibed will be treasured always. Thank you!

Thank you to my special supervisory committee: Drs. Karin Orsel and John Kastelic. Thank you for your valuable feedback and commitment to my work. You both have been very supportive throughout my entire PhD. A special thank you to Dr. John Kastelic for editing all my manuscripts and improving them with your magic wand; it was a tremendous help. To both of you, a resounding thanks you for your dedication and belief in my work!

My heartfelt gratitude goes to my esteemed mentors: Drs. Hinayah Rojas, Pablo Fonseca, Tati Chud, Gerson de Oliveira, Emhimad Abdalla, Pauline Martin, and Francesca Malchiodi. Your unwavering support and sage guidance were instrumental in the successful completion of my PhD.

During challenging moments, your insights illuminated my path and enhanced the quality of my work. Your collective wisdom and mentorship have not only aided my immediate endeavors but have also inspired and shaped my growth as a researcher. A sincere thank you to each one of you for your invaluable contributions to my journey.

I want to thank all my friends from the Faculty of Veterinary Medicine, University of Calgary, and the Center for Genetic Improvement of Livestock (CGIL), University of Guelph, for their support. Special thanks to Ellen de Jong, Ali Naqvi, Diego Nobrega, Anne-Marieke Smid, Audrey Martin, Erin Massender and Shadi Nayeri. Thank you!

From the depths of my heart, I extend my profound gratitude to my family. To my husband, Sanjeewa Narayana, your unwavering presence during the highs and lows, and your invaluable assistance in computation and coding, have been pivotal in the completion of my PhD. I am deeply fortunate to share my life with you. To my parents, Mr. and Mrs. Gunasegaram, my sister Sanathani and my brother Jeygar Ramesh, your ceaseless support and encouragement from my earliest memories till now has been the backbone of my journey. My little rays of sunshine, Elina and Emilia Narayana, your endearing gestures of love – your smiles, hugs, and kisses – have been my daily source of strength and joy. Words may falter in expressing the depth of my gratitude, but I want you all to know how cherished you are. My heart is full of love and appreciation for each of you. Thank you.

The progress and completion of my work were notably facilitated by the generous support of various funding agencies. I am honored to have been awarded Alberta Innovates Graduate Student Scholarship from 2019 to 2022), Alberta Graduate Excellence Scholarship (AGES) in both 2019 and 2020, and NSERC CREATE in Milk Quality Scholarship (Canadian Mastitis Network) from 2018-2022. My heartfelt gratitude to all for their invaluable contributions!

I would like to express my thanks all other funders; NSERC Industrial Research Chair in Infectious Diseases of Dairy Cattle, Alberta Milk (Edmonton, AB, Canada), Dairy Farmers of Canada (Ottawa, ON, Canada), Lactanet (Guelph, ON, Canada) and Semex (Guelph, Canada).

I would like to thank **God,** who guided me throughout this long journey to get to where I am today. Thank you for all the opportunities and signs that you made for me.

Finally, I would like to thank all the people who were directly or indirectly involved in my life and gave me the strength to achieve this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CM	clinical mastitis
CNV	Copy Number Variant
DGV	direct genomic value
DIM	days in milk
EBV	estimated breeding value
GEBV	genomic estimated breeding value
GLMM	generalized linear mixed models
GS	genomic selection
GWAS	genome-wide association study
IMI	intramammary infection
kg	kilogram
mL	milliliter
NAS	non-*aureus* staphylococci
QTL	quantitative trait loci
REML	restricted Maximum Likelihood
SCC	somatic cell count
SCM	subclinical mastitis
SCS	somatic cell score
ssGBLUP	single-step genomic best linear unbiased prediction
SNP	single-nucleotide polymorphism
SE	Standard error

CHAPTER 1: General introduction

The escalating global population amplifies the demand for crucial food staples, notably dairy products like milk. This demand has not only led to an expansion of the global dairy herd but also emphasized the necessity to boost the milk yield per cow. However, a principal obstacle in increasing milk production in dairy cattle is udder health, with mastitis taking precedence. Mastitis stands as the most prevalent and economically impactful disease afflicting dairy cattle (Koeck et al., 2012a,b; Aghamohammadi et al., 2018).

1.1. What is mastitis?

Mastitis is a multifactorial and complex disease that occurs in response to the invasion of an infectious agent, mainly bacteria, in the udder, known as intramammary infection (IMI) (Berry and Meaney, 2006; Andersen et al., 2010; Constable et al., 2017). An IMI can manifest as subclinical (SCM) and as clinical mastitis (CM). It occurs in 3 stages: (i) invasion of the pathogen: pathogen transfer from teat end to milk inside the teat canal, (ii) infection: pathogens multiply and become established in the mammary gland of the host, and (iii) inflammation: a consequence of IMI and represents the stage of occurrence of CM (Constable et al., 2017; Ruegg, 2017). Bovine mastitis is, therefore, a disease resulting from the interaction between host, environment, and pathogens. Risk factors of mastitis can be grouped mainly into 3 categories: animal (genetics, age, stage of lactation, morphology and physiology of teat, nutrition, and milk production level of cow), environment (management of housing and bedding), and pathogens (ability to survive in the environment, virulence, and susceptibility to antimicrobials) (Constable et al., 2017).

1.1.1. Somatic cell count

Somatic cells have been used as an indicator of mastitis for a long time. Somatic cells in healthy cows consist of leucocytes (neutrophils (<11%), macrophages (66–88%), lymphocytes (10– 27%)), and a smaller percentage of epithelial cells (0–7%) (Sharma et al., 2011; Constable et al., 2017). Somatic cells are released in the mammary gland to provide a defence mechanism against infection in the udder (Sharma et al., 2011). During the IMI, particularly neutrophils (polymorphonuclear neutrophilic leukocytes (PMNL)) rapidly increase (≥ 65- 90 %; Schwarz et al., 2011; Constable et al., 2017) to provide the first line of defence against the infection (Rainard and Riollet, 2003, 2006), where these phagocytic cells engulf and kill the pathogens. One of the main factors that affect the somatic cell count (SCC) in milk is the level of infection in the mammary gland. Hence, SCC is used as a marker for the inflammatory process of the mammary gland. Healthy quarters generally have SCC <100,000 cells/mL with a low proportion of neutrophils (Rainard and Riollet, 2006; Constable et al., 2017). Any cut-off of SCC is not absolute though, e.g. a cut-off of 200,000 cells/mL has a sensitivity and a specificity of approximately 80% (Dohoo and Leslie, 1991).

1.1.2. Subclinical mastitis

Subclinical mastitis is defined based on SCC, where a cow is considered to have SCM if the SCC exceeds the defined cut-off value. Cows with SCM typically exhibit decreased milk yield, undesirable changes in milk composition, and inflammatory components in milk, but they do not have any physical abnormalities or visible changes in the milk (Harmon, 1994; Heringstad et al., 2000; Saidi et al., 2013). Even though elevated SCC is used to define SCM, there is no specific cut-off of SCC. Different studies have used different cut-off. The threshold of SCC used to classify

SCM ranges from 50,000 to 500,000 cells/mL, with 200,000 cells/mL being the most widely used cut-offs (Svendsen and Heringstad, 2006; Halasa et al., 2009; Koeck et al., 2012a; Krisanova et al., 2019).

1.1.3. Clinical mastitis

When the SCM is not resolved, it can lead to the clinical stage where the cow will start to show visible signs of infection, called CM. Clinical mastitis ranges from mild to severe. However, CM can also result from an IMI without first becoming subclinical. Symptoms of CM are swollen udder, abnormalities in milk such as clots, flakes and discoloration, and in severe cases a cow can exhibit fever, dehydration, anorexia, and stasis, and the CM can even result in death (Constable et al., 2017).

1.2. Pathogens

A broad spectrum of microbial species can cause mastitis; mainly bacteria are grouped as contagious, environmental, and opportunistic based on their epidemiology and pathophysiology (Constable et al., 2017). Moreover, based on the virulence and damaging potential of udder pathogens they could be grouped as major (e.g., *Staphylococcus aureus*, *Streptococcus agalactiae*, *Streptococcus uberis*, *Streptococcus dysgalactiae*, *Mycoplasma bovis*, and the coliforms) and minor pathogens (*Corynebacterium bovis* and non-*aureus* staphylococci (NAS)) (Reyher et al., 2012).

1.2.1. Contagious pathogens

Contagious pathogens are transmitted from infected quarters of an affected cow to other cows, especially during milking (Constable et al., 2017). *Streptococcus agalactiae* and *S. aureus* are the most common pathogens and IMI with these bacteria most often results in SCM (Keefe, 2012; Constable et al., 2017). Additionally, most of the contagious pathogens are considered to be major pathogens due to their effect on milk quality, production and SCC (Keefe, 2012). Less common contagious species are *Corynebacterium bovis* and *Mycoplasma* spp. (Constable et al., 2017).

1.2.2. Environmental pathogens

In herds where contagious pathogens have been controlled, CM is mostly caused by environmental pathogens (Constable et al., 2017). Environmental pathogens include coliforms (particularly *Escherichia coli* and *Klebsiella* spp.), environmental *Streptococcus* spp. such as *Streptococcus uberis*, and *Trueperella pyogenes.* The source of these pathogens is the cow's environment, where pathogens are transmitted from bedding, poor housing system and flies.

1.2.3. **Non-*aureus* staphylococci (NAS)**

A group of predominantly opportunistic pathogens, NAS, is the most frequent cause of IMI and is considered a minor pathogen (Condas et al., 2017). Non-*aureus* staphylococci consist of all *Staphylococcus* species other than *S. aureus*. Some species of NAS are shed through feces and can infect cows from the environment (Wuytack et al., 2019).

1.3. Occurrence and impact

Despite numerous research efforts and advancements in management practices, managing mastitis continues to be a major challenge (Ruegg et al., 2017). Globally, IMI prevalence is substantial, with variations observed during different lactation stages and parities, ranging from 20 to 75% for individual udder quarters and 49 to 97% for cows (Oliver et al., 2003; Sampimon et al., 2009a; Piepers et al., 2011). Regarding heifers, 12 to 57% of their udder quarters have IMI at the time of calving (Cook et al., 1992; Parker et al., 2007). A particular study in Belgium indicated that of 44,677 cows, 41% had at least one culture-positive quarter (Piepers et al., 2007). The occurrence of CM varies throughout lactation. In Canada, during the 1st, 2nd, and subsequent lactations, the incidence stands at 12.9, 18.6, and 22.2% respectively (CDN, 2013). It is noteworthy that mastitis incidence is particularly elevated during the initial 30 days of milk (DIM) in heifers as well as in multiparous cows. As cows enter early lactation, they experience major physiological, metabolic, immunological, and nutritional shifts, making them more prone to illnesses (Drackley, 1999; Mulligan and Doherty, 2008; Sordillo and Aitken, 2009). For Canadian Holsteins, quarter-level IMI prevalence is 33.4% for heifers and 25.8% for multiparous cows in the first 30 DIM. A substantial portion of these IMIs is attributed to NAS at 26.4% of affected quarters (Naqvi et al., 2018). When defining SCM based on a threshold of 200,000 cells/mL, the prevalence for heifers and multiparous cows during the first 30 DIM is 13.6 and 17.6%, respectively, in Canadian Holsteins (Naqvi et al., 2018). Further, Koeck et al. (2012a) reported prevalence of 13% of CM cases in first lactation Canadian Holsteins within 305 days after calving, with 33% of these first CM instances manifesting within the initial 30 DIM. Additionally, Naqvi et al. (2018) reported that CM affected 4% (0.73 cases per 100 quarters) of Canadian Holsteins heifers within 30 days after calving.

Mastitis has a huge economic impact on farm profitability through milk loss, physical changes in milk and culling and affects animal welfare (Aghamohammadi et al., 2018). Economic losses are caused mainly by loss of milk production, the replacement cost of culled cows, cost of control measures, and treatment costs (Constable et al., 2017). Early in the first lactation (5-14 DIM), a 1-unit increase in natural log-transformed SCC was associated with a decrease in 0.13 kg/d of milk yield in later lactation (De Vliegher et al., 2005). Moreover, milk losses due to SCM (defined based on an isolation of significant bacterial colony numbers and SCC >100,000 cells/mL) ranged from 0.07 to 1.4 kg/quarter milking and economic losses ranged from US $ 0.02 to 0.40/quarter milking based on pathogens (Gonçalves et al., 2018). In first lactation, following the diagnosis, cows with CM had substantial milk losses of 2.5, 8.4 3.2, 6.7. 7.6 kg/d which were caused by *Streptococcus* spp., *Staph. aureus*, *Staphylococcus* spp., *E. coli*, respectively (Grohn et al., 2004). Aghamohammadi et al. (2018) reported that expenses associated with mastitis were substantial, amounting to CAD 662/cow/year for an average Canadian dairy farm. A considerable amount of these expenses, accounting for 48%, could be traced back to SCM, while 34% and 15% were linked to CM and the adoption of preventative measures, respectively. The costs differ, however, depending on the mastitis pathogen involved. Sørensen et al. (2010) reported cost per CM case in Denmark of €231, 570, 380 and 206 for overall mastitis, *Staphylococcus aureus*, NAS and *E. coli*, respectively. Moreover, SCM has an impact on greenhouse gas (GHG) emission, where increased SCC is associated with increased GHG (Gülzari et al., 2018).

1.4. Prevention of mastitis using genetic selection

Despite the great effort taken on management practices to control and prevent mastitis, the incidence of CM has not reduced over time (Ruud et al., 2010; Richert et al., 2013; Bates and

Dohoo, 2016). Therefore, selective breeding for enhanced mastitis resistance should be emphasized as a sustainable way to control mastitis together with better management practices.

1.4.1. Genetic selection

The success of genetic selection depends on genetic variation and heritability. Although the heritability of mastitis resistance is low (<0.05; Koeck et al., 2012b; Jamrozik et al., 2013), many studies have explored the genetic variability of CM and concluded that although response to selection will be slow, there is enough genetic variability for genetic selection (Heringstad et al., 2003; Koeck et al., 2012b). Somatic cell score (SCS; log transformation of SCC; Ali and Shook, 1980) has been used as an indicator trait for CM for a long time due to its relatively higher heritability (0.16-0.18; Lactanet, 2021) and high correlation (0.69 and 0.68 in early and late lactation; Koeck et al., 2012a) with CM for genetic selection. As mastitis resistance has an antagonistic relationship with milk yield (0.29; Hinrichs et al., 2005), selection for higher milk yield will increase the susceptibility to mastitis; hence including mastitis resistance in breeding programs is necessary (Miglior et al., 2017). The first official run of genetic evaluation for mastitis resistance was introduced in August 2014 in Canada (Miglior et al., 2014; Jamrozik et al., 2013). Genetic evaluation for mastitis resistance index includes 11 traits; CM (calving to 150 DIM), mean SCS in early lactation (5-150 DIM), standard deviation of SCS (5-150 DIM), excessive test-day SCC (> 500,000 cells/mL) for first and later lactations and 3 type traits: fore udder attachment, udder depth and body condition score (Jamrozik et al., 2013). Mastitis resistance index has a heritability of 0.12 (Lactanet, 2021) and explains 72% of the genetic variation in mastitis resistance leading to high accuracy of genetic selection (CDN, 2013).

1.4.2. Genomic selection

Genomic selection (GS) is a better alternative to achieving breeding objectives for low-heritability traits. Genomic selection is a tool used to estimate genomic breeding values (GEBV) using dense markers covering the whole genome on which the selection decision will be based on (Goddard et al., 2007; Hayes et al., 2009). Compared to the conventional genetic selection method, GS is beneficial in improving prediction accuracy and reduced generation interval, thus leading to dramatic advancement in genetic selection of low heritability traits (Parker Gaddis et al., 2014). Genomic selection was implemented in Canada in 2009. With the implementation of GS annual lifetime performance index and pro$ changed from 45 (before genomics) to 97 points/yr (after 5 years of GS) and $102 (before genomics) to 246/yr (after 5 years of GS), respectively, in Canada (Lactanet, 2022). Total genetic gain realized in Canadian Holsteins before and after 5 years of GS for mastitis resistance, and SCS were 0.8 to 1.8 and 0.9 to 2.2 standard deviation units, respectively (Lactanet, 2022).

1.4.2.1. Single-step genomic prediction

In many countries, including Canada, the multi-step GS approach was introduced in 2009. The multi-step approach is complex and involves the prediction of traditional estimated breeding value (EBV) using pedigree information and subsequent de-regression, prediction of direct genomic values (DGV), and finally combination of DGV and parent average to generate GEBV (Sargolzaei and Schenkel, 2009). This approach can introduce bias and errors during de-regression (Legarra et al., 2014), as well as double-counting of information (Oliveira et al., 2019). In addition, as not all cattle are genotyped, and this preselection can introduce bias. To avoid constraints and limitations of the multi-step method, a single-step genomic best linear unbiased prediction (ssGBLUP) method has been introduced to estimate GEBV (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). Single-step GBLUP incorporates genomic, pedigree and phenotypic information into one model for genetic evaluation. Both genotyped and non-genotyped animals are simultaneously included in the evaluation through the H matrix (Legarra et al., 2009). The main benefits of ssGBLUP are accounting for pre-selection bias, accommodating more complicated models (multiple traits and threshold models), and increasing the accuracy of GEBV (Aguilar et al., 2010; Christensen et al., 2012; Baloche et al., 2014; Legarra et al., 2014; Zhang et al., 2016). Furthermore, ssGBLUP is more beneficial to low-heritability traits such as disease resistance and reproductive traits (Chen et al., 2011; Guarini et al., 2018; 2019).

1.4.2.2. Genome-wide association study (GWAS)

As mastitis is a complex disease genetically controlled by many causal genes with minor contributions to overall heritability (Pighetti and Elliott, 2011), identifying relevant genes associated with mastitis resistance is vital to understanding the underlying genetic architecture and can aid in GS to increase the frequency of favourable alleles in the population (Tiezzi et al., 2015). Genome-wide association studies offer a feasible approach for mapping underlying genes of complex traits such as mastitis resistance. However, classical GWAS using a multiple-step approach lacks the power to detect quantitative trait loci (QTLs) and precision for fine mapping, due to the introduction of bias and loss of information (Vitezica et al., 2011; Wang et al., 2012). As an alternative, ssGBLUP approach is faster, potentially more accurate and easier to implement to conduct GWAS (Wang et al., 2012). Moreover, several GWAS studies on mastitis-related traits have been conducted using different breeds, populations, approaches and sample sizes, resulting in a long list of candidate genes associated with mastitis resistance. Hence, it is difficult to identify the key functional candidate genes associated with mastitis-related traits.

1.4.2.3. Copy Number variants (CNVs)

To date, genomic variants such as SNPs have been used widely in GWAS to identify potential QTLs and functional candidate genes associated with mastitis resistance (Sahana et al., 2013; Welderufael et al., 2018). Copy Number Variants are a quantitative genomic structural variant and a major source of genomic diversity (Redon et al., 2006). Deletion or duplication of genomic sequence from 50 bp to 5 Mb is commonly referred to as CNV (Feuk et al., 2006; Mills et al., 2011). In humans, CNVs are associated with many diseases, including various forms of cancer, autoimmune disorders, schizophrenia and autism (Merikangas et al., 2009). Copy number variants have been reported to capture ¼ th of information that was not captured by SNPs (Xu et al., 2014). Due to challenges with CNV identification, few studies have explored CNV related with CM (Szyda et al., 2019; Żukowski et al., 2019) and SCS (Durán Aguilar et al., 2016; Ben Sassi et al., 2016; Prinsen et al., 2017; Zhou et al., 2018; Liu et al., 2019). However, so far, no study has identified CNVs associated with mastitis in the Canadian dairy cattle population due to difficulty in calling CNV either using SNP or short-read whole-genome DNA sequence data (Bickhart et al., 2012).

1.5. Thesis rationale

Mastitis has been a health problem for dairy cattle for several decades and has a large impact on the dairy sector. Although much research has been conducted on genetic aspects of mastitis to increase the mastitis resistance through genetic selection and to understand its genetic architecture in dairy cattle, there are still several research gaps. There is a lack of knowledge on genetic parameters of SCM with different thresholds in early lactation and pathogen-specific IMI in Canadian Holsteins. Furthermore, several GWAS studies have been conducted using SNPs to identify candidate genes associated with mastitis-related traits. However, there are apparently no published reports regarding CNVs associated with SCC in Canadian Holsteins. In addition, there is a need for a systematic review and gene prioritization analysis to bring in all the genes identified from several GWAS studies into one document and to prioritize candidate genes based on functionality to identify key functional candidate genes. Moreover, although studies have indicated that genomic prediction based on ssGBLUP approach is superior for low heritability traits, very little is known on the impact of ssGBLUP genomic prediction on SCM and SCS in early lactation in Canadian Holsteins.

1.6. Research objectives

1. Estimate genetic parameters for SCM in Canadian Holstein heifers in early lactation using alternative definitions of SCM based on test-day SCC and compare heritabilities and EBV for SCM resistance obtained from both threshold and linear models (Chapter 2);
2. Estimate genetic parameters for overall and pathogen-specific IMI and SCC (using SCS as the outcome parameter) in nonclinical primiparous and multiparous cows and determine genetic variation among sires for IMI-related traits (Chapter 3);
3. Conduct a systematic review of GWAS on resistance to mastitis and perform gene prioritization analysis of GWAS results of resistance to mastitis to identify potential functional candidate genes (Chapter 4);
4. Identify CNVs associated with SCC (using SCS as the outcome parameter) (Chapter 5);

5. Conduct genomic prediction using ssBGLUP approach for CM, SCM and SCC (using SCS as the outcome parameter) (Chapter 6).

1.7. References

Aghamohammadi, M., D. Haine, D. F. Kelton, H. W. Barkema, H. Hogeveen, G. P. Keefe, and S. Dufour. 2018. Herd-level mastitis-associated costs on Canadian dairy farms. Front. Vet. Sci. 5:1–12.

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

Ali, A. K. A., and G. E. Shook. 1980. An optimum transformation for somatic cell concentration in milk. J. Dairy Sci. 63:487–490.

Andersen, S., I. R. Dohoo, R. Olde Riekerink, H. Stryhn, and Mastitis Research Workers' Conference. 2010. Diagnosing intramammary infections: Evaluating expert opinions on the definition of intra- mammary infection using conjoint analysis. J. Dairy Sci. 93:2966– 2975.

Baloche, G., A. Legarra, G. Sallé, H. Larroque, J. M. Astruc, C. Robert-Granié, and F. Barillet. 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. J. Dairy Sci. 97:1107–1116.

Bates, A. J., and I. Dohoo. 2016. Risk factors for peri-parturient farm- er diagnosed mastitis in New Zealand dairy herds: Findings from a retrospective cohort study. Prev. Vet. Med. 127:70–76.

Ben Sassi, N., Ó. González-Recio, R. de Paz-Del Río, S. T. Rodríguez- Ramilo, and A. I. Fernández. 2016. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. J. Dairy Sci. 99:6371–6380.

Berry, D. P., and W. J. Meaney. 2006. Interdependence and distribution of subclinical mastitis and intramammary infection among udder quarters in dairy cattle. Prev. Vet. Med. 75:81–91.

https:

CDN. 2013. Official genetic evaluation for mastitis resistance. Accessed May 1, 2023. https://www.cdn.ca/articles.php?_postback_=1&_formname_=form1&fromdate=01%2F01%2F1999&todate=02%2F20%2F2020&keywords=Official+genetic+evaluation+for+mastitis+resistance&search=Search&_returnurl_=https%3A%2F%2Fwww.cdn.ca%2Fhome.php

Chen, C. Y., I. Misztal, I. Aguilar, S. Tsuruta, T. H. E. Meuwissen, S. E. Aggrey, T. Wing, and W. M. Muir. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. J. Anim. Sci. 89:23–28.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571.

Condas, L. A. Z., J. De Buck, D. B. Nobrega, D. A. Carson, S. Naushad, S. De Vliegher, R. N. Zaoks, J. R. Middleton, S. Dufour, J. P. Kastelic, and H. W. Barkema. 2017. Prevalence of non-*aureus* staphylococci species causing intramammary infections in Canadian dairy herds. J. Dairy Sci. 100:5592–5612.

Constable, P., K. Hinchcliff, S. Done, W. Grünberg, and O. Radostits. 2017. Veterinary medicine: A textbook of the diseases of cattle, horses, sheep, pigs and goats (Edition 11). Elservier, St. Louis, MO, USA.

Cook, W. F., E. A. Fiez, and L. K. Fox. 1992. Mastitis in first lactation Southwest Idaho dairy cows. J. Dairy Sci. 75:158.

De Vliegher, S., H. W. Barkema, H. Stryhn, G. Opsomer, and A. De Kruif. 2005. Impact of early lactation somatic cell count in heifers on milk yield over the first lactation. J. Dairy Sci. 88:938–947.

Dohoo, I. R., and K. E. Leslie. 1991. Evaluation of changes in somatic cell counts as indicators of new intramammary infections. Prev. Vet. Med. 10:225-237.

Drackley, J. K. 1999. Biology of dairy cows during the transition period: the final frontier. J. Dairy Sci. 82: 2259-2273.

Durán Aguilar, M., S. I. Román Ponce, F. J. Ruiz López, E. González Padilla, C. G. Vásquez Peláez, A. Bagnato, and M. G. Strillacci. 2016. Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. J. Anim. Breed. Genet. 134:49–59.

Feuk, L., A. R. Carson, and S. W. Scherer. Structural variation in the human genome. 2006. Nat. Rev. Genet. 7:85-97.

Goddard, M. E., and B. J. Hayes. 2007. Genomic selection. J. Anim. Breed. Genet.124:323-330.

Gonçalves, J. L., C. Kamphuis, C.M.M.R. Martins, J. R. Barreiro, T. Tomazi, A. H. Gameiro, H. Hogeveen, and M. V. Dos Santos. 2018. Bovine subclinical mastitis reduces milk yield and economic return. Livest. Sci. 210:25-32.

Grohn, Y. T., D. J. Wilson, R. N. Gonzalez, J. A. Hertl, H. Schulte, G. Bennett, and Y. H. Schukken. 2004. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. J. Dairy Sci. 87:3358-3374.

Guarini, A. R., D. A. L. Lourenco, L. F. Brito, M. Sargolzaei, C. F. Baes, F. Miglior, I. Misztal, and F. S. Schenkel. 2018. Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. J. Dairy Sci. 101:8076-8086.

Guarini, A. R., D. A. L. Lourenco, L. F. Brito, M. Sargolzaei, C. F. Baes, F. Miglior, I. Misztal, and F. S. Schenkel. 2019. Genetics and genomics of reproductive disorders in Canadian Holstein cattle. J. Dairy Sci. 102:1341-1353.

Gülzari, Ş. Ö., B. V. Ahmadi, and A. W. Stott. 2018. Impact of subclinical mastitis on greenhouse gas emissions intensity and profitability of dairy cows in Norway. Prev. Vet. Med. 150:19-29.

Halasa, T., M. Nielen, A. P. W. De Roos, R. Van Hoorne, G. De Jong, T.J.G.M. Lam, T. Van Werven, and H. Hogeveen. 2009. Production loss due to new subclinical mastitis in Dutch dairy cows estimated with a test-day model. J. Dairy Sci. 92:599–606.

Harmon, R. J. 1994. Physiology of mastitis and factors affecting somatic cell counts. J. Dairy Sci. 77:2103–2112.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433–443.

Heringstad, B., G. Klemetsdal, and J. Ruan. 2000. Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. Livest. Prod. Sci. 64:95–106.

Heringstad, B., G. Klemetsdal, and T. Steine. 2003. Selection responses for clinical mastitis and protein yield in two Norwegian dairy cattle selection experiments. J. Dairy Sci. 86:2990–2999.

Hinrichs, D., E. Stamer, W. Junge, and E. Kalm. 2005. Genetic analyses of mastitis data using animal threshold models and genetic correlation with production traits. J. Dairy Sci. 88:2260-2268

Jamrozik, J., A. Koeck, F. Miglior, G. J. Kistemaker, F. S. Schenkel, D. F. Kelton, and B. J. Van Doormaal. 2013. Genetic and genomic evaluation of mastitis resistance in Canada. Interbull

Bull. 47:43-51.

Keefe, G., 2012. Update on control of *Staphylococcus aureus* and *Streptococcus agalactiae* for management of mastitis. Vet. Clin. North Am. Food Anim. Pract. 28:203-216.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012a. Alternative somatic cell count traits to improve mastitis resistance in Canadian Holsteins. J. Dairy Sci. 95:432–439.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012b. Health recording in Canadian Holsteins: Data and genetic parameters. J. Dairy Sci. 95:4099-4108.

Kirsanova, E., B. Heringstad, A. Lewandowska-Sabat, and I. Olsaker. 2019. Alternative subclinical mastitis traits for genetic evaluation in dairy cattle. J. Dairy Sci. 102:5323-5329.

Lactanet. 2021. Heritability estimates used for genetic evaluation in Canada. Accessed May 1, 2023. https://lactanet.ca/en/heritability-estimates-used-for-genetic-evaluation-in-canada/

Lactanet. 2022. The impact of genomics. Accessed May 1, 2023. https://lactanet.ca/en/impact-genomics/

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54–65.

Liu, M., L. Fang, S. Liu, M. G. Pan, E. Seroussi, J. B. Cole, L. Ma, H. Chen, and G. E. Liu. 2019. Array CGH-based detection of CNV regions and their potential association with reproduction and other economic traits in Holsteins. BMC Genom. 20:1-10.

Merikangas, A. K., A. P. Corvin, and L. Gallagher. 2009. Copy-number variants in neurodevelopmental disorders: promises and challenges. Trends Genet. 25:536–544.

Miglior, F., A. Fleming, F. Malchiodi, L. F. Brito, P. Martin and C. F. Baes. 2017. A 100-Year

Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100:10251-10271.

Miglior, F., A. Koeck, J. Jamrozik, F. S. Schenkel, D. F. Kelton, G. J. Kistemaker, B. J. Van Doormaal. 2014. Index for mastitis resistance and use of BHBA for evaluation of health traits in Canadian Holsteins. Interbull Bull. No. 48. Berlin, Germany. Accessed September 4, 2023. https://journal.interbull.org/index.php/ib/article/view/1586/1588

Mills, R. E., K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, A. Chinwalla, D. F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L. M. Iakoucheva, Z. Iqbal, S. Kang, J. M. Kidd, M. K. Konkel, J. Korn, E. Khurana, D. Kural, H. Y. K. Lam, J. Leng, R. Li, Y. Li, C. Y. Lin, R. Luo, X. J. Mu, J. Nemesh, H. E. Peckham, T. Rausch, A. Scally, X. Shi, M. P. Stromberg, A. M. Stütz, A. E. Urban, J. A. Walker, J. Wu, Y. Zhang, Z. D. Zhang, M. A. Batzer, L. Ding, G. T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, K. Ye, E. E. Eichler, M. B. Gerstein, M. E. Hurles, C. Lee, S. A. McCarroll, and J. O. Korbel. 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470:59–65.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655.

Mulligan F. J., M. L. Doherty. 2008. Production diseases of the transition cow. Vet. J. 176:3-9.

Naqvi, S.A., J. De Buck, S. Dufour, and H. W. Barkema. 2018. Udder health in Canadian dairy heifers during early lactation. J. Dairy Sci. 101:3233–3247.

Olde Riekerink, R. G. M., H. W. Barkema, D. F. Kelton, D. T. Scholl. 2008. Incidence rate of clinical mastitis on Canadian dairy farms. J. Dairy Sci. 91:1366-1377.

Oliveira, H. R., L. F. Brito, M. Sargolzaei, F. F. de Silva, J. Jamrozik, D. A. L. Lourenco, and F.

S. Schenkel. 2019. Impact of including information from bulls and their daughters in the training population of multiple-step genomic evaluations in dairy cattle: A simulation study. J. Anim. Breed. Genet. 137:1–12.

Oliver, S. P., M. J. Lewis, B. E. Gillespie, H. H. Dowlen, E. C. Jae- nicke, and R. K. Roberts. 2003. Prepartum antibiotic treatment of heifers: Milk production, milk quality and economic benefit. J. Dairy Sci. 86:1187–1193.

Parker Gaddis, K. L., J. B. Cole, J. S. Clay, and C. Maltecca. 2014. Genomic selection for producer-recorded health event data in US dairy cattle. J. Dairy Sci. 97:3190-3199.

Parker, K. I., C. Compton, F. M. Anniss, A. Weir, C. Heuer, and S. McDougall. 2007. Subclinical and clinical mastitis in heifers following the use of a teat sealant precalving. J. Dairy Sci. 90:207–218.

Piepers, S., L. De Meulemeester, A. De Kruif, G. Opsomer, H. W. Barkema, and S. De Vliegher. 2007. Prevalence and distribution of mastitis pathogens in subclinically infected dairy cows in Flanders, Belgium. J. Dairy Res. 74:478–483.

Piepers, S., K. Peeters, G. Opsomer, H. W. Barkema, K. Frankena, and S. De Vliegher. 2011. Pathogen group specific risk factors at herd, heifer and quarter levels for intramammary infections in early lactating dairy heifers. Prev. Vet. Med. 99:91–101.

Pighetti, G. M., and A. A. Elliott. 2011. Gene polymorphisms: The keys for marker assisted selection and unraveling core regulatory pathways for mastitis resistance. J. Mammary Gland Biol. Neoplasia 16:421–432.

Prinsen, R.T.M.M., A. Rossoni, B. Gredler, A. Bieber, A. Bagnato, and M. G. Strillacci. 2017. A genome wide association study between CNV and quantitative traits in Brown Swiss cattle. Livest. Sci. 202:7–12.

Rainard, P., and C. Riollet. 2003. Mobilization of neutrophils and de-fense of the bovine mammary gland. Reprod. Nutr. Dev. 43:439–457.

Rainard, P., and C. Riollet. 2006. Innate immunity of the bovine mammary gland. Vet. Res. 37:369–400.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A.R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. González, M. Gratacòs, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, J. Zhang, L. Armengol, D.F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K.W. Jones, S.W. Scherer, and M. E. Hurles. 2006. Global variation in copy number in the human genome. Nature 444:444–454.

Reyher, K. K., D. Haine, I. R. Dohoo, and C. W. Revie. 2012. Examining the effect of intramammary infections with minor mastitis pathogens on the acquisition of new intramammary infections with major mastitis pathogens—A systematic review and meta-analysis. J. Dairy Sci. 11:6483-6502.

Richert, R. M., K. M. Cicconi, M. J. Gamroth, Y. H. Schukken, K. E. Stiglbauer, and P. L. Ruegg. 2013. Risk factors for clinical mastitis, ketosis, and pneumonia in dairy cattle on organic and small conventional farms in the United States. J. Dairy Sci. 96:4269– 4285.

Ruegg, P. L. 2017. A 100-year review: Mastitis detection, management, and prevention. J. Dairy Sci. 100:10381–10397.

Ruud, L. E., K. E. Bøe, and O. Østerås. 2010. Associations of soft flooring materials in free stalls with milk yield, clinical mastitis, teat lesions, and removal of dairy cows. J. Dairy Sci.

93:1578–1586.

Sahana, G., B. Guldbrandtsen, B. Thomsen, and M. S. Lund. 2013. Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. Anim. Genet. 44:620–626.

Saidi, R., D. Khelef, and R. Kaidi. 2013. Subclinical mastitis in cattle in Algeria: Frequency of occurrence and bacteriological isolates. J. S. Afr. Vet. Assoc. 84:E1–5.

Sampimon, O., H. W. Barkema, I. Berends, J. Sol, and T.J.G.M. Lam. 2009. Prevalence of intramammary infection in Dutch dairy herds. J. Dairy Res. 76:129–136.

Sargolzaei, M., and F. S. Schenkel. 2009. Canadian implementation of genomic evaluations. Interbull Bull. 40:214-218.

Schwarz, D., U. S. Diesterbeck, S. Konig, K. Brugemann, K. Schlez, M. Zschock, W. Wolter, and C. P. Czerny. 2011. Microscopic differential cell counts in milk for the evaluation of inflammatory reactions in clinically healthy and subclinically infected bovine mammary glands. J. Dairy Res. 78:448–455.

Sharma, N., N. K. Singh, and M. S. Bhadwal. 2011. Relationship of somatic cell count and mastitis: An overview. Asian-Australasian. J. Anim. Sci. 24:429-438.

Sordillo, L. M, S. L. Aitken. 2009. Impact of oxidative stress on the health and immune function of dairy cattle. Vet. Immunol. Immunopathol. 128.:104-109.

Sørensen, L. P., T. Mark, M. K. Sørensen, and S. Østergaard. 2010. Economic values and expected effect of selection index for pathogen-specific mastitis under Danish conditions. J. Dairy Sci. 93:358–369.

Svendsen, M., and B. Heringstad. 2006. Somatic cell count as an indicator of subclinical mastitis. Genetic parameters and correlations with clinical mastitis. Interbull Bull. 35:12–16.

Szyda, J., M. Mielczarek, M. Frąszczak, G. Minozzi, J. L. Williams, and K. Wojdak-Maksymiec. 2019. The genetic background of clinical mastitis in Holstein-Friesian cattle. Animal 13:2156-2163.

Tiezzi, F., K. L. Parker-Gaddis, J. B. Cole, J. S. Clay, and C. Maltecca. 2015. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PLoS One 10:e0114919.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. 93:357–366.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. 94:73–83.

Welderufael, B. G., P. Løvendahl, D. J. de Koning, L. L. G. Janss, and W. F. Fikse. 2018. Genome-wide association study for susceptibility to and recoverability from mastitis in Danish Holstein cows. Front. Genet. 9:141.

Wuytack, A., A. De Visscher, S. Piepers, F. Boyen, F. Haesebrouck, and S. De Vliegher. 2019. Non-*aureus* staphylococci in fecal samples of dairy cows: First report and phenotypic and genotypic characterization. J. Dairy Sci. 102:9345-9359.

Xu, L., J. B. Cole, D. M. Bickhart, Y. Hou, J. Song, P. M. VanRaden, T. S. Sonstegard, C. P. Van Tassell, and G. E. Liu. 2014. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genom. 15:683.

Zhang, X., D. Lourenco, I. Aguilar, A. Legarra, and I. Misztal. 2016. Weighting strategies for single-step genomic BLUP: An iterative approach for accurate calculation of GEBV and GWAS. Front. Genet. 7:1–14.

Zhou, Y., E. E. Connor, G. R. Wiggans, Y. Lu, R. J. Tempelman, S. G. Schroeder, H. Chen, and

G. E. Liu. 2018. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. BMC Genom. 19:314.

Żukowski, K., P. Topolski, and A. Żarnecki. 2019. Genome-wide patterns of copy number variation in case-control mastitis study of Polish HF cows using high-density Illumina BovineSNP50 v3 BeadChip array. Translational Res. Vet. Sci. 2:9-22.

CHAPTER 2: Genetic analysis of subclinical mastitis in early lactation of heifers using both linear and threshold models

2.1. Abstract

Subclinical mastitis (SCM) causes economic losses for dairy producers by reducing milk production and leading to higher incidence of clinical mastitis and premature culling. The prevalence of SCM in first lactation heifers is highest during early lactation. The objective of this study was to estimate genetic parameters for SCM in early lactation in first-parity Holsteins. Somatic cell count test-day records were collected monthly in 91 Canadian herds participating in the National Cohort of Dairy Farms of the Canadian Bovine Mastitis Research Network. Only the first test-day record available between 5 and 30 DIM was considered for analysis. The final dataset contained 8,518 records from first lactation Holstein heifers. Six alternative traits were defined as indicators of SCM, using various cut-off values of SCC, ranging from 150,000 to 400,000 cells/mL. Both linear and threshold animal models were used. Overall prevalence of SCM using the 6 traits ranged from 13 to 24%. Heritability estimates (SE) from linear and threshold models ranged from 0.037 to 0.057 (0.015 to 0.018) and from 0.040 to 0.051 (0.017 to 0.020), respectively. There were strong genetic correlations (SE) among alternative SCC traits, ranging from 0.90 to 0.99 (0.013 to 0.069), indicating that these 6 traits were genetically similar. Despite low heritability, based on estimated breeding values (EBV) predicted from both models, there was exploitable genetic variation among sires. Higher EBV of SCM resistance corresponded to sires with a higher percentage of daughters without SCM. Based on a linear model (all 6 traits), percentage of daughters with SCM ranged from 5 to 13% and from 19 to 33% for the top 10% and

worst 10 % of 69 sires with minimum 20 daughters in at least 5 herds, respectively. Spearman's rank correlations among EBVs of sires predicted from linear (from 0.75 to 0.95) and threshold (from 0.74 to 0.95) models were moderate to high, respectively. Very high rank correlations (0.98 to 0.99) between EBV predicted for the same trait from linear and threshold model indicated that re-ranking of sires based on model used was minimal. In conclusion, despite low heritability, there was utilizable genetic variation in early lactation of heifers. Hence, genetic selection to improve genetic resistance to SCM in early lactation of heifers was deemed possible.

2.2. Introduction

Mastitis is one of the most prominent and costly diseases of dairy cattle (Hortet et al., 1999; Koeck et al., 2012a; Saidi et al., 2013). Mastitis is an intramammary infection (**IMI**) that causes udder inflammation, abnormalities in the milk, reduced milk production, and a high incidence of involuntary culling (Halasa et al., 2007). Cows with subclinical mastitis (**SCM**), defined as SCC exceeding a cut-off value, do not have any physical abnormalities (Saidi et al., 2013), but have decreased milk production, undesirable changes in milk composition, and inflammatory components in milk (Heringstad et al., 2000). Moreover, increased prevalence of SCM is associated with a higher incidence of clinical mastitis (Rupp et al., 2000; Steeneveld et al., 2008), which can result in involuntary culling and consequently has a great economic impact. Herd-level economic losses caused by SCM are substantial and even exceed those caused by clinical mastitis (Huijps et al., 2008). The main economic loss associated with SCM is milk loss. De Vliegher et al. (2005) reported that early in the first lactation (5 to 14 DIM), a 1 unit of increase in natural log-transformed SCC was associated with decrease in 0.13 kg/d of milk yield later in lactation.

Elevated SCC is commonly used to detect SCM (Harmon, 1994; Pyörälä, 2003). However, the SCC cut-off for definition of SCM is not uniform; therefore, prevalence of SCM is not consistent across studies. The threshold of SCC used to classify SCM ranges from 100,000 to 500,000 cells/mL, with 200,000 cells/mL being the most widely used cut-off (Svendsen and Heringstad., 2006; Halasa et al., 2009; Koeck et al., 2012b). The reported prevalence of SCM ranges from 11 to 52% (Gianneechini et al., 2002; Svendsen and Heringstad, 2006; Santman-Berends and Sampimon, 2011).

In first-lactation heifers, the prevalence of SCM is highest in early lactation and it is associated with increased SCC throughout the first lactation (De Vliegher et al., 2004). Koeck et al. (2012a) reported that 35% of CM occurred during the first month after calving, perhaps a consequence of not detecting SCM. The high prevalence of SCM in early lactation is mainly attributed to unresolved IMI before the first calving (Trinidad et al., 1990) and less frequently to a new IMI. Hence, breeding for heifers less susceptible to SCM, in combination with efficient management practices, is a sustainable strategy to prevent and control mastitis. Some genetic studies have regarded SCM as a binary trait (Koivula et al., 2004; Svendsen and Heringstad, 2006; De Haas et al., 2008; Urioste et al., 2012). Most genetic studies have focused on mean lactational SCC and somatic cell score (**SCS**) (Shook and Schutz, 1994; Koeck et al., 2012b; Alam et al., 2015). Reported heritability of SCM ranges from 0.03 to 0.17 (Koivula et al., 2004; Svendsen and Heringstad, 2006; De Haas et al., 2008; Urioste et al., 2012). However, as the definition of SCM varies, it is difficult to compare genetic parameters across studies. Moreover, genetic studies on SCM that focus on early lactation in heifers are very scarce.

Objectives of this study were to: 1) estimate genetic parameters for SCM in Canadian Holstein heifers in early lactation, using alternative definitions of SCM based on test-day SCC;

and 2) compare heritabilities and estimated breeding values (**EBV**) for SCM resistance obtained from both threshold and linear models.

2.3. Material and Methods

2.3.1. Data

Rationale and details regarding selection of farms were described (Reyher et al. 2011). Briefly, test-day records for SCC were collected monthly between 2005 and 2009 in 91 Canadian dairy herds participating in the National Cohort of Dairy Farms of the Canadian Bovine Mastitis Research Network (CBMRN). To represent inherent diversity of Canadian dairy herds, commercial dairy farms were selected from 6 provinces: Alberta, Ontario, Québec, Prince Edward Island, Nova Scotia, and New Brunswick. All participating herds had > 80% Holstein cows, milked twice per day, and participated in a regular DHI recording system.

The initial dataset contained 352,101 records of 19,319 Holstein cattle from 91 Canadian dairy herds. For genetic analysis, herds were required to have a minimum of 50 lactating females. Test-day records from first lactation Holstein heifers with an age at first calving from 18 to 36 mo were considered. Only the first test-day record available between 5 to 30 DIM was included in analyses. The final dataset contained records from 8,518 Holstein first lactation heifers from 90 herds. However, the pedigree file contained a total of 43,344 animals, generated by tracing back 7 generations.

2.3.2. Trait Definition

Six traits were defined as an indicator of SCM, based on SCC in composite DHI milk samples. All traits were coded as binary variables (0 or 1), with 1 was assigned to the presence of

SCM. For the first 3 traits: SCM_{150}, SCM_{200} and SCM_{250}, SCC cut-off values of 150,000, 200,000 and 250,000 cells/mL from $5 \leq DIM \leq 30$ were used, respectively. Heifers with SCC exceeding cut-off were considered affected (1), with all other heifers designated healthy (0). For the latter 3 traits, namely $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$, heifers were considered affected (1) if SCC > 400,000 cells/ml within $5 \leq DIM \leq 10$ and SCC > 150,000, 200,000 and 250,000 cells/ml, respectively, within $11 \leq DIM \leq 30$. Different thresholds for different periods of DIM were used to account for dilution effect and fluctuation due to DIM (Miller et al., 1993; Harmon, 1994; Schepers et al., 1997). A similar approach was used by Koeck et al. (2012b). Thresholds and intervals were selected based on mean SCC distribution from 5 to 30 DIM (Figure 2-1b).

2.3.3. Statistical Models

Subclinical mastitis was analyzed using both linear and probit threshold models. In theory, threshold models are more appropriate for analysis of binary response traits, as they account for the binary nature of the data (Gianola, 1982; Gianola and Foulley, 1983). Existence of an underlying continuous non-observable variable, called liability (λ) is assumed in a threshold-liability model. If the λ is larger than a fixed threshold (τ), then the observed binary response takes the value of 1, otherwise it is designated 0.

Given the mean and variance, liability was assumed to be normally distributed. As threshold (τ) and the residual variance (σ^2_e) were not identifiable, these parameters were assigned arbitrary values of 0 and 1, respectively. When a linear model is used for analysis of binary data, heritability estimates change according to the frequency of the trait (Gianola, 1982). However, when the progeny group size is large, binary trait analyses using threshold linear models yield similar sire rankings, with respect to single-trait genetic evaluation (Hagger and Hofer, 1990;

Heringstad et al., 2003; Negussie et al., 2008). Therefore, linear models are often implemented for routine genetic evaluation of binary traits due to computational efficiency.

Variance and covariance components were estimated using Restricted Maximum Likelihood (REML) procedure for linear models and generalized linear mixed models (GLMM) with a probit link function for the threshold model in ASReml (Gilmour et al., 2009). The following probit link function was used: $E(y_n) = \Phi(\lambda_n)$, where $E(\lambda_n)$ is the expected value for the n^{th} observation (0 or 1); Φ is the normal cumulative density function; and λ_n is n^{th} fitted value on the underlying normal scale.

The following linear model was applied to all SCM traits:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_{hy}\mathbf{hy} + \mathbf{Z}_a\mathbf{a} + \mathrm{e}$$

where **y** is a vector of observations; **β** is a vector of systematic effects, including fixed effects for age at first calving (10 classes) and year-season of calving (20 classes); **hy** is a vector of random herd-year of calving effects (441 levels); a is a vector of random additive genetic animal effects (43,344 levels); **e** is a vector of random residuals; and $\mathbf{X}$, $\mathbf{Z}_{hy}$, and $\mathbf{Z}_a$ are corresponding incidence matrices. The linear model used in the current study was similar to the model used for genetic evaluation of mastitis in Canada (Jamrozik et al., 2013) and by Koeck et al. (2012b).

First and last classes of age at first calving were < 24 and > 31 mo, whereas other classes were single months. Year of calving classes were from < 2005 to 2009. Year-season of calving classes were formed, from January 2005 to December 2009 by combining 3 consecutive months: January to March, April to June, July to September, and October to December.

Random effects were assumed to be normally distributed, with means equal to zero and covariance structure equal to:

$$v\begin{bmatrix}\mathbf{hy}\\ \mathbf{a}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{I}\sigma^2_{\mathrm{hy}} & 0 & 0\\ 0 & \mathbf{A}\sigma^2_{\mathrm{a}} & 0\\ 0 & 0 & \mathbf{I}\sigma^2_{\mathrm{e}}\end{bmatrix}$$

where σ^2_{hy}, σ^2_{a}, and σ^2_{e} are the herd-year of calving, animal additive genetic, and residual variances, respectively. **I** and **A** are the identity matrix and additive relationship matrix, respectively.

Subsequently, bivariate linear animal models were used to estimate phenotypic, genetic, and residual correlations between the 6 analyzed traits. Bivariate analyses using a threshold model were not performed, as similarity in estimated genetic correlations from linear and threshold models have been reported (Gianola, 1982; Negussie et al., 2008). Random effects were assumed to be normally distributed with mean equal to zero and a (co)variance structure as follows:

$$v\begin{bmatrix}\mathbf{hy}\\ \mathbf{a}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\boldsymbol{HY}\otimes\boldsymbol{I} & 0 & 0\\ 0 & \mathbf{G}\otimes\mathbf{A} & 0\\ 0 & 0 & \mathbf{R}\otimes\mathbf{I}\end{bmatrix}$$

Where:

$$\mathbf{HY}=\begin{bmatrix}\sigma^2_{\mathrm{hy1}} & \sigma_{\mathrm{hy12}}\\ \sigma_{\mathrm{hy12}} & \sigma^2_{\mathrm{hy2}}\end{bmatrix}$$

is the (co)variance (2×2) matrix between traits due to herd-year calving effects;

$$\mathbf{G}=\begin{bmatrix}\sigma^2_{\mathrm{a1}} & \sigma_{\mathrm{a12}}\\ \sigma_{\mathrm{a12}} & \sigma^2_{\mathrm{a2}}\end{bmatrix}$$

is the covariance (2×2) matrix between traits due to animal additive genetic effects, and

$$\mathbf{R}=\begin{bmatrix}\sigma^2_{\mathrm{e1}} & \sigma_{\mathrm{e12}}\\ \sigma_{\mathrm{e12}} & \sigma^2_{\mathrm{e2}}\end{bmatrix}$$

is the residual covariance (2×2) matrix between traits; ⊗ is the Kronecker product; **A** is the additive relationship matrix; and **I** is an identity matrix.

Heritability (h^2) for each trait was calculated from univariate analysis as follows:

$$h^2 = \sigma_a^2 / (\sigma_{hy}^2 + \sigma_a^2 + \sigma_e^2)$$

where σ_{hy}^2, σ_a^2, and σ_e^2 are the herd-year of calving, animal additive genetic and residual variances, respectively.

Heritability estimates from threshold model were transformed to an observed scale for comparison purposes, using the Dempster and Lerner (1950) method:

$$h^2_{obs} = h^2_{und} \times (z_{ord})^2 / p \times (1-p)$$

where h^2_{obs} is heritability on an observed scale, h^2_{und} is heritability on the underlying scale estimated with the threshold model, z_{ord} is the ordinate height of the normal distribution at the threshold point corresponding to *p*, and *p* is the prevalence of SCM_{150}, SCM_{200}, SCM_{250}, $SCM_{400,150}$, $SCM_{400,200}$ or $SCM_{400,250}$.

Spearman rank correlations between estimated breeding values (EBV) predicted using the 6 SCM traits, and between EBV predicted using 2 statistical approaches (linear vs threshold model) were calculated using the CORR procedure of SAS (SAS Institute Inc., 2013). Only the 69 sires with at least 20 daughters in 5 herds were included in this part of the analysis.

2.4. Results

2.4.1. Descriptive Statistics

The mean SCC was high in first-lactation heifers during early lactation, then declined gradually, and became stable over the remainder of the first lactation (Figure 2-1a). In early lactation, mean SCC was high from 5 to 10 DIM and then decreased from 11 to 30 DIM (Figure 2-1b). Among the first 3 traits (SCM_{150}, SCM_{200} and SCM_{250}), which were defined by single threshold values from 5 to 30 DIM, SCM_{150} had the highest overall SCM prevalence of 24.3% followed by SCM_{200} (18.7%), and SCM_{250} (15.2%) (Table 2-1). As expected, prevalence of SCM decreased with an increasing threshold value of SCC. Likewise, among the other 3 traits ($SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$), which accounted for dilution effect, $SCM_{400,150}$ had the highest overall prevalence of 19.6% followed by $SCM_{400,200}$ (15.9%) and $SCM_{400,250}$ (13.4%). These traits also had a similar pattern as the first 3 traits (SCM_{150}, SCM_{200} and SCM_{250}). Across provinces, prevalence of SCM from 5 to 30 DIM ranged from 10.4 to 27.1% for the 6 SCM traits. Prevalence of SCM was highest in Ontario (range: 15.0 to 26.8%) and Prince Edward Island (range: 14.2 to 27.1%), but lowest in Alberta (range: 10.4 to 19.0%). In general, overall and across-provinces prevalence of SCM was high in early lactation but decreased as the threshold used to differentiate affected and not affected heifers increased.

2.4.2. Genetic Parameters of SCM

Heritability estimates (SE) from the linear model ranged from 0.037 to 0.057 (0.015 to 0.018) (Table 2-2). Among the first 3 traits, highest heritability was estimated for SCM_{150} (0.057), followed by SCM_{200} (0.053) and SCM_{250} (0.037). As heritability is frequency dependent, it decreased with increasing threshold of SCC used to define SCM traits. Likewise, heritability of the latter 3 traits ($SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$) that accounted for the dilution effect were in the same range as heritability of the first 3 traits. Also, heritability decreased with decreasing average prevalence of SCM in a trait. Heritability of $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ was 0.051, 0.048 and 0.040, respectively (Table 2-2).

Heritabilities estimated (SE) from the threshold model (underlying scale) ranged from 0.040 to 0.051 (0.017 to 0.020). These heritabilities estimates were lower compared to the estimates from linear model. However, threshold model estimates were not significantly different from estimates from linear model. Heritability was highest for SCM_{200} (0.051), followed by SCM_{150} (0.046) and SCM_{250} (0.040). Traits $SCM_{400,150}$ and $SCM_{400,200}$ had the same heritability estimation of 0.047, whereas $SCM_{400,250}$ had a lower estimation of 0.043. However, pattern of reduction in the heritability estimates from linear model with increase of threshold of SCC used to define SCM traits were not observable here, as heritability estimation from the threshold model does not depend on the frequency. For comparison purposes of heritability estimations from linear and threshold models, threshold model estimations were converted to an observed scale (Dempster and Lerner, 1950). Converted heritabilities were lower than from linear models and ranged from 0.017 to 0.025 for the 6 studied traits.

Genetic correlations (SE) estimated from linear animal models among the 6 traits were very strong and ranged from 0.906 to 0.998 (0.013 to 0.069) (Table 2-3). Among the first 3 traits (SCM_{150}, SCM_{200} and SCM_{250}), genetic correlations were stronger when thresholds of these traits were closer to each other. Likewise, the latter 3 traits ($SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$) had a similar pattern. With respect to genetic correlations between the first and last 3 traits, the strongest value was between SCM_{250} and $SCM_{400, 200}$ (0.993), whereas the lowest estimation was between SCM_{150} and $SCM_{400, 150}$ (0.906).

Moderate to strong phenotypic correlations were estimated among all 6 traits (Table 2-3). Similar to the pattern genetic of correlations, phenotypic correlations were stronger among traits with a threshold close to each other. The strongest correlation (0.929) was between SCM_{250} and $SCM_{400,250}$. The lowest correlation of 0.695 was estimated between SCM_{150} and $SCM_{400,250}$. Estimated residual correlations were moderate to high (0.715 to 0.926) and tended to be similar to genetic correlations in direction (Table 2-4).

2.4.3. Prediction of Breeding Values

Moderate to high Spearman rank correlations were estimated among EBVs of the 6 analyzed traits predicted from linear and threshold models (Table 2-5). With respect to EBVs predicted from linear models, there was a similar high Spearman rank correlation of 0.95 between EBV of SCM_{200} and SCM_{250}, and $SCM_{400,200}$ and $SCM_{400,250,}$ respectively. The lowest correlation (0.75) was estimated between EBV of SCM_{150} and $SCM_{400,250}$. Rank correlations within EBV for the same trait predicted from threshold models ranged from 0.74 to 0.95. Furthermore, rank correlation between EBV predicted from linear and threshold models ranged from 0.98 to 0.99.

Despite low heritability, there were substantial variations in sires' EBV (Table 2-6). Estimated breeding values were reversed in sign in order to relate higher EBV to sires with more healthy daughters. Mean percentage of daughters with SCM ranged from 12.5 to 18.4% for the analyzed 6 traits. Based on the linear model estimation, the percentage of daughters with SCM ranged from 5.6 to 13.6% and from 19.5 to 33.3% for the best (10% decile), and the worst (90% decile) sires, respectively (within 5 to 30 DIM). Similarly, percentage of daughters with SCM in early lactation estimated based on a threshold model varied from 5.6 to 14.9% and from 18.2 to 33.3% for the best and worst sires, respectively. These sires were the best 10% and worst 10 % of the 69 sires with minimum 20 daughters in at least 5 herds. Although heritability of SCM was low, there was a large exploitable variation in percentage of healthy daughters per sire's EBV within 5 to 30 DIM (Figure 2-2).

2.5. Discussion

Mean SCC was high in early lactation and then gradually decreased and remained stable at approximately 35 DIM (Figures 2-1a and 2-1b). A similar pattern of SCC was observed by Koeck et al. (2012a) for first-lactation heifers with and without clinical mastitis (5 to 305 DIM). However, SCC of cows with clinical mastitis was higher throughout lactation. In the current study mean SCC of diseased cows were relatively higher than the SCC of healthy cows from 5 to 30 DIM for analyzed 6 traits. Likewise, Hagnestam-Nielsen et al. (2009) reported a similar trend for geometric mean SCC for primiparous cows with SCM and without SCM. Authors defined the SCM using a threshold of 175,000 and 50,000 cells/mL at 1 week and 2 to 44 weeks of lactation, respectively. Distribution of SCC throughout lactation is affected by several factors, including IMI status, herd, sampling time, quarter location, stage of lactation, parity, season, milk production, and cow itself (Harmon, 1994). Elevated SCC in early lactation is mainly due to IMI (Harmon, 1994; Laevens et al., 1997; Barkema et al., 1999). Intramammary infection at calving, particularly caused by major pathogens (*Staphylococcus aureus*, *Streptococcus agalactiae*, *Streptococcu*s *dysgalactiae*, *Streptococcus uberis*, and *Escherichia coli)*, resulted in more elevated SCC compared to IMI caused by minor pathogens and culture-negative quarter milk samples (Barkema et al., 1999). Intramammary infection often occurs earlier in life (Trinidad et al., 1990) and persists for an extended interval if unresolved, resulting in elevated SCC in early lactation. Additionally, negative energy balance status in early lactation (Nyman et al., 2008) increases susceptibility to new IMI, due to impaired immune functions (Wathes et al., 2009; Nyman et al., 2008; Esposito et al., 2014).

The dilution effect of milk also has a role in increased SCC in early lactation, if no IMI is present. At the onset of lactation, milk production is low, which results in high SCC in milk (Schepers et al., 1997). Therefore, in the present study, traits $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ were defined to account for the dilution effect (based on Figure 2-1b) by having a higher threshold value of 400,000 cells/mL within 5 to 10 DIM and slightly lower values from 11 to 30 DIM (150,000, 200,000 and 250,000 cells/mL). Somatic cell count of quarter milk is elevated immediately postpartum, regardless of infection status, but SCC of quarters without IMI rapidly declines, likely as a result of the dilution effect (Harmon, 1994; Barkema et al., 1999). Emanuelson and Funke (1991) estimated a dilution effect of 11% when estimated conditionally on mastitis prevalence. Reports on the interval after calving reported a physiological elevated SCC in non-infected quarters, which differed from 3 to 35 DIM (Reichmuth, 1975; Sheldrake et al., 1983), yet this elevation of SCC was lower compared to SCC from infected quarters (Harmon, 1994; Barkema et al., 1999). However, the presence of a dilution effect was non-conclusive across studies. In a recent study by Boland and More (2013), there was no evidence of a dilution effect in Irish Dairy cattle. Regardless, based on Figure 2-1b traits, $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ were defined to account for a potential dilution effect in the present study. Therefore, they had high threshold of SCC 400,000 cells/mL from 5 to 10 DIM and lower thresholds (150,000, 200,000 and 250,000 cells/mL) from 11 to 30 DIM. A similar approach was taken in a study by Koeck et al. (2012b).

Several studies have reported a wide range of prevalence for SCM. Due to inconsistencies in defining SCM and the interval under consideration, it is difficult to compare prevalence among studies (Ruegg, 2017). However, Jaeger et al. (2017) recently reported that a SCC threshold of 150,000 cells/mL was an optimal threshold to identify IMI with all possible pathogens, with sensitivity of 90.3% and specificity of 71.8%. Moreover, a quarter-milk SCC is preferred over a composite milk SCC to identify SCM, as increases in quarter SCC gets diluted in the latter (Barkema et al., 1999). Besides, several studies also reported that SCC patterns provided additional information compared to mean lactation SCC, for genetic evaluation of mastitis resistance (de Haas et al., 2003; Koeck et al., 2012b).

Estimated overall and across-provinces prevalence of early lactation SCM in heifers was comparable to other studies, despite variations among studies. A recent review reported a high quarter-level early postpartum SCM prevalence of 12 to 57% (De Vliegher et al., 2012). Similar to the lactation stage (5 to 30 DIM) analyzed in the present study, Naqvi et al. (2018) reported SCM prevalence of 13, 15 and 20% in low (SCC <150,000 cells/mL), medium (SCC 150,000 to 300,00 cells/mL) and high (SCC>300,00 cells/mL) bulk milk SCC (BMCC) categories within the first 30 DIM. In agreement with the presented study, those authors also reported Alberta had the lowest prevalence of SCM. Those authors defined SCM using a threshold of 200,000 cells/mL. In comparison to the present study, Santman-Berends and Sampimon (2011) reported that 26% of cows per herd had SCM from 4 to 100 DIM, based on a threshold of 150,000 cells/mL. Additionally, in the present study, prevalence of SCM decreased as the threshold used to differentiate affected and non-affected heifers increased. De Vliegher et al. (2001) also reported a similar pattern of reducing prevalence (65, 16, 13 and 6%, respectively) with increasing threshold in early lactation (5 to 14 DIM), with thresholds ranging from 150,000 to 1,000,000 cells/mL. In support of this view, Svendsen and Heringstad, (2006) estimated prevalences of 45.5, 25, 17, and

12% for SCM with thresholds of 50,000, 100,000, 150,00 and 200,000 cells/ml, respectively, from 30 to 270 DIM in first-lactation heifers.

In many studies, genetic evaluation of SCC was based on lactational average SCC or SCS. Moreover, comparisons among estimated heritabilities of SCM is difficult, as it varies with definition of SCM and on the interval considered. However, estimated heritabilities (0.037 to 0.057) from a linear model in the present study were in line with other studies. In general, heritability of SCM using a linear model ranged from 0.03 to 0.17 (Koivula et al., 2004; Svendsen and Heringstad, 2006; De Haas et al., 2008; Urioste et al., 2012). Most studies have focused on the entire first lactation, whereas very few have explored only early lactation (Koivula et al., 2004). A heritability of 0.05 was reported by Koeck et al. (2012b) using a linear model for a trait defined as cows with at least 1 SCC test-day above 500,000, 300,000, and 200,000 cells/mL within 5 to 10, 11 to 30, and 31 to 305 DIM, respectively, considered as affected (1), otherwise as healthy (0). These authors defined SCM with higher thresholds for 5 to 10 DIM and 11 to 30 DIM than used in the present current study and their analysis was done for the entire lactation (until 305 DIM). In addition, the same authors reported a low heritability of 0.03 for high SCC trait defined as at least 1 SCC test-day > 500,000 cells/ml from 5 to 305 DIM. Comparable to the present study, Svendsen and Heringstad (2006) reported lower heritabilities with an increasing threshold used to define SCM. Estimated heritabilities were 0.086, 0.078, 0.059 and 0.045 for SCM with threshold ranges of 50,000 to 200,000 cells/ml, respectively, in first lactation from 30 to 270 DIM using a linear sire model. This may be due to decreased frequency of affected cows and frequency dependency of linear models (Gianola, 1982). A similar pattern was reported by Koeck et al. (2012b).

De Haas et al. (2008) reported a low heritability (0.03 to 0.04) for SCM for primiparous cows over the course of first lactation using a linear model. These authors defined SCM as a transition from 2 consecutive SCC recordings below and above thresholds of 150,000 cells/ml. They also reported a heritability of 0.06 for trait 'suspicion of IMI,' defined as absence or presence of test-day SCC >150,000 cells/mL. This was similar to the present study for SCM_{150} with a linear model (0.057). Koivula et al. (2004) reported heritabilities of 0.045 and 0.094 from 5 to 20 and 31 to 60 DIM, respectively, for $\log_e$ SCC in first lactation for Holsteins using a linear model. Also, authors stated that the low heritability during early lactation was partly due to larger environmental variation. In contrast to other reports, Urioste et al. (2012) estimated a higher heritability of 0.17 for first lactation. This may have been due to SCM defined as number of periods from DIM > 45 with a SCC > 150,000 cells/mL and without treatment for clinical mastitis.

Threshold models are theoretically more appropriate for analysis of binary traits, as they assume existence of an underlying normally distributed variable (Gianola and Foulley, 1983). Moreover, threshold models provide a better comparison between traits and studies with different prevalence, as it is not as frequency-dependent as a linear model (Gianola, 1982). However, computational demands may pose a problem for routine genetic evaluation, based on a large dataset using a threshold animal model. Application of linear models to binomial traits is a common approach in routine genetic evaluation, as it is less computationally demanding. Furthermore, there was no difference in sire ranking using both models when progeny group size was large (Heringstad et al., 2003; Negussie et al., 2008). In the present study, heritabilities estimated from threshold models ranged from 0.040 to 0.051. However, there are very few publications on estimates of heritability for SCM, treating it as binary trait and using a threshold probability model for genetic analysis. Haugaard et al. (2013) reported heritabilities (underlying scale) of 0.04 (0.01) to 0.14 (0.03) for the pathogen-specific SCM using threshold sire model. However, authors defined the SCM based on test-day SCC records (threshold of

100,00cells/mL) and bacteriological results of milk samples. Regardless, several studies reported heritability estimates for CM which ranged from 0.05 to 0.126 using threshold models (Kadarmideen et al., 2000; Heringstad et al., 2003; Hinrichs et al., 2005). In general, heritability estimation from threshold models on a probability scale is higher than estimation from a linear model (Heringstad et al., 1997; Kadarmideen et al., 2000); however, in our study, it was lower but not significantly different from linear model estimates. For comparison purposes of heritabilities estimated from linear and threshold models, threshold model estimations in the present study were converted to an observed scale (Dempster and Lerner, 1950). After transformation, heritability estimations from the threshold model were lower than the linear model and ranged from 0.017 to 0.025 for the 6 studied traits. Since a similar model was used for both statistical approaches with same dataset, lower heritability estimation from threshold model on underlying and observed scale might be due to the small sample size and the nature of the data used in the current study. Therefore, future studies with large datasets are required.

High genetic correlations among the 6 traits demonstrated that these traits were genetically similar. Similarly, Svendsen and Heringstad (2006) also reported high genetic correlations (range: 0.89 to 0.92) between the most different SCM traits defined with thresholds of 50,000 and 200,000 cells/mL. These authors also reported a moderate residual correlation (range, 0.31 to 0.37) between SCM traits.

Moderate to high Spearman rank correlations were estimated among EBV of 6 traits predicted from linear and threshold models; therefore, we inferred that there would be likely re-ranking of sires based on the chosen trait. In contrast, very high rank correlations between EBV predicted for the same trait from linear versus threshold models suggested sire rankings may not differ between threshold versus linear models. Hence, linear models are robust and could be used in genetic evaluation of binary traits. There is a paucity of reports comparing threshold and linear models for SCM within 30 DIM. Similar to the present study, Heringstad et al. (2003) also estimated a high rank correlation (> 0.99) between sire evaluation based on threshold and linear models for clinical mastitis. This was in agreement with other studies that explored various conditions, including lameness (Malchiodi et al., 2017) and dystocia (Renand et al., 1990). However, it was stated that limited information per sire could affect sire rankings using threshold versus linear models (Negussie et al. 2008). Importantly, the present study demonstrated that, despite low heritability, there was substantial variation in sires' EBV. Therefore, this represented an opportunity to use breeding to improve genetic resistance to SCM in early lactation.

2.6. Conclusions

Despite relatively low heritability, there was substantial exploitable genetic variability among sires for SCM of heifers early in their first lactation. Hence, there is opportunity for improvement of heifer mastitis through long-term genetic selection. Moreover, high genetic correlations among analyzed 6 traits (SCM_{150}, SCM_{200}, SCM_{250}, $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$) indicated that these traits were genetically similar, irrespective of different thresholds. However, sire rankings may differ between traits used to define SCM. Therefore, selection of traits to define SCM should be done with caution. Hence, in future studies, estimating sensitivity and specificity are required to find the best threshold of SCC to define SCM. Regardless, estimated very high Spearman rank correlations between EBV predicted from both linear and threshold models indicated there will be no difference in sire rankings when using threshold versus linear models. Hence, a linear model could be used for routine genetic evaluation of SCM instead of threshold model in the dairy industry, as they are less computationally demanding and robust.

2.7. References

Alam, M., C. I. Cho, T. J. Choi, B. Park, J. G. Choi, Y. H. Choy, S. S. Lee, and K. H. Cho. 2015. Estimation of genetic parameters for somatic cell scores of Holsteins using multi-trait lactation models in Korea. Asian-Australalasian J. Anim. Sci. 28:303-310.

Barkema, H. W., H. A. Deluyker, Y. H. Schukken, and T.J.G.M. Lam. 1999. Quarter-milk somatic cell count at calving and at the first six milkings after calving. Prev. Vet. Med. 38:1–9.

Boland, F. and S. More. 2013. Investigating a dilution effect between somatic cell count and milk yield and estimating milk production losses in Irish dairy cattle. J. Dairy Sci. 96:1477-1484.

De Haas, Y., H. W. Barkema, Y. H. Schukken, and R. F. Veerkamp. 2003. Genetic associations for pathogen-specific clinical mastitis and patterns of peaks in somatic cell count. Anim. Sci. 77:187–195.

De Haas, Y., W. Ouweltjes, J. Ten Napel, J. J. Windig, and G. De Jong. 2008. Alternative somatic cell count traits as mastitis indicators for genetic selection. J. Dairy Sci 91:2501–2511.

Dempster, E. R., and I. M. Lerner. 1950. Heritability of threshold characters. Genetics 35:212–236.

De Vliegher, S., H. W. Barkema, H. Stryhn, G. Opsomer, and A. De Kruif. 2004. Impact of early lactation somatic cell count in heifers on somatic cell counts over the first lactation. J. Dairy Sci. 87:3672–3682.

De Vliegher, S., H. W. Barkema, H. Stryhn, G. Opsomer, and A. De Kruif. 2005. Impact of early lactation somatic cell count in heifers on milk yield over the first lactation. J. Dairy Sci. 88:938–947.

De Vliegher, S., L. K. Fox, S. Piepers, S. McDougall, and H. W. Barkema. 2012. Invited review: Mastitis in dairy heifers: Nature of the disease, potential impact, prevention, and control. J.

Dairy Sci. 95:1025–1040.

De Vliegher, S., H. Laevens, G. Opsomer, E. De Muêlenaere, and A. De Kruif. 2001. Somatic cell counts in dairy heifers during early lactation. Vlaams Diergeneeskd. Tijdschr. 70:212–215.

Emanuelson, U., and H. Funke. 1991. Effect of milk yield on relationship between bulk milk somatic cell count and prevalence of mastitis. J. Dairy Sci. 74:2479–2483.

Esposito, G., P. C. Irons, E. C. Webb, and A. Chapwanya. 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. Anim. Reprod. Sci. 144:60–71.

Gianneechini, R., C. Concha, R. Rivero, I. Delucci, and J. Moreno López. 2002. Occurrence of clinical and sub-clinical mastitis in dairy herds in the West Littoral region in Uruguay. Acta Vet. Scand. 43:221–230.

Gianola, D. 1982. Theory and analysis of threshold characters. J. Anim. Sci. 54:1079-1096.

Gianola, D., and J. L. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201-224.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml user guide release 3.0. VSN Int Ltd.

Hagger, C., and A. Hofer. 1990. Genetic analyses of calving traits in the Swiss Black and White, Braunvieh and Simmental breeds by REML and MAPP procedures. Livest. Prod. Sci. 24:93–107.

Hagnestam-Nielsen, C., U. Emanuelson, B. Berglund, and E. Strandberg. 2009. Relationship between somatic cell count and milk yield in different stages of lactation. J. Dairy Sci. 92:3124-3133.

Halasa, T., M. Nielen, A. P. W. De Roos, R. Van Hoorne, G. De Jong, T.J.G.M. Lam, T. Van

Werven, and H. Hogeveen. 2009. Production loss due to new subclinical mastitis in Dutch dairy cows estimated with a test-day model. J. Dairy Sci. 92:599–606.

Halasa, T., K. Huijps, and H. Hogeveen. 2007. Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29:18–31.

Harmon, R. J. 1994. Physiology of mastitis and factors affecting somatic cell counts. J. Dairy Sci. 77:2103–2112.

Haugaard, K., B. Heringstad, and A. C. Whist. 2013. Genetic associations between somatic cell score and pathogen-specific subclinical mastitis in Norwegian Red cows. J. Anim. Breed. Genet. 130:98–105.

Heringstad, B., A. Karlsen, and G. Klemetsdal. 1997. Preliminary results from a genetic analysis of clinical mastitis data. Interbull Bull. 15:45-49

Heringstad, B., G. Klemetsdal, and J. Ruan. 2000. Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. Livest. Prod. Sci. 64:95–106.

Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal, and K. A. Weigel. 2003. Genetic change for clinical mastitis in Norwegian cattle: a threshold model analysis. J. Dairy Sci. 86:369–375.

Hinrichs, D., E. Stamer, W. Junge, and E. Kalm. 2005. Genetic analyses of mastitis data using animal threshold models and genetic correlation with production traits. J. Dairy Sci. 88:2260–2268.

Hortet, P., F. Beaudeau, H. Seegers, and C. Fourichon. 1999. Reduction in milk yield associated with somatic cell counts up to 600 000 cells/mL in French Holstein cows without clinical mastitis. Livest. Prod. Sci. 61:33–42.

Huijps, K., T. J. Lam, and H. Hogeveen. 2008. Costs of mastitis: facts and perception. J. Dairy

Res. 75:113–120.

Jaeger, S., F. Virchow, P. Torgerson, M. Bischoff, B. Biner, S. Hartnack, and S. Rüegg. 2017. Test characteristics of milk amyloid A ELISA, somatic cell count, and bacteriological culture for detection of intramammary pathogens that cause subclinical mastitis. J. Dairy Sci 100:7419-7426.

Jamrozik. J., A. Koeck, F. Miglior, G. Kistemaker, F. Schenkel, D. Kelton, B. Van Doormaal. 2013. Genetic and genomic evaluation of mastitis resistance in Canada. Interbull Bull., 47: 43-51.

Kadarmideen, H. N., R. Thompson, and G. Simm. 2000. Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. Anim. Sci. 71:411–419

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012a. Health recording in Canadian Holsteins: Data and genetic parameters. J. Dairy Sci. 95:4099–4108.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012b. Alternative somatic cell count traits to improve mastitis resistance in Canadian Holsteins. J. Dairy Sci. 95:432–439.

Koivula, M., E. Negussie, and E. A. Mäntysaari. 2004. Genetic parameters for test-day somatic cell count at different lactation stages of Finnish dairy cattle. Livest. Prod. Sci. 90:145–157.

Laevens, H., H. Deluyker, Y. H. Schukken, L. De Meulemeester, R. Vandermeersch, E. De Muelenaere, and A. De Kruif. 1997. Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. J. Dairy Sci. 80:3219–3226.

Malchiodi, F., A. Koeck, S. Mason, A. M. Christen, D. F. Kelton, F. S. Schenkel, and F. Miglior. 2017. Genetic parameters for hoof health traits estimated with linear and threshold models using alternative cohorts. J. Dairy Sci. 100:2828-2836

Miller, R. H., M. J. Paape, L. A. Fulton, and M. M. Schutz. 1993. The relationship of milk somatic cell count to milk yields for Holstein heifers after first calving. J. Dairy Sci. 76:728–733.

Naqvi, S. A., J. De Buck, S. Dufour, and H. W. Barkema. 2018. Udder health in Canadian dairy heifers during early lactation. J. Dairy Sci. 101: 3233-3247.

Negussie, E., I. Strandén, and E. A. Mäntysaari. 2008. Genetic analysis of liability to clinical mastitis, with somatic cell score and production traits using bivariate threshold–linear and linear–linear models. Livest. Sci. 117:52–59.

Nyman, A. K., U. Emanuelson, K. Holtenius, K. L. Ingvartsen, K. L. Ingvartsen, and K. P. Waller. 2008. Metabolites and immune variables associated with somatic cell counts of primiparous dairy cows. J. Dairy Sci 91:2996–3009.

Pyörälä, S. 2003. Indicators of inflammation in the diagnosis of mastitis. Vet. Res 34:565–578.

Reichmuth, J., 1975. Somatic cell counting-Interpretation of results. In: Proc. Seminar on Mastitis Control, 7-11 April. International Dairy Federation, Brussels, 85:93-109

Renand, G., L. L. G. Janss, and J. Gaillard. 1990. Sire evaluation for direct effects on dystocia by linear and threshold models.. Proc. 4th World Congr. Genet. Appl. to Livest. Prod. Edinburgh 23-27 July 1990. XIII. Plenary Lect. Mol. Genet. mapping, Sel. Predict. Estim. 465–467.

Reyher, K. K., S. Dufour, H. W. Barkema, L. Des Côteaux, T. J. DeVries, I. R. Dohoo, G. P. Keefe, J.-P. Roy, and D. T. Scholl. 2011. The National Cohort of Dairy Farms—A data collection platform for mastitis research in Canada. J. Dairy Sci. 94:1616–1626.

Ruegg, P.L. 2017. A 100-year review: Mastitis detection, management, and prevention. J. Dairy Sci. 100:10381–10397.

Rupp, R., F. Beaudeau, and D. Boichard. 2000. Relationship between milk somatic-cell counts in the first lactation and clinical mastitis occurrence in the second lactation of French Holstein cows. Prev. Vet. Med. 46:99–111.

Saidi, R., D. Khelef, and R. Kaidi. 2013. Subclinical mastitis in cattle in Algeria: Frequency of occurrence and bacteriological isolates. J. S. Afr. Vet. Assoc. 84: 1-5.

Santman-Berends, I. and O. C. Sampimon. 2011. Subclinical mastitis in Dutch dairy heifers in early lactation and associated risk factors. Pages 325-330 in Udder health and communication. Wageningen Academic Publishers, Wageningen, Netherlands.

SAS Institute. 2013. SAS® 9.4 Guide to software updates. SAS Institute Inc., Cary, NC.

Schepers, A. J., T.J.G.M. Lam, Y. H. Schukken, J. B. M. Wilmink, and W. J. A. Hanekamp. 1997. Estimation of variance components for somatic cell counts to determine thresholds for uninfected quarters. J. Dairy Sci. 80:1833–1840.

Sheldrake, R. F., R. J. T. Hoare, and G. D. McGregor. 1983. Lactation stage, parity and infection affecting somatic cells, electrical conductivity and serum albumin in milk. J. Dairy Sci. 66:542-547.

Shook, G. E., and M. M. Schutz. 1994. Selection on somatic cell score to improve resistance to mastitis in the United States. J. Dairy Sci. 77:648–658.

Steeneveld, W., H. Hogeveen, H. W. Barkema, J. van den Broek, and R. B. M. Huirne. 2008. The influence of cow factors on the incidence of clinical mastitis in dairy cows. J. Dairy Sci. 91:1391–1402.

Svendsen, M., and B. Heringstad. 2006. Somatic cell count as an indicator of subclinical mastitis. Genetic parameters and correlations with clinical mastitis. Interbull Bull. 35:12-16.

Trinidad, P., S. C. Nickerson, and T. K. Alley. 1990. Prevalence of intramammary infection and teat canal colonization in unbred and primigravid dairy heifers. J. Dairy Sci. 73:107–114.

Urioste, J., J. Franzén, J. Windig, and E. Strandberg. 2012. Genetic relationships among mastitis and alternative somatic cell count traits in the first 3 lactations of Swedish Holsteins. J. Dairy

Sci. 95:3428–3434.

Wathes, D. C., Z. Cheng, W. Chowdhury, M. A. Fenwick, R. Fitzpatrick, D. G. Morris, J. Patton, and J. J. Murphy. 2009. Negative energy balance alters global gene expression and immune responses in the uterus of postpartum dairy cows. Physiol. Genomics. 39:1–13.

Table 2-1. Prevalence (%) of subclinical mastitis (SCM) in first-lactation heifers from 5 to 30 days in milk across Canadian provinces for alternative SCM traits[2]

	Provinces[1]						
Trait	AB	NB	NS	ON	PE	QC	Total
SCM_{150}	19.0	23.3	21.8	26.8	27.1	25.8	24.3
SCM_{200}	14.4	19.0	17.5	20.9	21.1	19.5	18.7
SCM_{250}	12.0	15.4	14.8	17.0	16.7	15.4	15.2
$SCM_{400,150}$	15.4	17.5	17.6	22.0	21.6	20.6	19.6
$SCM_{400,200}$	11.9	15.1	14.8	17.9	17.6	16.6	15.9
$SCM_{400,250}$	10.4	12.9	12.9	15.0	14.2	14.1	13.4

[1]AB = Alberta, NB = New Brunswick, NS = Nova Scotia, ON = Ontario, PE = Prince Edward Island, QC = Quebec

[2]SCM_{150}, SCM_{200} and SCM_{250} = presence of SCC > 150,000, 200,000 and 250,000 cells/mL within $5 \leq DIM \leq 30$, respectively. $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ = presence of SCC > 400,000 cells/mL within $5 \leq DIM \leq 10$ and SCC > 150,000, 200,000 and 250,000 cells/mL, respectively, between $11 \leq DIM \leq 30$.

Table 2-2. Heritability[1] on the underlying scale (h_{und}^2) and heritability transformed into an observed scale through the Dempster and Lerner (1950) method ($h_{obs\text{-}t}^2$) from univariate probit threshold animal model, and heritability on the observed scale ($h_{obs\text{-}l}^2$) from univariate linear animal model for alternative SCM traits[2]. Standard errors are provided in parentheses.

Trait	h_{und}^2	$h_{obs\text{-}t}^2$	$h_{obs\text{-}l}^2$
SCM_{150}	0.046 (0.017)	0.024	0.057 (0.018)
SCM_{200}	0.051 (0.018)	0.025	0.053 (0.017)
SCM_{250}	0.040 (0.019)	0.017	0.037 (0.015)
$SCM_{400,150}$	0.047 (0.018)	0.023	0.051 (0.017)
$SCM_{400,200}$	0.047 (0.019)	0.021	0.048 (0.016)
$SCM_{400,250}$	0.043 (0.020)	0.017	0.040 (0.015)

[1]Heritability (h^2) for each trait was calculated from univariate analysis as follows:

$$h^2 = \sigma_a^2 / (\sigma_{hy}^2 + \sigma_a^2 + \sigma_e^2)$$

where σ_{hy}^2, σ_a^2, and σ_e^2 are the herd-year of calving, animal additive genetic, and residual variances, respectively.

[2]SCM_{150}, SCM_{200} and SCM_{250} = presence of SCC > 150,000, 200,000 and 250,000 cells/mL within $5 \leq DIM \leq 30$, respectively. $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ = presence of SCC > 400,000 cells/mL within $5 \leq DIM \leq 10$ and SCC > 150,000, 200,000 and 250,000 cells/mL, respectively, between $11 \leq DIM \leq 30$.

Table 2-3. Genetic correlations (above the diagonal; SE in parentheses) and phenotypic correlations (below the diagonal; SE in parentheses) estimated from bivariate linear animal models for alternative SCM traits[1]

Trait	SCM_{150}	SCM_{200}	SCM_{250}	$SCM_{400,150}$	$SCM_{400,200}$	$SCM_{400,250}$
SCM_{150}		0.998 (0.017)	0.941 (0.052)	0.906 (0.046)	0.932 (0.052)	0.909 (0.069)
SCM_{200}	0.848 (0.003)		0.991 (0.019)	0.990 (0.028)	0.988 (0.016)	0.973 (0.031)
SCM_{250}	0.748 (0.005)	0.882 (0.002)		0.949 (0.054)	0.993 (0.025)	0.976 (0.020)
$SCM_{400,150}$	0.871 (0.003)	0.787 (0.004)	0.731 (0.005)		0.996 (0.015)	0.944 (0.043)
$SCM_{400,200}$	0.767 (0.005)	0.905 (0.002)	0.839 (0.003)	0.8792 (0.003)		0.995 (0.013)
$SCM_{400,250}$	0.695 (0.006)	0.820 (0.004)	0.929 (0.002)	0.797 (0.004)	0.907 (0.002)	

[1]SCM_{150}, SCM_{200} and SCM_{250} = presence of SCC > 150,000, 200,000 and 250,000 cells/mL within $5 \leq DIM \leq 30$, respectively. $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ = presence of SCC > 400,000 cells/mL within $5 \leq DIM \leq 10$ and SCC > 150,000, 200,000 and 250,000 cells/mL, respectively, between $11 \leq DIM \leq 30$.

Table 2-4. Residual correlations (above the diagonal; SE in parentheses) estimated from bivariate linear animal models for alternative SCM traits[1]

Trait	SCM_{200}	SCM_{250}	$SCM_{400, 150}$	$SCM_{400, 200}$	$SCM_{400, 250}$
SCM_{150}	0.836	0.7330	0.867	0.7530	0.680
	(0.004)	(0.007)	(0.004)	(0.006)	(0.008)
SCM_{200}		0.876	0.772	0.899	0.810
		(0.003)	(0.006)	(0.003)	(0.005)
SCM_{250}			0.715	0.826	0.926
			(0.007)	(0.005)	(0.002)
$SCM_{400,150}$				0.871	0.785
				(0.003)	(0.006)
$SCM_{400,200}$					0.900
					(0.003)

[1]SCM_{150}, SCM_{200} and SCM_{250} = presence of SCC > 150,000, 200,000 and 250,000 cells/mL within $5 \leq DIM \leq 30$, respectively. $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ = presence of SCC > 400,000 cells/mL within $5 \leq DIM \leq 10$ and SCC > 150,000, 200,000 and 250,000 cells/mL, respectively, between $11 \leq DIM \leq 30$.

Table 2-5. Rank correlations between estimated breeding values predicted using linear (above the diagonal) and threshold animal models (below the diagonal) for alternative SCM traits[1] for 69 sires with at least 20 daughters in 5 herds

Trait	SCM_{150}	SCM_{200}	SCM_{250}	$SCM_{400,150}$	$SCM_{400,200}$	$SCM_{400,250}$
SCM_{150}		0.92	0.86	0.86	0.84	0.75
SCM_{200}	0.92		0.95	0.85	0.92	0.85
SCM_{250}	0.86	0.94		0.84	0.90	0.91
$SCM_{400,150}$	0.88	0.85	0.86		0.94	0.90
$SCM_{400,200}$	0.84	0.92	0.92	0.94		0.95
$SCM_{400,250}$	0.74	0.83	0.91	0.90	0.95	

[1]SCM_{150}, SCM_{200} and SCM_{250} = presence of SCC > 150,000, 200,000 and 250,000 cells/mL within $5 \leq DIM \leq 30$, respectively. $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ = presence of SCC > 400,000 cells/mL within $5 \leq DIM \leq 10$ and SCC > 150,000, 200,000 and 250,000 cells/mL, respectively, between $11 \leq DIM \leq 30$.

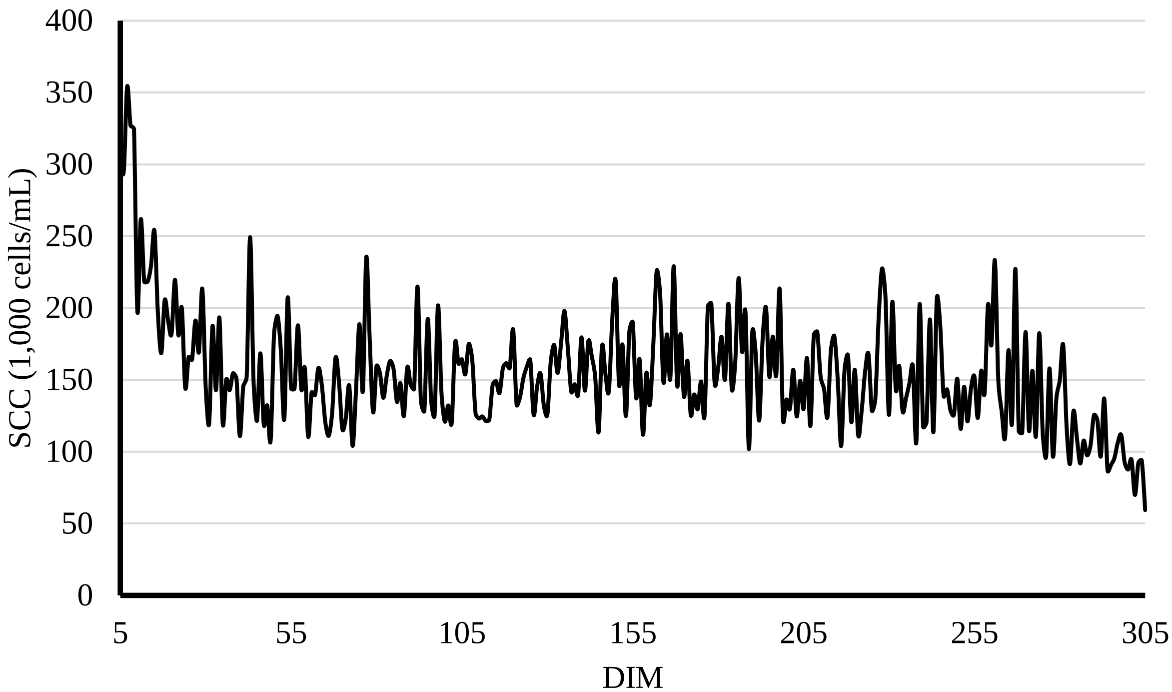
Table 2-6. Percentage of daughters with subclinical mastitis from all (average), the best (10% decile), and the worst (90% decile) sires (with at least 20 daughters in > 5 herds (n = 69)) according to their estimated breeding value for resistance to subclinical mastitis predicted with univariate linear and probit threshold models for alternative SCM traits[1] within 5 to 30 DIM

		Linear model		Threshold model	
Trait	Average	10% decile	90% decile	10% decile	90% decile
SCM_{150}	17.9	13.6	33.3	14.9	33.3
SCM_{200}	14.2	10.1	28.5	11.0	27.5
SCM_{250}	18.4	5.7	22.4	7.2	20.9
$SCM_{400,150}$	18.4	10.3	28.9	11.2	28.7
$SCM_{400,200}$	14.9	6.6	23.2	7.6	23.2
$SCM_{400,250}$	12.5	5.6	19.5	5.6	18.2

[1]SCM_{150}, SCM_{200} and SCM_{250} = presence of SCC > 150,000, 200,000 and 250,000 cells/mL within $5 \leq DIM \leq 30$, respectively. $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ = presence of SCC > 400,000 cells/mL within $5 \leq DIM \leq 10$ and SCC > 150,000, 200,000 and 250,000 cells/mL, respectively, between $11 \leq DIM \leq 30$.

Figure 2-1. Mean somatic cell count (SCC) in first-lactation Holstein heifers from a) 5 to 305 DIM; b) and from 5 to 30 DIM.

(a)



(b)

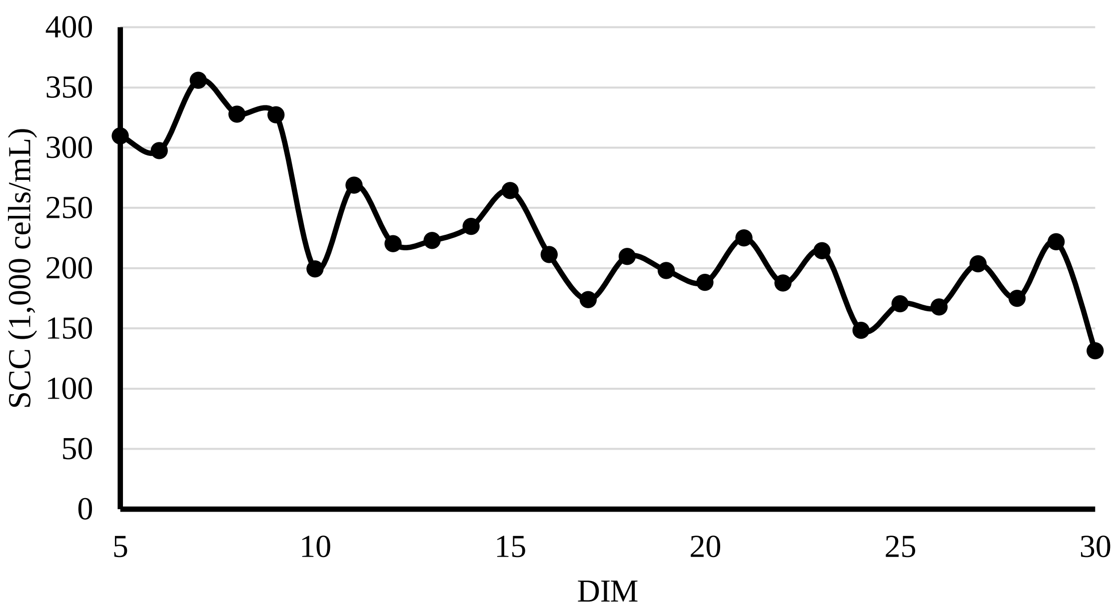
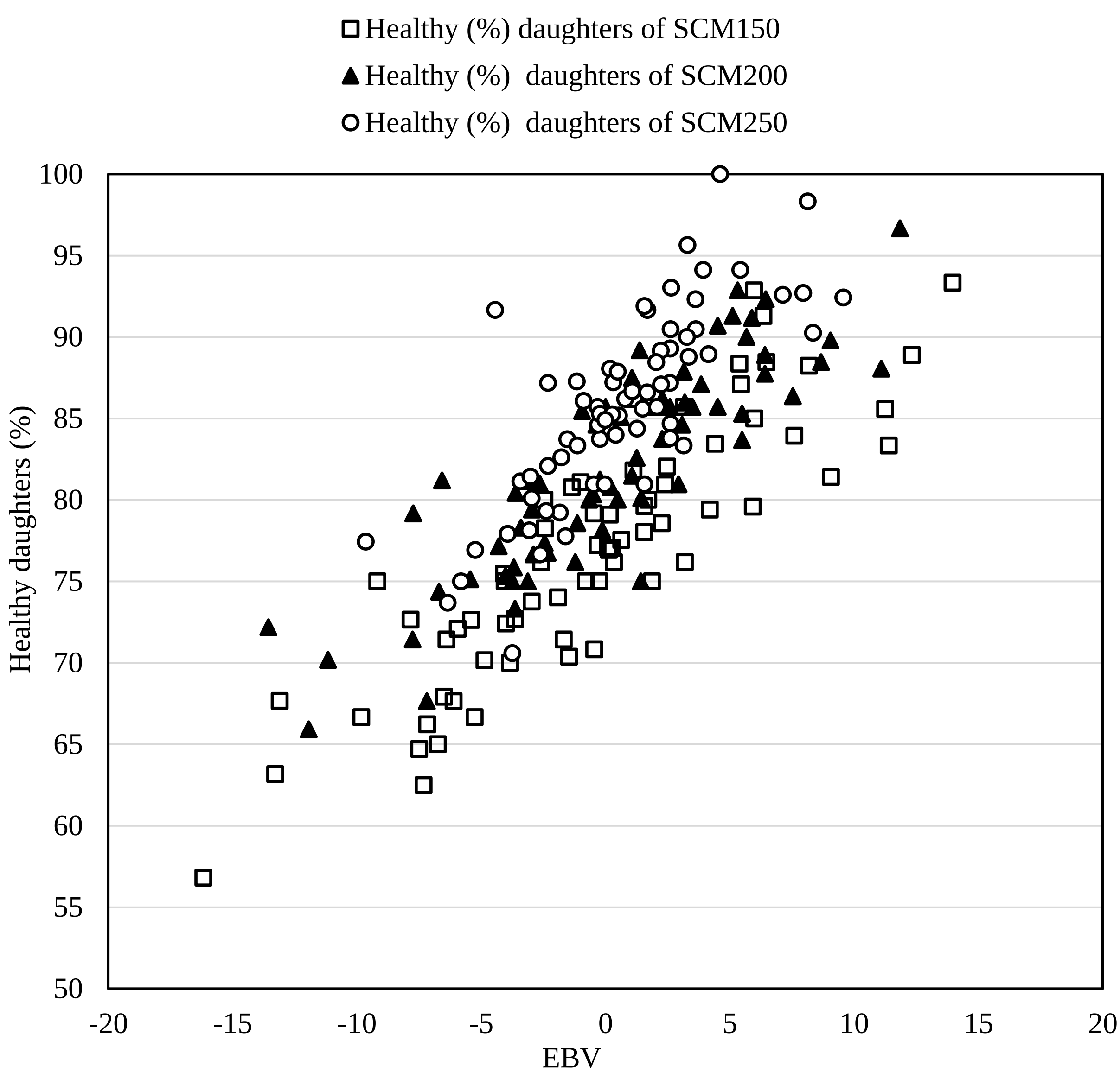
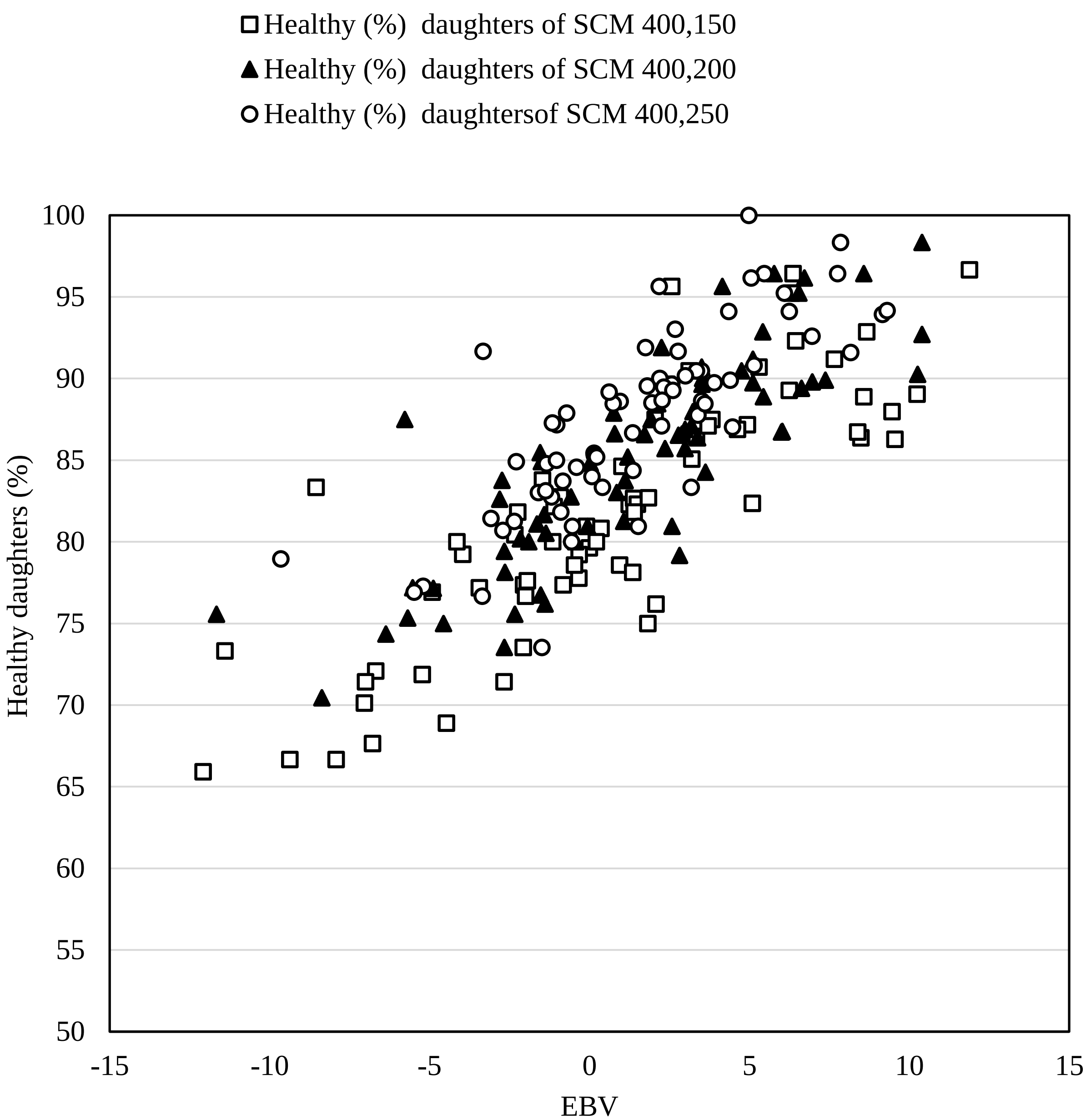


Figure 2-2. Percentage of healthy daughters according to the estimated breeding value (EBV) estimated using linear model of subclinical mastitis resistance of sires (within 5 to 30 DIM) with at least 20 daughters in > 5 herds (n = 69) for (a) SCM_{150}, SCM_{200} and SCM_{250} = presence of SCC > 150,000, 200,000 and 250,000 cells/mL within 5 ≤ DIM ≤ 30, respectively, and (b) $SCM_{400,150}$, $SCM_{400,200}$ and $SCM_{400,250}$ = presence of SCC > 400,000 cells/mL within 5 ≤ DIM ≤ 10 and SCC > 150,000, 200,000 and 250,000 cells/mL, respectively, between 11 ≤ DIM ≤ 30.

(a)



(b)



Healthy (%) daughters of SCM 400,150
Healthy (%) daughters of SCM 400,200
Healthy (%) daughtersof SCM 400,250
Healthy daughters (%)
EBV
100
95
90
85
80
75
70
65
60
55
50
-15
-10
-5
0
5
10
15

CHAPTER 3: Genetic analysis of pathogen-specific intramammary infections in dairy cows

3.1. Abstract

Mastitis is one of the most common diseases in dairy cattle, causing severe economic losses to dairy farmers. Mastitis usually occurs due to intramammary infection (IMI) caused by a variety of pathogenic bacteria. Although good progress has been made in understanding genetics of pathogen-specific clinical mastitis, studies involving genetic analysis of pathogen-specific IMI are scarce. The overall objective of this study was, therefore, to assess genetic variation of overall and pathogen-specific IMI in non-clinical primiparous and multiparous cows using bacterial culture. Data and milk samples were collected over a 2-year interval as part of the Canadian Bovine Mastitis Research Network. The final dataset contained records of 46,900 quarter milk samples from 3,382 clinically healthy primiparous and multiparous Holstein cows from 84 dairy herds. For the genetic analysis, we considered the following 7 traits: overall IMI, non-*aureus* staphylococci (NAS) IMI, contagious pathogen IMI, environmental pathogen IMI, major pathogen IMI, minor pathogen IMI and somatic cell score (SCS). Data were analyzed at the quarter level using a threshold probit model via Gibbs sampling in BLUPF90. Prevalence of IMI traits at quarter-level in multiparous cow from 0-400 DIM ranged from 6.8 to 45.5%. Posterior mean of quarter heritability estimates (on the underlying scale, posterior SD in brackets) of overall IMI and pathogen-specific IMI traits ranged from 0.017 to 0.073 (±0.009 to 0.030). Weak to strong genetic correlations (ranging from 0.18 to 0.97 (±0.01 to 0.29)) among pathogen-specific IMI traits and with overall IMI indicated that not all of these traits were genetically similar. Weak to moderate Spearman rank correlations between EBVs for overall IMI and pathogen-specific IMI traits (from 0.31 to 0.87) indicated possible substantial re-ranking of sires. The percentage of daughters with

IMI caused by various pathogen groups ranged from 13 to 80% and from 38 to 94% for the best (10% decile) and worst sires (90% decile) according to their IMI trait-specific EBV, respectively. Pathogen-specific IMI traits and overall IMI had weak to moderate positive genetic correlations (ranging from 0.11 to 0.81 (±0.11 to 0.22)) with SCS. Therefore, selection for lower SCS will improve resistance to IMI. However, based on the observed weak to moderate rank correlations (0.04 to 0.47) between pathogen-specific IMI traits and SCS, selection for lower SCC will not improve resistance to IMI from every pathogen-specific IMI group in the same manner. Therefore, despite low heritability estimates, there was sizeable genetic variation for pathogen-specific IMI traits, indicating that long-term direct genetic selection for pathogen-specific IMI can improve pathogen-specific IMI resistance.

3.2. Introduction

Mastitis remains, a challenging disease in dairy cattle, causing considerable economic losses to dairy producers and the industry as a whole. For a typical Canadian dairy farm, cost associated with mastitis was $66,178 CAD/100 cows-year (Aghamohammadi et al., 2018). The cost associated with mastitis differs, however, based on the mastitis pathogen involved, with estimated costs per clinical mastitis case of €231, 570, 380 and 206 for overall mastitis, *Staphylococcus aureus*, non-*aureus* staphylococci (**NAS**) and *Escherichia coli*, respectively (Sørensen et al. 2010). Milk yield reduction caused by *E. coli*, *S. aureus* and NAS CM was estimated at 3.5, 2.3 and 1.8 kg/d, respectively (Heikkilä et al., 2018).

Worldwide, the prevalence of intramammary infection, determined in various stages of lactation and parities, is high and ranges from 20 to 75% at quarter-level and 49 to 97% at cow-level (Oliver et al., 2003; Sampimon et al., 2009; Piepers et al., 2011). At calving, 12 to 57% of udder quarters of heifers harbor an IMI (Cook et al., 1992; Parker et al., 2007). In a study in Belgium, 41% of 44,677 cows had at least 1 culture-positive quarter (Piepers et al., 2007).

Whereas intramammary infection (**IMI**) refers to the presence of bacteria in the udder (Berry and Meaney, 2006; Andersen et al., 2010), mastitis is an inflammation of the udder, which can be subclinical (**SCM**) or clinical (**CM**). Distribution of pathogens differs between IMI and CM (Fox, 2009); major pathogens such as *E. coli* and *S. aureus* are most frequently involved in CM (Sørensen et al., 2009b), whereas NAS, a group of minor pathogens, is the most prevalent cause of IMI (Dufour et al., 2012; Piepers et al., 2011; De Visscher et al., 2016). In developed dairy countries with a relatively low bulk tank SCC, > 50% of all IMI are caused by NAS (e.g. Piepers et al., 2007; Fox, 2009). In a Canadian study, prevalence of quarter-level NAS IMI was 26% and NAS prevalence was higher in heifers than in multiparous cows (36 versus 23%, respectively; Condas et al., 2017b).

Selecting and breeding for cows that have a high resistance to IMI, together with better management, have potential to be an excellent strategy to combat IMI. Although good progress has been made in comprehension of genetics of pathogen-specific CM (De Haas et al., 2002a,b; Sørensen et al., 2009a; Haugaard et al., 2012; Holmberg et al., 2012), studies involving genetic analysis of pathogen-specific IMI are scarce (Wanner et al., 1998; Schafberg et al., 2006). Wanner et al. (1998) reported heritabilites that ranged from 0.02 to 0.66 for pathogen-specific IMI incidence. However, the SE around these estimations were large, due to a small sample size (756 first-lactation cows). Moreover, genetic correlations among pathogen-specific CM and between SCC (or SCS) and pathogen-specific CM were moderate to strong and positive (De Haas et al., 2003; Haugaard et al., 2012). Because costs associated with mastitis vary based on causative pathogens (Sørensen et al., 2010), breeding cows based on a pathogen-specific udder-health index would be advantegeous.

A large cohort study conducted during 2007 and 2008 by the Canadian Bovine Mastitis Research Network (CBMRN) provides an excellent opportunity to estimate genetic variability of IMI in clinically healthy cows at quarter-level and to address knowledge gaps. Therefore, objectives of this study were to: 1) estimate genetic parameters for overall and pathogen-specific IMI and SCS in non-clinical primiparous and multiparous cows; and 2) determine genetic variation among sires for IMI related traits.

3.3. Material and Methods

3.3.1. Data

Rationale and details regarding selection of farms and data collection were described (Reyher et al., 2011). Milk samples were collected in 3 independent sampling periods and 1 intensive sampling period from lactating cows that had no CM at the first sampling of the series. Independent sampling periods consisted of 3 visits per 3 weeks and were conducted during spring 2007 (March-May), winter 2008 (January-March), and summer 2008 (June-August). There was an intensive sampling period (7 weekly samplings) during summer 2007 (June-August). Quarter milk samples were collected by trained technicians from 15 cows per period in 91 Canadian dairy herds participating in the National Cohort of Dairy Farms of the CBMRN. To represent inherent diversity of Canadian dairy herds, commercial dairy farms were selected from 6 provinces: Alberta, Ontario, Québec, Prince Edward Island, Nova Scotia, and New Brunswick. All participating herds had > 80% Holstein cows, milked twice per day, and participated in a regular DHI recording system.

The initial dataset contained 80,340 records of 3,773 Holstein cattle from 91 Canadian dairy herds. Contaminated and composite milk samples (combined milk from all 4 quarters) were excluded from the dataset. For the genetic analysis, only Lactations 1 to 5 with specific age limits at time of the milk sampling (570-1,790, 900-2,120, 1,230-2,450, 1,560-2,780, and 1,890-3,110 days for Lactations 1-5, respectively) were considered. Only test-day records from 0 to 400 DIM were included in the analyses. Additionally, culture and SCC results were included if milk samples were available from all 4 quarters and the interval between milk sample collections was ≥ 7 days. Furthermore, herds with < 20 registered Holstein cows were excluded from the analysis. The final dataset contained 46,900 quarter-level records of 3,382 primiparous and multiparous Holstein cows from 84 herds. The pedigree file was obtained from the Canadian Dairy Network (Guelph, ON, Canada) and contained a total of 26,109 animals, generated by tracing back 7 generations.

3.3.2. Definition of IMI Traits

A quarter was considered having an IMI if ≥ 1 colony was cultured from 10 μL of inoculum in mixed or pure culture (Dohoo et al., 2011b) and the quarter was assigned infected (coded as 1). If no bacterial pathogen was isolated, a quarter was assigned uninfected (coded as 0). A milk sample was considered contaminated if ≥ 3 species of bacteria were cultured. Moreover, a cow was considered to have IMI if ≥ 1 quarter was infected (coded as 1); if not, a cow was coded as 0 (uninfected).

Pathogens were grouped into 6 categories based on transmission mode, coagulase test and effect on udder. Traits were overall IMI (all udder pathogens), contagious pathogens (*S. aureus*, *Streptococcus dysgalactiae, Streptococcus agalactiae* or *Corynebacterium* spp.), environmental pathogens (*E. coli, Klebsiella* spp., *Enterobacter* spp., *Nocardia* spp., *Prototheca* spp., *Trueperella pyogenes*, *Streptococcus uberis*, *Bacillus* spp., *Streptococcus* spp., *Serratia* spp., *Citrobacter* spp., *Proteus* spp., *Salmonella* spp., *Pseudomonas* spp., or *Pasteurella multocida*), NAS, major pathogens (*S. aureus, Strep. agalactiae, Strep. dysgalactiae, Streptococcus* spp., *Strep. uberis, E. coli, Klebsiella* spp., *Enterobacter* spp., or *Trueperella pyogenes*) and minor pathogens (*Corynebacterium* spp., NAS). As IMI by up to 2 pathogens was considered possible, a single quarter could be included in multiple pathogen groups. All pathogen-specific IMI traits were treated as binary.

Another trait included in the analysis was quarter-level SCS. In order to achieve normal distribution, somatic cell count was converted to SCS with the following formula: SCS = $\log_2$(SCC/100,000)+3 (Shook, 1993).

3.3.3. Statistical Models

Intramammary infection related binary traits were analyzed using a probit-threshold mixed model, whereas SCS was analyzed using a linear animal mixed model. In theory, threshold models are more appropriate for analysis of binary response traits, as they account for the binary nature of the data (Gianola, 1982; Gianola and Foulley, 1983). In a threshold model, existence of an underlying continuous non-observable variable, called liability (λ), is assumed. If λ is larger than a fixed threshold (τ), the observed binary response takes the value of 1, otherwise it is 0. Given the mean and variance, the liability is assumed to be normally distributed. For binary data, as threshold (τ) and the residual variance (σ^2_e) are not identifiable, these parameters are assigned arbitrary values of 0 and 1, respectively.

Heritabilities and breeding values for all IMI-related binary traits were estimated using a Bayesian approach (Markov Chain Monte Carlo (MCMC)) via Gibbs sampling algorithm using THRGIBBS1F90 application of BLUPF90 suite of programs (Misztal, 2002; Tsuruta and Misztal, 2006). The following repeatability animal probit-threshold model was used:

$$\boldsymbol{\lambda} = \mathbf{X\beta} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_{htd}\mathbf{htd} + \mathbf{Z}_{pew}\mathbf{pew} + \mathbf{Z}_{pea}\mathbf{pea} + \mathbf{e}$$

where $\boldsymbol{\lambda}$ is a vector of underlying liabilities corresponding to the binary observations in y (0= uninfected and 1= infected); $\boldsymbol{\beta}$ is a vector of systematic effects, including fixed effects for DIM (13 classes; classes of 30 d each with the last class from 361- 400 DIM) and parity (5 classes); a is a vector of random additive genetic animal effects (26,109 levels); **htd** is a vector of random herd-sampling date effects (1,153 levels); **pew** is a vector of random within-lactation permanent environmental effect (3,382 levels); **pea** is a vector of random across-lactation permanent environmental effect (3,777 levels); **e** is a vector of random residuals; and $\mathbf{X}$, $\mathbf{Z}_a$ $\mathbf{Z}_{htd}$, $\mathbf{Z}_{pew}$ and $\mathbf{Z}_{pea}$ are corresponding incidence matrices.

Heritability and breeding values for SCS were estimated using a similar, but linear repeatability animal model via Gibbs sampling algorithm in BLUPF90 using THRGIBBS1F90 application.

Genetic correlations among IMI-related traits and, between IMI-related traits and SCS were estimated using bivariate analysis; these analyses were carried out in BLUPF90 using THRGIBBS1F90 application. All random effects were assumed to be independent of each other within trait but correlated between traits.

For all univariate and bivariate analyses, Gibbs sampler was run for 1,000,000 rounds, with the first 200,000 rounds excluded as a burning-in period. A thinning interval of 100 rounds was used to reduce the lag correlation among thinned samples. Therefore, genetic parameters and respective posterior standard deviations were calculated using the 8,000 post-Gibbs effective samples. Initial test of significance for fixed effects included in the models were performed in SAS (SAS Institute Inc., 2013).

Heritability (h^2) for each trait at quarter-level was calculated from univariate analysis as follows:

$$h^2 = \frac{\sigma_{\mathrm{a}}^2}{\sigma_{\mathrm{a}}^2 + \sigma_{\mathrm{htd}}^2 + \sigma_{\mathrm{pew}}^2 + \sigma_{\mathrm{pea}}^2 + \sigma_{\mathrm{e}}^2}$$

Threshold model heritability estimates (per quarter) were converted to heritability estimates on the observed scale (Dempster and Lerner, 1950), as follows:

$$h_{obs}^2 = \frac{h_{und}^2 \times (z_{ord})^2}{p \times (1 - p)}$$

where h_{obs}^2, h_{und}^2, z_{ord}, and p are the heritability estimate on the observed scale, heritability estimated on the underlying scale, ordinate height of the normal distribution at the threshold point corresponding to p, and p is the prevalence of IMI-related traits.

Repeatability within-lactation (Rep_w) was calculated as follows:

$$\mathrm{Rep}_w = \frac{\sigma_{\mathrm{a}}^2 + \sigma_{\mathrm{pew}}^2 + \sigma_{\mathrm{pea}}^2}{\sigma_{\mathrm{a}}^2 + \sigma_{\mathrm{htd}}^2 + \sigma_{\mathrm{pew}}^2 + \sigma_{\mathrm{pea}}^2 + \sigma_{\mathrm{e}}^2}$$

Repeatability across-lactation (Rep_a) was calculated as follows:

$$\mathrm{Rep}_a = \frac{\sigma_{\mathrm{a}}^2 + \sigma_{\mathrm{pea}}^2}{\sigma_{\mathrm{a}}^2 + \sigma_{\mathrm{htd}}^2 + \sigma_{\mathrm{pew}}^2 + \sigma_{\mathrm{pea}}^2 + \sigma_{\mathrm{e}}^2}$$

where $\sigma_{\mathrm{a}}^2, \sigma_{\mathrm{htd}}^2, \sigma_{\mathrm{pew}}^2, \sigma_{\mathrm{pea}}^2$, and σ_{e}^2 are the animal additive genetic, herd-sampling date, within-lactation permanent environmental, across-lactation permanent environmental, and residual variances, respectively.

Genetic correlations (r_g) were calculated as follows:

$$r_g = \frac{\mathrm{Cov}(t_1, t_2)}{\sqrt{v1 \times v2}}$$

where Cov is the additive genetic covariance between trait 1 (t_1) and trait 2 (t_2), and $v1$ and $v2$ are the additive genetic variance for trait 1 and trait 2, respectively. Phenotypic correlations within lactation were estimated using correlations and variance ratios for the random effects in the model, except for the across-lactation permanent environmental effect (Falconer and Mackay, 1996).

Spearman rank correlations among estimated breeding values (EBV) for the 6 IMI traits, and between EBV for SCS and for the 6 IMI traits at quarter-level were calculated using SAS (SAS Institute Inc., 2013). A total 55 sires with ≥ 10 daughters in ≥ 5 herds were included in this part of the analysis. For estimation of percentage of uninfected daughters, only first records from 0 to 305 DIM were used. Moreover, a weighted regression of percentage of uninfected daughters on sires' EBV was performed, using total number of daughters per sire as the weighting factor in R (Version 3.4.4, R Core Team., 2018).

3.4. Results

3.4.1. Prevalence

Prevalence of IMI traits at quarter-level in multiparous cows from 0-400 DIM ranged from 6.8 (for major pathogens) to 45.5% (for overall IMI) (Table 3-1). Among pathogen-specific IMI traits, quarter-level prevalence of NAS was highest (30.8%), followed by environmental pathogens (14.0%). Contagious pathogen IMI had the lowest quarter-level prevalence (7.5%). Minor pathogen IMI were most prevalent (34.0%) at the quarter level compared to major pathogen IMI (7.0%). Distribution of pathogen groups across provinces was similar, where NAS and minor pathogen IMI had the highest quarter-level prevalence, whereas contagious pathogen IMI had the lowest prevalence. Across provinces, quarter-level prevalence of overall IMI was highest in Ontario (53.0%), followed by Alberta (49.0%), Quebec (45.0%), Nova Scotia (33.0%), and New Brunswick (32.0%), whereas Prince Edward Island had the lowest overall IMI prevalence (30.0%).

3.4.2. Genetic Parameters of IMI

Posterior mean of heritability estimates (± posterior SD) per quarter of all analyzed IMI traits using the threshold model on the underlying scale ranged from 0.017 to 0.073 (±0.009 to 0.030; Table 3-2). Heritability of overall IMI was 0.041 (±0.013) on the underlying scale. Among pathogen-specific IMI traits, contagious pathogens (0.073) had the highest heritability followed by NAS (0.035). Lowest heritability estimate was for environmental pathogen IMI (0.017). Heritability estimates for major pathogens (0.043) and minor pathogen IMI (0.041) were similar.

Converted heritability (per quarter) estimates on the observed scale ranged from 0.007 to 0.026. The highest heritability on the observed scale was estimated for overall IMI (0.026), followed by contagious pathogens (0.021) and NAS IMI (0.020), respectively. Environmental pathogen IMI had the lowest heritability estimate (0.007). When heritability estimates were converted from underlying scale to observed scale, minor pathogen IMI (0.025) had higher heritability than major pathogen IMI (0.012). Moreover, estimated heritability (per quarter) for SCS on the observed linear scale was 0.047 (±0.014).

Repeatability within lactation (Rep_w) for all traits was higher (0.09 to 0.40) compared to repeatability across lactation (Rep_a) (0.05 to 0.30). Contagious pathogen IMI had the highest Rep_w and Rep_a. Similarly, major pathogen IMI had a higher Rep_w and Rep_a than minor pathogen IMI. Repeatability of SCS within and across lactations was 0.29 and 0.20, respectively. Furthermore, repeatability estimations were higher than heritability estimations for all traits.

Posterior mean of genetic correlation estimates (± posterior SD) were positive among all analyzed traits (Table 3-3). Genetic correlation between overall IMI and other pathogen-specific IMI traits were moderate to strong (ranged from 0.67 to 0.90 (±0.05 to 0.014)) except for environmental pathogen IMI (0.18 ±0.25). A strong genetic correlation of 0.97 (±0.01) was estimated between NAS and minor pathogen IMI. Likewise, contagious pathogen IMI had strong genetic correlation (0.91 ±0.09) with major pathogen IMI. A moderate genetic correlation (0.30 ±0.25) was estimated between major and minor pathogen IMI. Pathogen-specific IMI traits had weak to moderate genetic correlations (ranging from 0.11 to 0.81 (±0.11 to 0.22) with SCS. Contagious pathogen IMI had the strongest genetic correlation of 0.81 (±0.11) with SCS.

Phenotypic correlations (within lactation) among IMI and pathogen-specific IMI traits ranged from -0.02 to 0.91 (Table 3-3). Intramammary infection caused by NAS and minor pathogens had the strongest phenotypic correlations with overall IMI (0.90 and 0.91, respectively). Intramammary infection caused by NAS had the strongest phenotypic correlation with minor pathogen IMI (0.92). Overall IMI and pathogen-specific IMI traits had weak to moderate phenotypic correlations (0.08 to 0.28) with SCS. Within pathogen groups, contagious and major pathogen IMI had the strongest phenotypic correlations with SCS (0.24 and 0.28), respectively.

3.4.3. Prediction of Breeding Values

Weak to moderate (0.31 to 0.87) Spearman rank correlations were estimated between EBV for overall IMI and pathogen-specific IMI traits (Table 3-4). Non-*aureus* staphylococci (0.83) and minor pathogens IMI (0.87) had the strongest rank correlation with overall IMI. Moreover, rank correlation of 0.96 and 0.53 were estimated between NAS and minor pathogen IMI, and contagious and major pathogen IMI, respectively. In addition, overall IMI and pathogen-specific IMI traits had weak to moderate (0.04 to 0.47) rank correlations with SCS. Contagious pathogen IMI had the strongest rank correlation with SCS (0.47) among the pathogen-specific IMI traits analyzed.

Despite low heritability, substantial variation in sires' EBV were observed (Table 3-5). Estimated breeding values were reversed in sign to relate higher EBVs to sires with more uninfected daughters. Mean percentage of daughters with IMI from various pathogen groups ranged from 24 to 85%. Of 55 sires with ≥ 10 daughters in ≥ 5 herds, percentage of daughters with IMI caused by various pathogen groups ranged from 13 to 80% and from 38 to 94% for the best (10% decile) and worst sires (90% decile), according to their trait-specific EBV, respectively. Although heritability of overall IMI and pathogen-specific IMI traits was low, variation in percentage of healthy daughters per sires' EBV seemed exploitable (Figure 2-1). Variation was the highest for contagious and major pathogen IMI.

3.5. Discussion

Prevalence estimated for overall IMI at quarter-level (45.5%) and pathogen-specific IMI in the current study were within the range of reported IMI prevalence. Piepers et al. (2007) reported a lower prevalence of 17.0% of IMI at quarter-level in dairy cows, whereas in Dutch dairy cows, slightly lower prevalence of 37.0 and 21.0% bacterial growth in quarter milk samples of high and low SCC cows were reported (Sampimon et al., 2009a). However, da Costa et al. (1996) estimated a higher quarter-level prevalence of 57.0% for overall IMI in heifers. Futhermore, prevalence of IMI is higher in late gestation, at calving and in early lactation in heifers (Oliver et al., 2003; Piepers et al., 2011). Quarter-level prevalences of IMI were 76.0, 47.0 and 29.0% at 7 days before calving and at 3 and 10 DIM, respectively (Oliver et al., 2003). Differences in reported prevalence of overall IMI may also be due to differences in definition of IMI, with varying sensitivity and specificity. Moreover, in agreement with other studies (Middleton et al., 2005; Sampimon et al., 2009b; Condas et al., 2017a), prevalence of NAS IMI (30.8%) was higher compared to other pathogen groups in the current study.

Our objective was to assess the genetic variation of IMI at quarter-level. Most of the genetic parameters for overall and pathogen-specific IMI have been reported at cow-level using either composite or quarter milk samples (e.g. Weller et al., 1992; Wanner et al., 1998;Schafberg et al., 2006). Use of composite milk samples is considered more economical than quarter milk samples. However, composite milk samples will reduce sensitivity by diluting the milk of infected quarters with milk from healthy quarters (Reyher and Dohoo, 2011), resulting in more false-negative cows and affecting genetic parameter estimations. Identification of new IMI is also less accurate when a composite milk sample is used (Reyher and Dohoo, 2011). Hence, estimating genetic parameters at quarter-level will be more beneficial.

Reports on estimated genetic parameters for pathogen-specific IMI of clinically healthy heifers and cows for mastitis are scarce. Furthermore, no heritability estimates on either the underlying or observed scales have been previously reported at quarter-level for overall IMI, as most reported heritability estimates have been determined on a cow-level. A quarter-level heritability of 0.041 on the underlying scale was estimated for overall IMI in the current study. Reported cow-level heritability for overall IMI on the underlying scale ranged from 0.045 to 0.21 in first and multiparous lactations (Weller et al., 1992; Schafberg et al., 2006). Differences in the reported heritability were attributed mainly due to differences in the reporting level (single vs multiple quarters).

On the observed scale at cow-level, Weller et al. (1992) reported a similar heritability of 0.024 to the present study (0.026; quarter-level) for overall IMI. However, those authors reported lower prevalence (22.8%) and used a linear model, which is frequency dependent. In contrast, a much higher cow-level heritability of 0.21 (± 0.11) was reported for overall IMI using linear sire model in a sample of 756 US Holstein cows (Wanner et al., 1998).

Comparing heritability estimates for pathogen-specific IMI groups with those from other studies is difficult, as grouping of pathogens differs among studies and most heritability estimates are reported for a single pathogen at the cow-level (Weller et al., 1992; Detilleux et al., 1995; Wanner et al., 1998; Schafberg et al., 2006). Regardless, heritability estimates for pathogen-specific IMI groups in the current study were in the low range compared to estimates in other studies (Weller et al., 1992; Detilleux et al., 1995; Wanner et al., 1998; Schafberg et al., 2006). Estimated heritabilities were likely lower because they were for single quarters, whereas in other studies, estimations were done for all 4 quarters. Heritability estimates for SCS of 0.047 per quarter (and a composite estimate of 0.13 (4 quarters) - data not shown) were in line with other studies that reported a heritability of SCS ranging from 0.08 to 0.17 (Koeck et al., 2010; Haugaard et al., 2012; Koeck et al., 2012).

Culture of serial milk samples is necessary to accurately determine IMI status of an udder quarter (Dohoo et al., 2011a). Additionally, composite sampling provides less accurate information about IMI status compared to sampling all 4 quarters, due to dilution effects (Dohoo et al., 2011a; Reyher and Dohoo, 2011; Souza et al., 2016). That Rep_w and Rep_a were both < 1 confirmed the importance of using repeated sampling and collecting samples from each quarter for genetic evaluation of pathogens-specific IMI. Moreover, due to large contributions of permanent environmental effects, repeatability estimations were higher than the heritability in the present study and may be regarded as upper limit of heritability (Young et al., 1960; Detilleux et al., 1995). Additionally, high repeatability of the contagious and major pathogen IMI groups was due to overlapping of pathogens. Repeatability estimated in present study was in line with repeatability estimated for overall IMI (0.24 ± 0.06) by Young et al. (1960). Detilleux et al. (1995) and Luttinen and Juga (1997) reported a high Rep_a of 0.31 and 0.43 for SCS, higher than in our study (0.20), attributed to a large contribution of permanent environmental effects across parities.

To our knowledge, no estimated genetic correlations among IMI caused by different pathogen groups have been reported. However, Sørensen et al. (2009a) and Haugaard et al. (2012) have reported a positive genetic correlation between pathogens-specific CM traits. In the present study, genetic correlation < 1 between IMI and pathogen-specific IMI traits indicated that these traits were genetically not exactly the same. High genetic correlations between IMI caused by NAS and minor pathogens, and contagious and major pathogens, were due to overlap in grouping of pathogens. A moderate genetic correlation (0.47±0.29) was estimated between environmental and contagious pathogen IMI in the present study. Similarly, Sørensen et al. (2009a) estimated genetic correlation of 0.45 (±0.17) between CM caused by *E. coli* and *S. aureus* based on CM and SCM data. However, Haugaard et al. (2012) estimated a high genetic correlation of 0.80 (±0.10) and 0.75 (±0.14) between CM caused by *E. coli*, a major environmental udder pathogen, and the contagious major pathogens *S. aureus* and *Strep. dysgalactiae*, respectively. In the present study, a low genetic correlation (0.28 ±0.26) was estimated between NAS and contagious pathogen IMI. Genetic correlation of NAS with environmental pathogen IMI was inestimable in the current study. However, Sørensen et al. (2009a) reported a moderate genetic correlation of 0.61 (±0.14), 0.60 (±0.14) and 0.75 (±0.09) between NAS IMI and IMI with contagious pathogen *S. aureus* and the environmental pathogens *E. coli* and *Strep. uberis.*

All analyzed traits, except NAS IMI, had moderate to high positive genetic correlations (0.39 to 0.81) with SCS. Contagious pathogen IMI had the highest genetic correlation of 0.81 (±0.11) with SCS; this was attributed to NAS only moderately increasing SCC, whereas contagious major pathogens such as *S. aureus* and *Strep. dysgalactiae* are associated with elevated SCC, and often for a long duration (De Haas et al., 2002b; De Haas et al., 2004; Piepers et al., 2007). Therefore, a strong genetic correlation between contagious pathogen IMI and SCS was expected. In contrast to our study, Schafberg et al. (2006) reported a higher genetic correlation of 0.85 between NAS IMI and SCS compared to genetic correlation between *S. aureus* IMI and SCS (0.66); however, they only used information from 1 test day.

Using CM data, De Haas et al. (2003) also estimated a non-significant genetic correlation (0.26 ±0.25) between NAS CM and SCS. The same authors estimated a moderate genetic correlation with CM caused by contagious major pathogens *S. aureus, Strep. dysgalactiae* with SCS traits. Conversely, Haugaard et al. (2013) reported 0.76 (±0.06), 0.73 (±0.08) and 0.78 (±0.06) genetic correlation of SCM caused by *S. aureus, Strep. dysgalactiae* and NAS with lactation mean

SCS. However, these authors reported a wide 95% highest probability density (HPD) interval, indicating the uncertainty of those genetic correlation estimations. In contrast with the findings of other studies and the present study, Sørensen et al. (2009b) reported a low genetic correlation of 0.44 and 0.47 for two SCS traits; lactation average SCC ranged from 5 to 170 (LASCC_170) and 5 to 300 DIM (LASCC_300), respectively, with *S. aureus* IMI compared to other pathogens analyzed. Based on positive moderate to high genetic correlations between analyzed traits and SCS, we inferred that selection for low SCS will favor IMI resistance, regardless of causative pathogen group. However, effects will be higher on contagious pathogen IMI compared to other pathogen groups.

Estimated within lactation phenotypic correlation between NAS and minor pathogens IMI was high (0.92) because of the large overlap in grouping of pathogens. Moreover, contagious pathogen IMI had the highest phenotypic correlation with SCS (0.24) compared to NAS and environmental IMI pathogen groups. This was attributed to a strong association between IMI due to contagious pathogens, such as *S. aureus,* and increased SCC (De Haas et al., 2002b; De Haas et al., 2004). Similar to our study, De Haas et al. (2003) and Sørensen et al. (2009b) estimated high phenotypic correlation between IMI due to contagious pathogen *S. aureus* and 2 SCC traits (lactation average SCC up to 150 (SCS150) and 305 DIM (SCS305) in De Haas et al. (2003); lactation average SCC from 5 to 170 (LASCC_170) and 5 to 300 DIM (LASCC_300) in Sørensen et al., 2009b) compared to other pathogens, based on CM data. However, in their studies a lower phenotypic correlation for NAS CM with SCC traits was reported (0.09 and 0.08; 0.018 and 0.021, respectively) compared to the present study, potentially because of a lower incidence of NAS CM in their studies.

Spearman rank correlations between EBV for overall IMI and pathogen-specific IMI traits were weak to moderate. Therefore, we inferred that re-ranking of sires would likely occur, depending on the trait chosen (Table 3-4). Conversely, based on strong rank correlation between EBV for NAS and minor pathogen IMI, we inferred that sire rankings may not differ between these 2 groups of pathogens. Furthermore, none to weak rank correlation between overall IMI and pathogen-specific IMI traits with SCS implied that potential for substantial re-ranking of sires for IMI traits when selecting for lower SCC. Therefore, selection for lower SCC will not improve resistance to IMI for every pathogen-specific group in the same manner. Notably, the current study demonstrated that despite low heritability for overall IMI and IMI causing pathogen groups, there was substantial variation in daughters' IMI prevalence from sires with extreme EBV for IMI traits. Hence, there certainly is an opportunity to use breeding to improve genetic resistance to pathogen-specific IMI.

3.6. Implications

Sufficient knowledge about genetic variation, selection mechanisms and economic incentives are required for efficient genetic improvement (Gibson, 1991). The present study demonstrated that despite low heritability, there was exploitable genetic variation among sires in producing daughters with resistance to pathogen-specific IMI. Consequently, there is potential to breed cows more resistant to IMI caused by specific pathogens groups. Furthermore, the distribution of pathogens causing IMI differed among farms and regions. The approach, including genetic selection, therefore, needs to take this into consideration. Additionally, pathogen-specific IMI has a large economic impact on dairy industry (Sørensen et al., 2010; Heikkilä et al., 2018). Therefore, developing a pathogen group-specific udder health index may be advantageous.

However, routine milk sample collection and bacterial culture analysis will be a challenge, in terms of labor intensity and costs. Additionally, use of genomic information will enhance the accuracy of estimated breeding values for pathogen specific IMI traits and, perhaps, would allow for genetic evaluation with a lower number of phenotyped cows.

3.7. Conclusions

Prevalence of IMI was high at quarter-level, with a similar distributions of pathogen groups across the 6 Canadian provinces included in this study. Despite low heritability for pathogen-specific IMI traits, there was substantial exploitable genetic variation among daughters of sires, providing opportunity for improvement of pathogen-specific IMI resistance via long-term genetic selection. Weak to strong genetic correlations among pathogen-specific IMI traits and between overall IMI and pathogen-specific IMI traits indicated that not all these traits were genetically similar. Furthermore, weak to strong positive genetic correlation between pathogen-specific IMI traits and SCS confirmed that selection for lower SCS favored resistance to IMI. However, selection for lower SCS will not improve resistance to IMI from every pathogen group in the same manner, due to weak to moderate rank correlations between sire EBVs for pathogen-specific IMI trait and for SCS.

3.8. References

Aghamohammadi, M., D. Haine, D. F. Kelton, H. W. Barkema, H. Hogeveen, G. P. Keefe, and S. Dufour. 2018. Herd-level mastitis-associated costs on Canadian dairy farms. Front. Vet. Sci. 5:100.

Andersen, S., I. R. Dohoo, R. Olde Riekerink, and H. Stryhn., and Mastitis Research Workers' Conference. 2010. Diagnosing intramammary infections: Evaluating expert opinions on the definition of intramammary infection using conjoint analysis. J. Dairy Sci.93:2966–2975.

Berry, D. P. and W. J. Meaney. 2006. Interdependence and distribution of subclinical mastitis and intramammary infection among udder quarters in dairy cattle. Prev. Vet. Med. 75:81–91.

Cook, W. F., E. A. Fiez, and L. K. Fox. 1992. Mastitis in first lactation Southwest Idaho dairy cows. J. Dairy Sci. 75:158.

Condas, L. A., J. De Buck, D. B. Nobrega, D. A. Carson, J. P. Roy, G. P. Keefe, T. J. DeVries, J. R. Middleton, S. Dufour, and H. W. Barkema. 2017a. Distribution of non-*aureus* staphylococci species in udder quarters with low and high somatic cell count, and clinical mastitis. J. Dairy Sci. 100:5613–5627.

Condas, L. A. Z., J. De Buck, D. B. Nobrega, D. A. Carson, S. Naushad, S. De Vliegher, R. N. Zadoks, J. R. Middleton, S. Dufour, J. P. Kastelic, and H. W. Barkema. 2017b. Prevalence of non-*aureus* staphylococci species causing intramammary infections in Canadian dairy herds. J. Dairy Sci. 100:5592–5612.

Da Costa, E. O., P. A. Melville, A. R. Ribeiro, E. Watanabe, F. C. Viani, and C. R. White. 1996. Prevalence of intramammary infections in primigravid Brazilian dairy heifers. Prev. Vet. Med. 29:151-155.

De Haas, Y., H. W. Barkema, and R. F. Veerkamp. 2002a. Genetic parameters of pathogen-specific incidence of clinical mastitis in dairy cows. Anim. Sci. 74:233–242.

De Haas, Y., H. W. Barkema, and R. F. Veerkamp. 2002b. The effect of pathogen-specific clinical mastitis on the lactation curve for somatic cell count. J. Dairy Sci 85:1314–1323.

De Haas, Y., H. W. Barkema, Y. H. Schukken, and R. F. Veerkamp. 2003. Genetic associations for pathogen-specific clinical mastitis and patterns of peaks in somatic cell count. Anim. Sci. 77:187-195.

De Haas, Y., R. F. Veerkamp, H. W. Barkema, Y. T. Gröhn, and Y. H. Schukken. 2004. Associations between pathogen-specific cases of clinical mastitis and somatic cell count patterns. J. Dairy Sci 87:95–105.

Dempster, E. R. and I. M. Lerner. 1950. Heritability of threshold characters. Genetics 35:212–236.

Dempster, J. C., M. E. Kehrli, A. E. Freeman, L. K. Fox, and D. H. Kellev. 1995. Mastitis of periparturient Holstein cattle: a phenotypic and genetic study. J. Dairy Sci. 78:2285–2293.

De Visscher, A., S. Piepers, F. Haesebrouck, and S. De Vliegher. 2016. Intramammary infection with coagulase-negative staphylococci at parturition: Species-specific prevalence, risk factors, and effect on udder health. J. Dairy Sci. 99:6457–6469.

Dohoo, I., S. Andersen, R. Dingwell, K. Hand, D. Kelton, K. Leslie, Y. Schukken, and S. Godden. 2011a. Diagnosing intramammary infections: comparison of multiple versus single quarter milk samples for the identification of intramammary infections in lactating dairy cows. J. Dairy Sci. 94:5515–5522.

Dohoo, I. R., J. Smith, S. Andersen, D. F. Kelton, S. Godden, and Mastitis Research Workers' Conference. 2011b. Diagnosing intramammary infections: Evaluation of definitions based on a single milk sample. J. Dairy Sci. 94:250–261.

Dufour, S., I. R. Dohoo, H. W. Barkema, L. DesCôteaux, T. J. DeVries, K. K. Reyher, J. P. Roy, and D. T. Scholl. 2012. Epidemiology of coagulase-negative staphylococci intramammary infection in dairy cattle and the effect of bacteriological culture misclassification. J. Dairy Sci. 95: 3110-3124.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics, Ed 4. Longmans

Green, Harlow, Essex, UK.

Fox, L. K. 2009. Prevalence, incidence and risk factors of heifer mastitis. Vet. Microbiol. 134: 82-88.

Gianola, D. 1982. Theory and analysis of threshold characters. J. Anim. Sci. 54:1079–1096.

Gianola, D. and J. L. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201.

Gibson, J. P. 1991. The potential for genetic change in milk fat com- position. J. Dairy Sci. 74:3258–3266.

Haugaard, K., B. Heringstad, and A. Whist. 2012. Genetic analysis of pathogen-specific clinical mastitis in Norwegian Red cows. J. Dairy Sci. 95:1545–1551.

Haugaard, K., B. Heringstad, and A.C. Whist. 2013. Genetic associations between somatic cell score and pathogen-specific subclinical mastitis in Norwegian Red cows. J. Anim. Breed. Genet. 130:98–105.

Heikkilä, A.-M., E. Liski, S. Pyörälä, and S. Taponen. 2018. Pathogen-specific production losses in bovine mastitis. J. Dairy Sci. 101: 9493-9504.

Holmberg, M., W. F. Fikse, L. Andersson-Eklund, K. Artursson, and A. Lundén. 2012. Genetic analyses of pathogen-specific mastitis. J. Anim. Breed. Genet. 129:129–137.

Koeck, A., B. Heringstad, C. Egger-Danner, C. Fuerst, P. Winter, and B. Fuerst-Waltl. 2010. Genetic analysis of clinical mastitis and somatic cell count traits in Austrian Fleckvieh cows. J. Dairy Sci. 93:5987–5995.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012. Alternative somatic cell count traits to improve mastitis resistance in Canadian Holsteins. J. Dairy Sci. 95:432–439.

Luttinen, A. and J. Juga. 1997. Genetic Relationships between milk yield, somatic cell count, mastitis, milkability and leakage in Finnish dairy cattle population. Interbull Bull. 15:78–83.

Middleton, J. R., L. L. Timms, G. R. Bader, J. Lakritz, C. D. Luby, and B. J. Steevens. 2005. Effect of prepartum intramammary treatment with pirlimycin hydrochloride on prevalence of early first-lactation mastitis in dairy heifers. J. Am. Vet. Med. Assoc. 227:1969-1974.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). Proc. 7th World Congr. Genet. Appl. Livest. Prod. 33:743-744.

Oliver, S. P., M. J. Lewis, B. E. Gillespie, H. H. Dowlen, E. C. Jaenicke, and R. K. Roberts. 2003. Prepartum antibiotic treatment of heifers: milk production, milk quality and economic benefit. J. Dairy Sci. 86:1187–1193.

Parker, K. I., C. Compton, F. M. Anniss, A. Weir, C. Heuer, and S. McDougall. 2007. Subclinical and clinical mastitis in heifers following the use of a teat sealant precalving. J. Dairy Sci. 90:207–218.

Piepers, S., L. De Meulemeester, A. De Kruif, G. Opsomer, H. W. Barkema, and S. De Vliegher. 2007. Prevalence and distribution of mastitis pathogens in subclinically infected dairy cows in Flanders, Belgium. J. Dairy Res. 74: 478-483.

Piepers, S., K. Peeters, G. Opsomer, H. W. Barkema, K. Frankena, and S. De Vliegher. 2011. Pathogen group specific risk factors at herd, heifer and quarter levels for intramammary infections in early lactating dairy heifers. Prev. Vet. Med. 99:91–101.

Piepers, S., S. De Vliegher, A. De Kruif, G. Opsomer, and H. W. Barkema. 2009. Impact of intramammary infections in dairy heifers on future udder health, milk production, and culling. Vet. Microbiol. 134:113-120.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reyher, K. K. and I. R. Dohoo. 2011. Diagnosing intramammary in- fections: Evaluation of composite milk samples to detect intrama- mmary infections. J. Dairy Sci. 94:3387–3396.

Reyher, K. K., S. Dufour, H. W. Barkema, L. Des Côteaux, T. J. DeVries, I. R. Dohoo, G. P. Keefe, J. P. Roy, and D. T. Scholl. 2011. The National Cohort of Dairy Farms—A data collection platform for mastitis research in Canada. J. Dairy Sci. 94:1616–1626.

Sampimon, O. C., H. W. Barkema, I. Berends, J. Sol, and T.J.G.M. Lam. 2009a. Prevalence of intramammary infection in Dutch dairy herds. J. Dairy Res. 76:129-136.

Sampimon, O. C., H. W. Barkema, I.M.G.A. Berends, J. Sol, and T.J.G.M. Lam. 2009b. Prevalence and herd-level risk factors for intramammary infection with coagulase-negative staphylococci in Dutch dairy herds. Vet. Microbiol. 134:37-44

SAS Institute. 2013. SAS 9.4 Guide to software updates. SAS Institute Inc., Cary, NC.

Schafberg, R., F. Rosner, and H. H. Swalve. 2006. Examinations on intramammary infections in dairy cows based on pathogen-specific data. Proc. 8th World Congr. Genet. Appl. Livest. Prod. 15-13.

Shook, G. E. 1993. Genetic improvement of mastitiss with through selection on somatic cell count. Vet. Clin. North Am. Food Anim. Pract. 9:563–577.

Sørensen, L. P., P. Madsen, T. Mark, and M. S. Lund. 2009a. Genetic parameters for pathogen-specific mastitis resistance in Danish Holstein Cattle. Animal 3:647–656.

Sørensen, L. P., T. Mark, P. Madsen, and M. S. Lund. 2009b. Genetic correlations between pathogen-specific mastitis and somatic cell count in Danish Holsteins. J. Dairy Sci. 92:3457–3471.

Sørensen, L. P., T. Mark, M. K. Sørensen, and S. Østergaard. 2010. Economic values and expected effect of selection index for pathogen-specific mastitis under Danish conditions. J. Dairy Sci.

93:358–369.

Souza, F. N., A. F. Cunha, D. L. Rosa, M. A. V. Brito, A. S. Guimarães, L. C. Mendonça, G. N. Souza, A. P. Lage, M. G. Blagitz, A. M. Della Libera, and M. B. Heinemann. 2016. Somatic cell count and mastitis pathogen detection in composite and single or duplicate quarter milk samples. Pesq. Vet. Bras. 36:811-818.

Tsuruta, S. and I. Misztal. 2006. THRGIBBS1F90 for estimation of variance components with threshold-linear models. Threshold 3:4.

Wanner, J. M., G. W. Rogers, M. E. Kehrli, and J. B. Cooper. 1998. Genetics and breeding: intramammary infections in primiparous Holsteins: heritabilities and comparisons of bovine leukocyte adhesion deficiency carriers and noncarriers. J. Dairy Sci. 81:3293–3299.

Weller, J. I., A. Saran, and Y. Zeliger. 1992. Genetic and environmental relationships among somatic cell count, bacterial infection, and clinical mastitis. J. Dairy Sci. 75:2532-2540.

Young, C. W., J. E. Legates, and J. G. Lecce. 1960. Genetic and phenotypic relationships between clinical mastitis, laboratory criteria, and udder height. J. Dairy Sci. 43:54–62.

Table 3-1. Quarter-level prevalence (%) of intramammary infection-related traits[1] in multiparous cows from 0 to 400 DIM across Canadian provinces.

Trait[1]	Province[2]						Total
	AB	NB	NS	ON	PE	QC	
IMI	49.2	31.9	32.8	53.2	29.9	45.3	45.5
NAS	30.7	18.3	19.3	38.5	20.5	30.9	30.8
CON	9.5	5.1	4.6	6.4	2.1	9.9	7.5
ENV	18.3	11.6	11.1	16.1	9.8	11.4	13.9
Major	5.9	8.9	5.9	7.3	4.5	7.5	6.8
Minor	35.9	19.5	21.6	41.0	21.5	35.3	34.1
CN	50.8	68.1	67.2	46.8	70.1	54.7	54.5

[1]IMI = overall intramammary infection, NAS = non-*aureus staphylococci*, CON = contagious pathogens, ENV = environmental pathogens, Major = major pathogens, Minor = minor pathogens, CN = culture-negative.

[2]AB = Alberta, NB = New Brunswick, NS = Nova Scotia, ON = Ontario, PE = Prince Edward Island, QC= Québec.

Table 3-2. Posterior mean of heritability (per quarter) of intramammary infection-related traits on the underlying scale, heritability (per quarter) of somatic cell score on the observed scale and repeatability within and across lactation (posterior SD in parentheses).

Trait[1]	Heritability (Under)	Heritability (Obs)	Rep_w[2]	Rep_a[3]
IMI	0.041 (0.013)	0.026	0.175 (0.007)	0.125 (0.019)
NAS	0.035 (0.012)	0.020	0.166 (0.007)	0.108 (0.018)
CON	0.073 (0.030)	0.021	0.399 (0.014)	0.297 (0.044)
ENV	0.017 (0.009)	0.007	0.089 (0.007)	0.048 (0.016)
Major	0.043 (0.018)	0.012	0.337 (0.012)	0.239 (0.036)
Minor	0.041 (0.013)	0.025	0.177 (0.008)	0.112 (0.019)
SCS	--	0.047 (0.014)	0.290 (0.006)	0.200 (0.021)

[1]IMI = overall intramammary infection, NAS = non-*aureus* staphylococci, CON = contagious pathogens, ENV = environmental pathogens, Major = major pathogens, Minor = minor pathogens, SCS = somatic cell score.

[2]Repeatability within lactation.

[3]Repeatability across lactations.

Table 3-3. Posterior mean of genetic correlations (above the diagonal; posterior SD in parentheses) and phenotypic correlations within lactation (below the diagonal) among intramammary infection-related traits and somatic cell score estimated from bivariate analysis.

Trait[1]	IMI	NAS	CON	ENV	Major	Minor	SCS
IMI		0.79	0.78	0.18	0.67	0.90	0.67
		(0.08)	(0.11)	(0.25)	(0.14)	(0.05)	(0.12)
NAS	0.90		0.28	-	0.19	0.97	0.11
			(0.26)		(0.29)	(0.01)	(0.19)
CON	0.75	-0.06		0.47	0.91	0.56	0.81
				(0.29)	(0.09)	(0.18)	(0.11)
ENV	0.86	0.02	-0.07		0.28	-	0.59
					(0.29)		(0.21)
Major	0.79	-0.02	0.51	0.71		0.30	0.48
						(0.25)	(0.22)
Minor	0.91	0.92	0.41	0.08	-0.03		0.39
							(0.16)
SCS	0.28	0.19	0.24	0.08	0.28	0.16	

[1]IMI = overall intramammary infection, NAS = non-*aureus* staphylococci, CON = contagious pathogens, ENV = environmental pathogens, Major = major pathogens, Minor = minor pathogens, SCS = somatic cell score.

Table 3-4. Rank correlations between EBV for alternative intramammary infection (IMI) traits and somatic cell score for 55 sires with at least 10 daughters in 5 herds.

Trait[1]	NAS	CON	ENV	Major	Minor	SCS
IMI	0.83	0.42	0.31	0.42	0.87	0.43
NAS		0.20	0.06	0.08	0.96	0.32
CON			0.01	0.53	0.33	0.47
ENV				0.51	0.03	0.04
Major					0.11	0.41
Minor						0.40

[1]IMI = overall intramammary infection, NAS = non-*aureus* staphylococci, CON = contagious pathogens, ENV = environmental pathogens, Major = major pathogens, Minor = minor pathogens, SCS = somatic cell score.

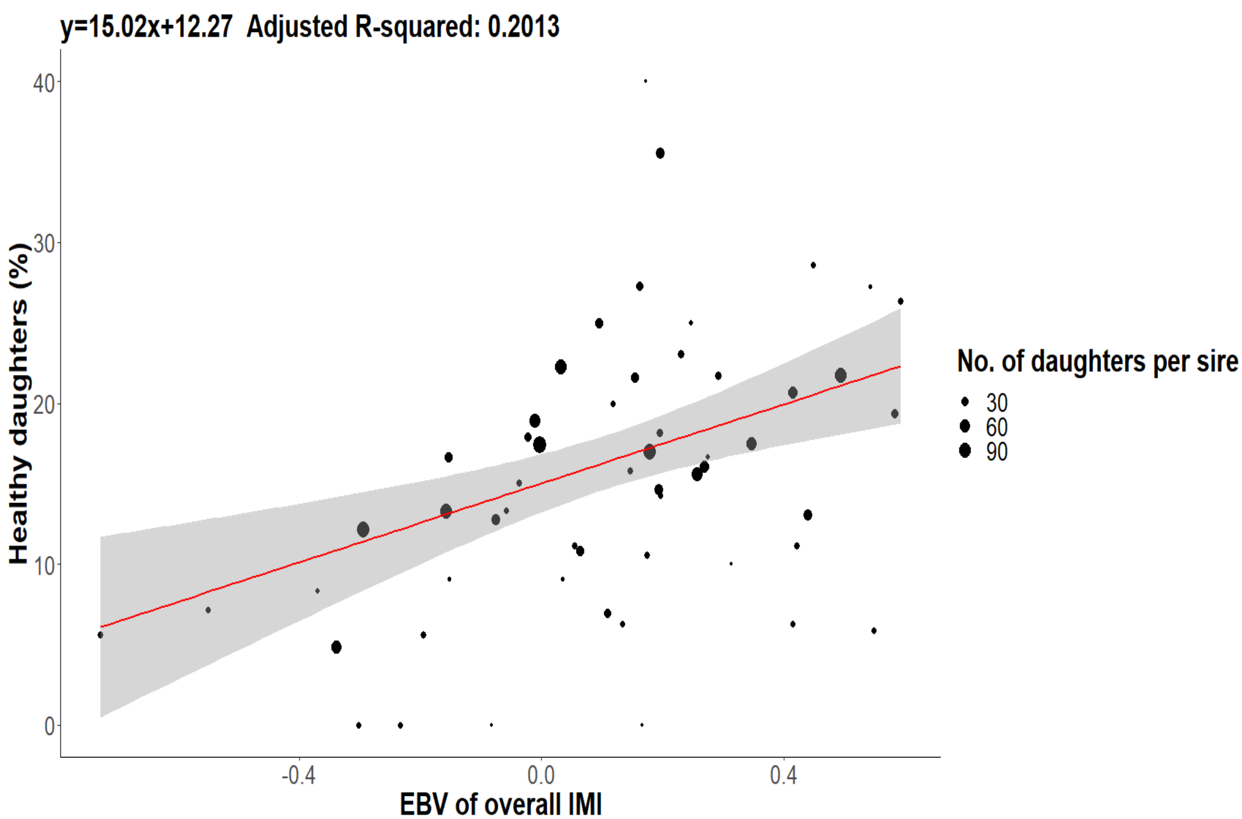
Table 3-5. Percentage of daughters with intramammary infection (IMI) from all (average), the best (10% decile), and the worst (90% decile) sires (with at least 10 daughters in > 5 herds (n = 55)) according to their estimated breeding value for resistance to IMI predicted with a univariate probit threshold model for IMI traits.

Trait[1]	Average	10% decile	90% decile
IMI	85	80	94
NAS	75	63	82
CON	24	13	38
ENV	39	28	50
Major	24	13	40
Minor	74	65	86

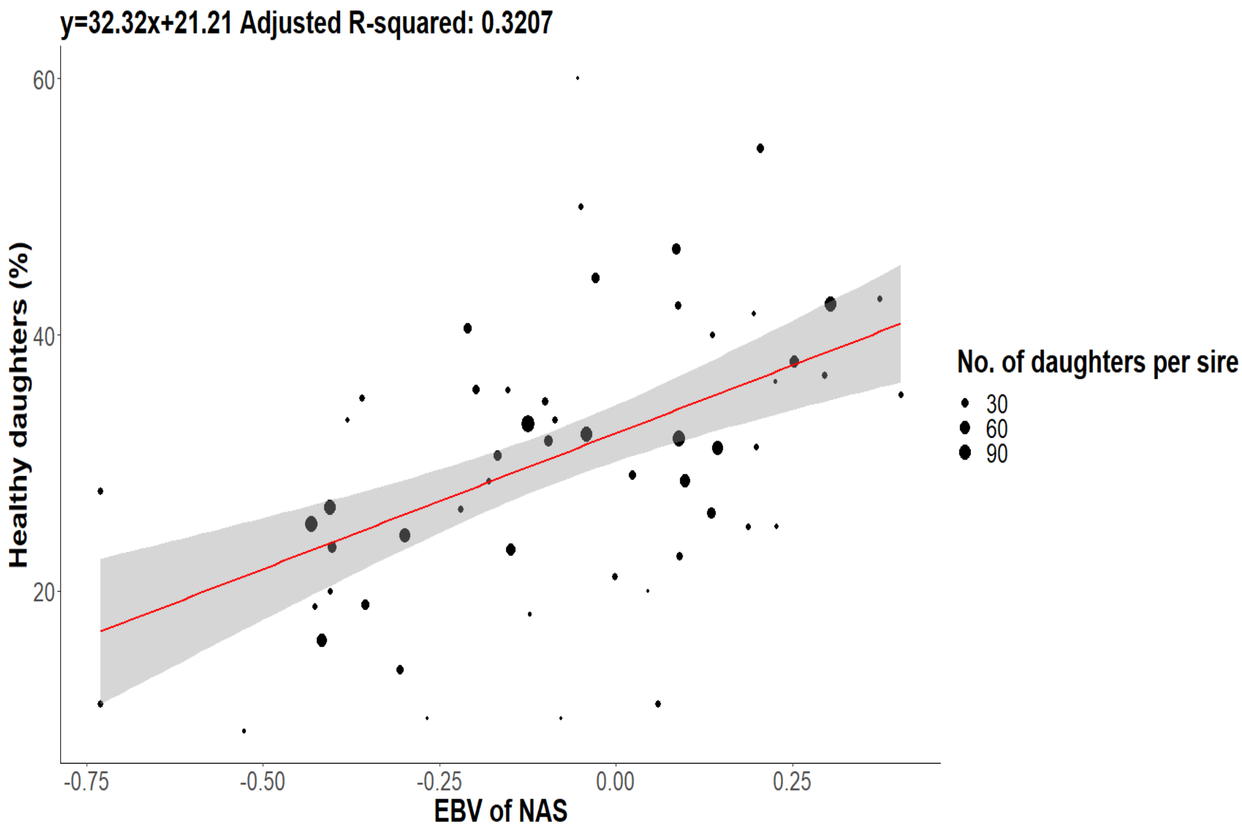
[1]IMI = overall intramammary infection, NAS = non-*aureus* staphylococci, CON = contagious pathogens, ENV = environmental pathogens, Major = major pathogens, Minor = minor pathogens.

Figure 3-1. Percentage of healthy daughters according to the estimated breeding value (EBV) of pathogen-specific IMI resistance of sires with at least 10 daughters in > 5 herds (n = 55) for a) IMI = overall intramammary infection; b) NAS = non-*aureus* staphylococci; c) CON = contagious pathogens; d) ENV = environmental pathogens; e) Major = major pathogens; and f) Minor = minor pathogens.

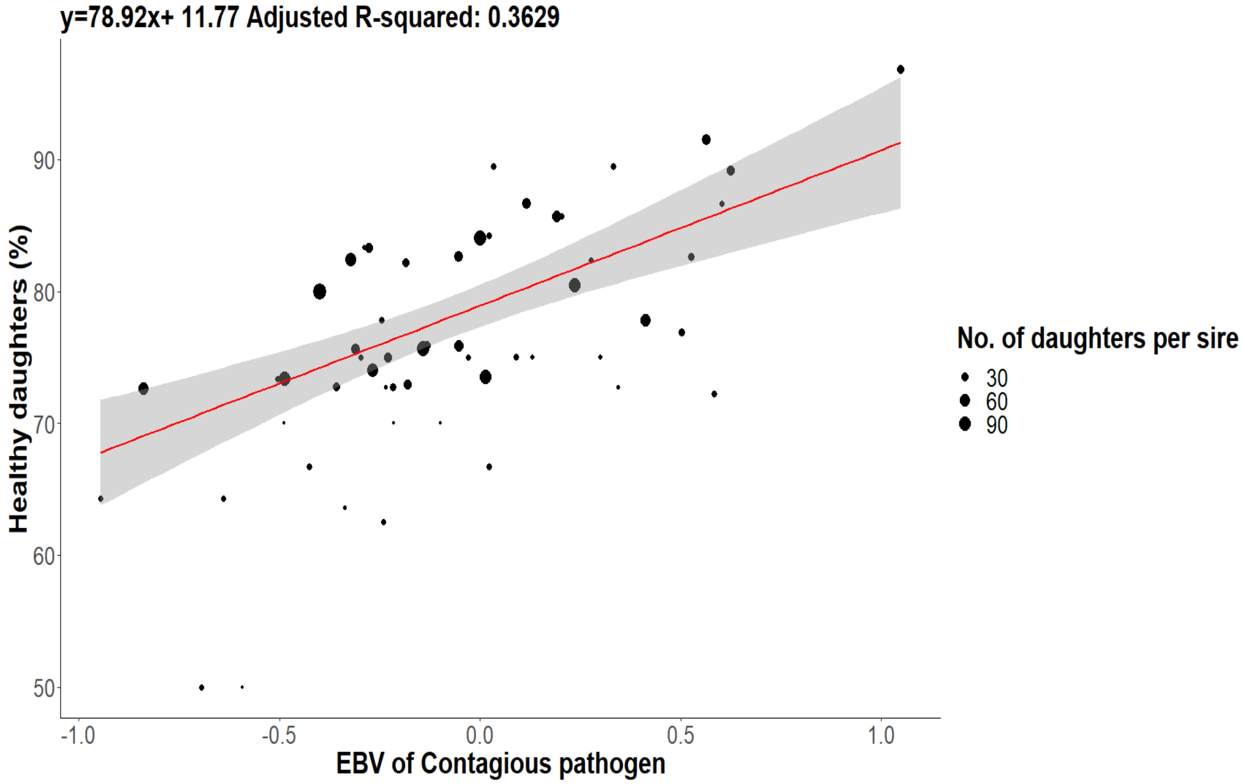
(a)



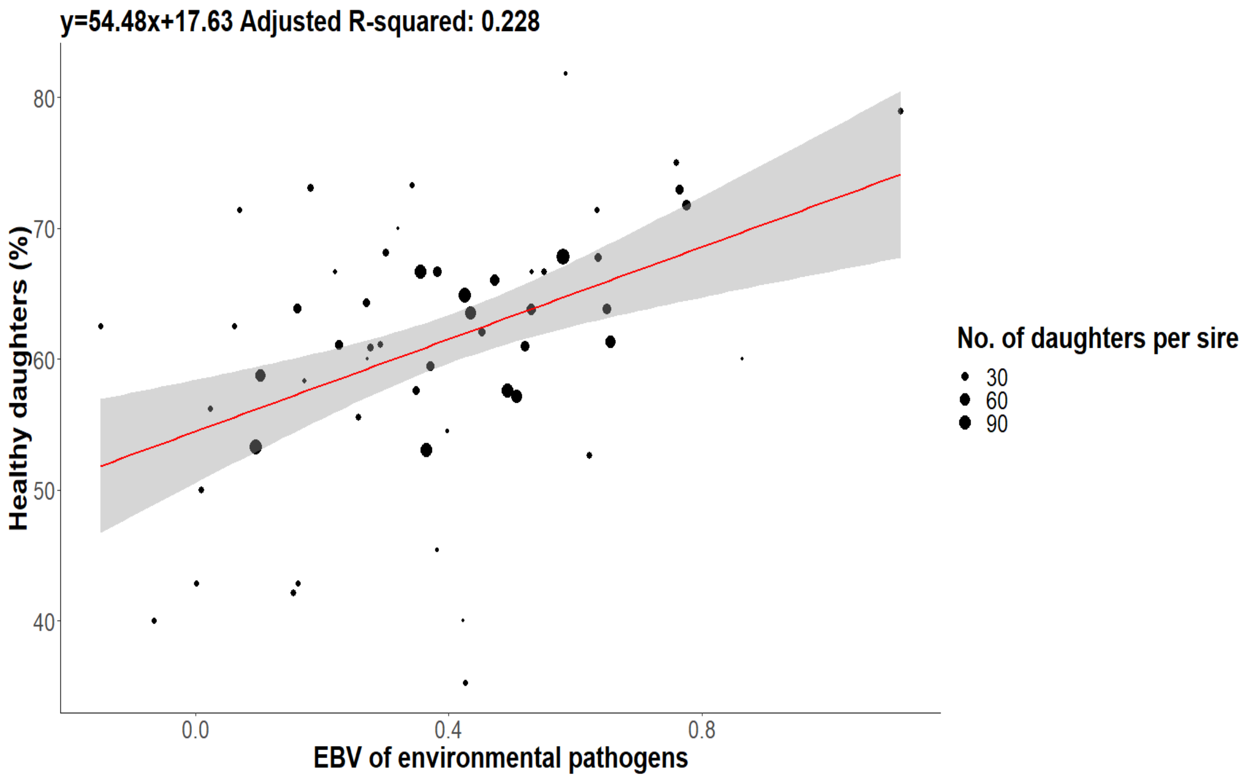
(b)


y=32.32x+21.21 Adjusted R-squared: 0.3207
Healthy daughters (%)
60
40
20
-0.75
-0.50
-0.25
0.00
0.25
EBV of NAS
No. of daughters per sire
30
60
90

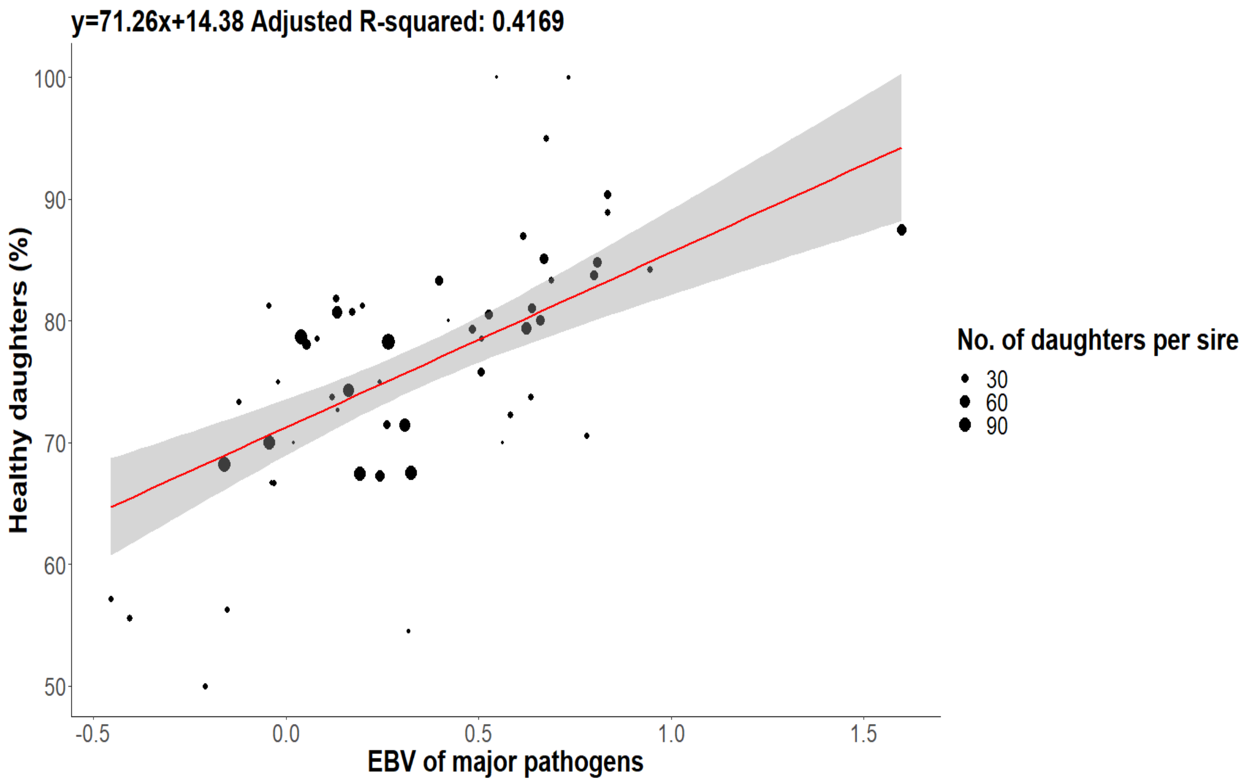
(c)


y=78.92x+ 11.77 Adjusted R-squared: 0.3629
Healthy daughters (%)
90
80
70
60
50
-1.0
-0.5
0.0
0.5
1.0
EBV of Contagious pathogen
No. of daughters per sire
30
60
90

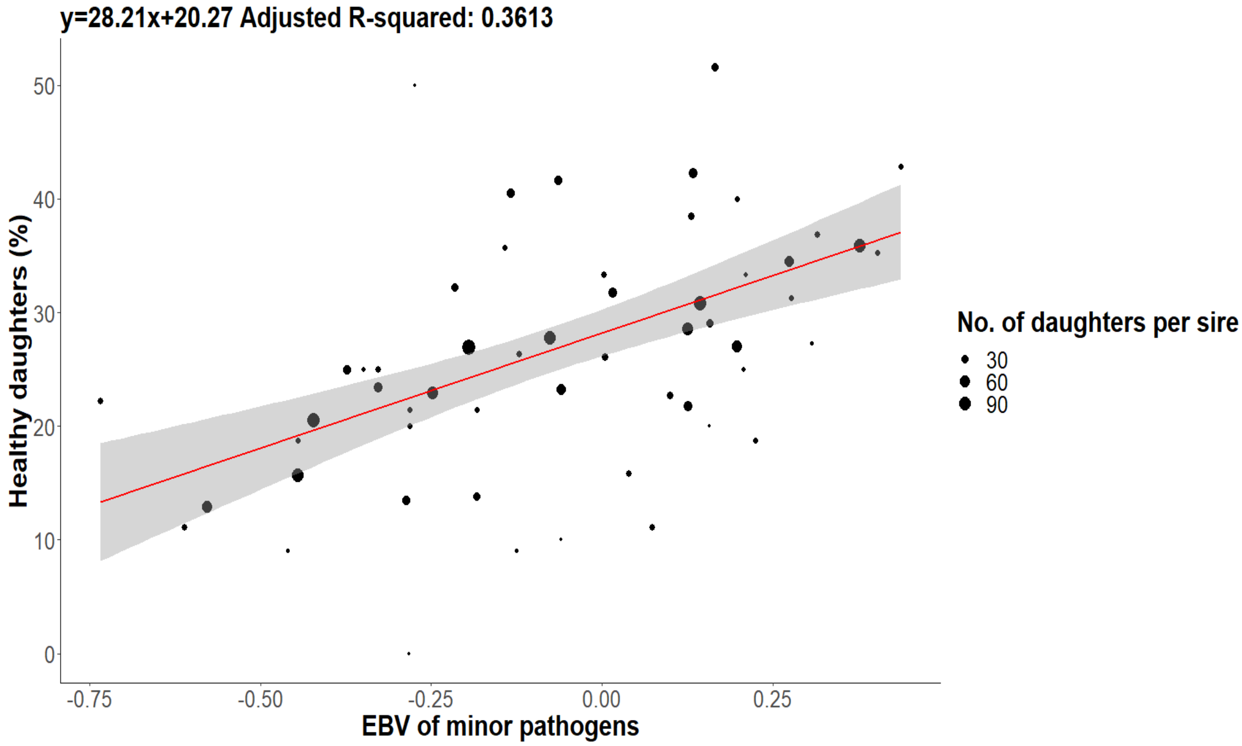
(d)


y=54.48x+17.63 Adjusted R-squared: 0.228
Healthy daughters (%)
80
70
60
50
40
0.0
0.4
0.8
EBV of environmental pathogens
No. of daughters per sire
30
60
90

(e)


y=71.26x+14.38 Adjusted R-squared: 0.4169
Healthy daughters (%)
100
90
80
70
60
50
-0.5
0.0
0.5
1.0
1.5
EBV of major pathogens
No. of daughters per sire
30
60
90

(f)


y=28.21x+20.27 Adjusted R-squared: 0.3613
Healthy daughters (%)
50
40
30
20
10
0
-0.75
-0.50
-0.25
0.00
0.25
EBV of minor pathogens
No. of daughters per sire
30
60
90

CHAPTER 4: Underlying genetic architecture of resistance to mastitis in dairy cattle: A systematic review and gene prioritization analysis of genome-wide association studies

4.1. Abstract

Mastitis is the most frequent disease in dairy cattle. Resistance to mastitis is a complex, polygenic trait controlled by many genes, each with small effects. Genome-wide association studies (GWAS) have been widely used to identify genomic variants associated with complex traits, including resistance to mastitis, to elucidate the underlying genetic architecture of the trait. However, no systematic review and gene prioritization analysis have been conducted to date on GWAS results for resistance to mastitis in dairy cattle. Hence, the objective was to perform a systematic review and gene prioritization analysis of GWAS studies to identify potential functional candidate genes associated with resistance to mastitis-related traits in dairy cattle. Four electronic databases were searched from inception to December 2020, supplemented with multiple sources of grey literature, to identify eligible articles. Annotation for genes and quantitative trait loci (QTLs), and QTL enrichment analysis were conducted using GALLO. Gene prioritization analysis was performed by a guilty-by-association approach using GUILDify and ToppGene. From 52 articles included within this systematic review, 30 articles were used for further functional analyses. Gene and QTL annotation resulted in 9,125 and 43,646 unique genes and QTLs, respectively, from 39 studies. In general, overlapping of genes across studies was very low (mean ± SD = 0.02% ± 0.07%). Most annotated genes were associated with somatic cell count-related traits and the Holstein breed. Within all annotated genes, 74 genes were shared among Holstein, Jersey and Ayrshire breeds. Nearly 7.5% of annotated QTL were related to QTL class “health.” Within the health QTL class, 2.6 and 2.2% of QTLs were

associated with clinical mastitis and somatic cell count-related traits. Enrichment analysis of QTLs demonstrated that many enriched QTLs were associated with somatic cell score located in *Bos taurus* autosomes 5, 6, 16, and 20. The prioritization analysis resulted in 427 significant genes after multiple test correction (false discovery rate of 5%) from 26 studies. Most prioritized genes were located in *Bos taurus* autosomes 19 and 7, and most top-ranked genes were from the cytokine superfamily (e.g., chemokines, interleukins, transforming growth factors, and tumor necrosis factor genes). Although most prioritized genes (397) were associated with somatic cell count-related traits, only 54 genes were associated with clinical mastitis-related traits. Twenty-four genes (*ABCC9, ACHE, ADCYAP1, ARC, BCL2L1, CDKN1A, EPO, GABBR2, GDNF, GNRHR, IKBKE, JAG1, KCNJ8, KCNQ1, LIFR, MC3R, MYOZ3, NFKB1, OSMR, PPP3CA, PRLR, SHARPIN, SLC1A3* and *TNFRSF25*) were reported for both somatic cell count and clinical mastitis-related traits. Prioritized genes were mainly associated with immune response, regulation of secretion, locomotion, cell proliferation, and development. In conclusion, this study provided a fine mapping of previously identified genomic regions associated with resistance to mastitis and identified key functional candidate genes for resistance to mastitis, which can be used to develop enhanced genomic strategies to combat mastitis by increasing mastitis resistance through genetic selection.

4.2. Introduction

Mastitis is the most common and expensive infectious disease in dairy cattle (Halasa et al., 2007). It is caused by IMI with pathogenic bacteria in the udder which, if not cleared, can manifest as subclinical mastitis (**SCM**) or clinical mastitis (**CM**). The prevalence of IMI ranges from 17 to 76% depending on lactation stage and parity (Oliver et al., 2003; Piepers et al., 2007; Narayana et al., 2021), whereas the prevalence of SCM in first-parity Canadian Holstein heifers in early lactation ranges from 13 to 24% (Narayana et al., 2018). Incidence rate of CM ranges from 23.7 cases per 100 cows per year in Canada (Levison et al., 2016), to 32.5 cases per 100 cows per year in the Netherlands (Santman-Berends et al., 2015) and 39.6 cases per 100 cows per year on large dairy farms in China (Gao et al., 2017). Mastitis can impact animal welfare and farm profitability through reduced milk yield, milk quality and treatment costs (Halasa et al., 2007). Costs associated with mastitis have been estimated at $66,178 CAD/100 cows per year for a typical Canadian dairy farm (Aghamohammadi et al., 2018).

Effectiveness of traditional methods to control mastitis, e.g., antimicrobial therapy and disinfectants, is relatively low (Pyörälä et al., 2002). Therefore, selective genetic breeding for enhanced mastitis resistance is a potential strategy to control mastitis, together with better management practices. Genetic improvement is cumulative and results in permanent and cost-efficient changes. However, the success of genetic selection of a trait depends on its heritability and additive genetic variation. Since the heritability of resistance to mastitis is low (≤0.05) (Koeck et al., 2012; Narayana et al., 2018; 2021), genomics provides exceptional opportunities to increase the frequency of favourable alleles in the population, compared to conventional pedigree-based breeding. This was demonstrated by an increase in genetic gain of 0.46 (standard units) when comparing the past 5 years to the 5 years before introduction of genomic selection for mastitis resistance (CDN, 2019). In addition, identifying genomic variants associated with mastitis resistance aids in understanding its genetic architecture (Tiezzi et al., 2015).

Genome-wide association studies are advanced techniques commonly used to detect significant genetic markers (SNPs, haplotypes) associated with complex traits, resulting in identification of QTLs and candidate genes. Resistance to mastitis is a complex and polygenic trait controlled by several genes, each with small effects (Rupp and Boichard, 2003; Tiezzi et al., 2015).

Therefore, numerous studies have performed GWAS of CM (Tiezzi et al., 2015; Welderufael et al., 2018; Kruz et al., 2019) and SCC, or its log transformation SCS, an indicator trait of SCM (Durán Aguilar et al., 2016; Oliveira et al., 2019) to identify genetic variants associated with resistance to mastitis and to understand its architecture. Several different as well as a few similar genetic variants associated with CM and SCC-related traits have been reported. However, these genetic variants depended on the phenotype, population, breed, genotyping and GWAS approach used for the analysis. This has resulted in a long list of positional candidate genes associated with resistance to mastitis, although only a few genes have been validated (Leyva-Baca et al., 2007; Chen et al., 2015). In the absence of appropriate prioritization analysis, a long list of candidate genes causes pitfalls in confirming potential functional candidate genes. A review of reported SNPs associated with SCC, using 7 published GWAS studies (Chen et al., 2015), did not use a systematic approach to select relevant literature and there were no formal comparisons between reported genes. More recently, several other studies have conducted GWAS and reported novel and similar genomic variants associated with resistance to mastitis. Therefore, there is a need for a systematic review of GWAS on resistance to mastitis and prioritization analyses to pinpoint potential functional candidate genes to improve mastitis resistance through genetic selection.

To our knowledge, there are no formal systematic reviews nor gene prioritization analyses on this topic. Given this knowledge gap, our objectives were to: 1) conduct a systematic review of GWAS on resistance to mastitis; and 2) perform gene prioritization analysis of GWAS results of resistance to mastitis to identify potential functional candidate genes.

4.3. Materials and Methods

4.3.1. Systematic Review Search strategy

This systematic review followed a pre-specified study protocol and the standards of preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement (Page et al., 2021). A combination of online databases and conference proceedings were searched to identify relevant literature. Online databases searched included CAB Abstracts (EBSCO), MEDLINE (OVID), Web of Science, and BIOSIS (Web of Science) from the database inception to December, 2020, with no language restrictions. Search terms consisted of both keywords and database-specific subject headings for the elements of the PPO framework: Population - dairy cattle; Prognostic tool - GWAS; and Outcome - mastitis/SCC-related traits.

Identifiers and synonyms for each framework element were combined using the Boolean operator "OR." Thereafter, the elements of the framework were combined using the Boolean operator "AND." The specific search terms used for CAB Abstracts, MEDLINE, Web of Science and BIOSIS are displayed in Supplementary Material 1 (https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022). Twenty-seven relevant articles identified using Google Scholar were used as a test set to check the recall of the search strategy. In addition, hand searches were conducted (from 2005 onwards) of the following conference proceedings: Plant and Animal Genome Conference (PAG); World Congress on Genetics Applied to Livestock Production (WGALP); and International Conference of Quantitative Genetics (ICQG). Searches were also conducted (from 2005 onwards) for proceedings of the following meetings: American Dairy Science Association (ADSA); National Mastitis Council (NMC); International Society for Animal Genetics (ISAG); and European Federation of Animal Science (EAAP). All identified articles were imported into Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia) and duplicates removed. Each abstract was independently screened by Saranya Narayana (**SN**) and Ellen de Jong (**EJ**). Abstracts were selected based on the following criteria: 1) original work (review articles, editorials, and commentaries were excluded); 2) population of interest: dairy cattle; and 3) mentioning of variants such as SNPs, sequences, copy number variation (**CNV**) and haplotypes used to conduct GWAS. During full-text review, articles were excluded if: 4) a GWAS technique was not used for the analysis; or 5) phenotypes defined did not include resistance to mastitis or SCC-related traits. Any disagreement between the 2 reviewers (SN and EJ) regarding inclusion of an abstract or full-text was resolved by consensus after discussion. The included articles were further screened for relevant references and a citation check was performed. If the full-text article was missing, or there was uncertainty regarding duplicates (results presented in both a conference and published work), the lead author was contacted. For each included full-text study, Microsoft Excel was used to organize the following data: phenotype, genotype, SNP selection method, GWAS approach, identified significant (as reported by study) genomic variants (SNPs/CNVs/haplotype) and their coordinates, SNP effects, standard errors, and genes.

4.3.2. Annotation for Genes and QTLs

4.3.2.1. *Conversion of genomic coordinates to ARS-UCD1.2*

Genomic coordinates of associated markers and windows extracted from the included articles were converted to new bovine genome assembly ARS-UCD1.2 (Medrano, 2017; Rosen et al., 2020) using Lift Genome Annotations, animalgenome.org (ARS-UCD1.2 Cow Genome Assembly: Mapping of all existing variants) and European Variation Archive, depending on available information. Because most studies did not report significant markers or windows per parity related to resistance to mastitis traits, duplicate markers or windows per parity per trait were deleted. Significant markers/windows with unknown chromosome and with negative positions were also deleted. Following conversion, if the start position of a SNP was higher than the end position, the orientation of coordinates was reversed for windows. If a study had used UMD 3.1 *Bos taurus* assembly, it was converted to UMD 3.1.1 *Bos taurus* assembly before converting to ARS-UCD1.2 in Lift Genome Annotations.

4.3.2.2. *Annotation of markers and windows*

After all genomic coordinates were retrieved, genes and QTLs mapped within 0.1 Mb interval of markers and windows were annotated using GALLO (Fonseca et al., 2020). Markers and windows were annotated for genes and QTLs separately. Subsequently, annotated genes and QTLs from markers and windows were combined to obtain unique genes and QTLs that were used for further analysis. Gene and QTLs coordinates of the new bovine assembly ARS-UCD1.2 were obtained from Ensembl.org and Animal QTLdb, respectively. Following this step, lists of unique genes identified in each article were compared, and proportions of shared unique genes between studies and traits estimated. Annotated QTLs information were used to estimate the percentage of QTL type in respective QTL classes. Moreover, QTL enrichment analysis was performed for each trait individually and at a chromosome-wise level in the GALLO package using a hypergeometric test.

4.3.3. Gene Prioritization Analysis

A “guilt by association” based gene prioritization analysis (Walker et al., 1999) was conducted using “trained” (genes obtained from GUILDify software (Guney et al., 2014; Aguirre-Plans et al., 2019) and “test” genes’ list (positional candidate genes derived from annotation of chosen articles from the systematic review) using GUILDify and ToppGene. The top 100 ranked trained genes (after removing the 36 common genes with test gene list, GUILD score ≥ 0.7322) were obtained from GUILDify v2.0 Web Server. GUILDify uses a Biologic Interaction and Network Analysis (BIANA) knowledge database to retrieve genes associated with the outcome (mastitis/SCC related traits), using given keywords related to the outcome. Keywords used in the present study were; "mastitis," "subclinical mastitis," "somatic cell count," "somatic cell score," "SCC," "SCS", "intramammary infection," "lactation," "milk," and "mammary gland." Subsequently, GUILDify uses selected genes and species-specific (*Homo sapiens*) protein interaction network and applies graph theory algorithms to prioritize genes.

Only genes annotated from studies with ≥ 400 samples were used for the prioritization analysis to minimize a potential sample size effect. To obtain the test gene list, first, for the positional candidate genes without gene name, human orthologues in Ensembl database 106 were used to derive gene names. If the percentage of id. query gene identical to target human was ≥ 70%, the human gene name was used for further analyses. Secondly, final sets of test and trained gene lists were simultaneously used in ToppGene (Chen et al., 2009), which performs functional annotation-based prioritization analysis using a fuzzy-based multivariate approach to estimate the relationship between any 2 genes, based on semantic annotations. Multivariate analyses were performed using functional information shared between test and trained gene sets derived from various sources including: Gene Ontology (**GO**) terms for molecular function (**MF**), biological process (**BP**), and cellular component (**CC**); human and mouse phenotypes; metabolic pathways; PubMed publications; co-expression pattern; and diseases. An overall *P*-value per functional candidate was estimated using a statistical meta-analysis where *P*-values for each functional annotation of each gene were combined into a final *P*-value. Finally, prioritized genes were selected after false discovery rate (**FDR**) multiple correction of 5% from significant genes.

4.3.4. Gene Ontology, Metabolic Pathway, and Gene network Analyses

Gene ontology and metabolic pathway analyses for *Bos taurus* were performed using Over-Representation analysis in WebGestalt (Liao et al., 2019) for prioritized genes. Gene ontology analyses were performed for 3 GO term categories: BP, MF and CC using a non-redundant function database. Additionally, metabolic pathway analyses were performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway within the WebGestalt database. Gene network analyses were conducted for the list of prioritized genes using STRING database (Szklarczyk et al., 2019) using co-expression as an interaction source with a medium confidence (0.400) interaction score. The resulting network was imported into Cytoscape 3.9.0 (Shannonet al., 2003), and the Cytoscape network analyzer plugin with undirected network criteria was used for network analyses. Key genes (nodes) were selected from proteins with degree ≥ 8, betweenness centrality (**Bc**) ≥ 0.28 and closeness centrality (**Cc**) ≥ 0.62.

4.4. Results

4.4.1. Systematic Review

A summary of the search strategy is displayed in Figure 4-1. A total of 4,877 studies were identified from the databases (CAB Abstracts, MEDLINE, Web of Science and BIOSIS) and grey literature (115) search. After exclusion of duplicates (n = 2,828), 2,164 abstracts were screened separately by SN and EJ from which 1,471 studies were excluded. A total of 693 studies were included in the full-text review, where 641 were excluded due to facts such as: not original study, duplication, analysed population was not dairy cattle, outcome was not resistance to mastitis, did not use GWAS, or did not have full-text available. In total, 52 unique publications met our inclusion criteria. Descriptive details of these 52 articles, including country, breed and phenotype (Table 4-1), were extracted. Most of the studies were from Europe (34 studies) and north America (11 studies). Of the 52 manuscripts, 38 and 40 reported on studies using SCC-related traits as phenotype and Holstein breed, respectively, for GWAS analysis. We also extracted other details regarding genotyping (SNP-chip, assembly), quality control summary (call rate, minor allele frequency (MAF) and Hardy-Weinberg threshold) (Table 4-2), and GWAS summary, e.g., approach used for GWAS, number of animals and SNPs used for GWAS, and significant SNPs/windows identified (Table 4-3). Most of the studies used 50K SNP chip (31 studies) with 3 different versions. Of 52 manuscripts, 27 and 7 studies used assembly UMD 3.1 and UMD 3.1.1, respectively, whereas only 2 studies used the new bovine assembly ARS-UCD 1.2. Number of animals used for GWAS ranged from 34 to 293,467.

Moreover, significant markers or windows associated with CM and SCC-related traits and their coordinates were extracted for further analyses. Functional analyses, such as gene and QTL annotation, QTL enrichment analysis, GO terms enrichment, and KEGG pathway analyses, were conducted for 40 studies, with 37 and 7 studies reported markers and windows, respectively, and 4 studies reporting both. Twelve studies were excluded in further analysis due to not reporting the significant makers/windows or inability to convert reported coordinates to the new bovine assembly, due to inadequate information.

4.4.2. Annotation for Genes and QTLs

4.4.2.1. Conversion of genomic coordinates to ARS-UCD1.2

All markers and windows identified in the 40 studies are provided in Supplementary Material 2 (https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022). During conversion of genomic coordinates to new bovine assembly (ARS-UCD1.2) 1, 7, 1, and 1 significant markers from Jiang et al. (2019), Oliveira et al. (2019), Pegolo et al. (2020), and Tribout et al. (2020), respectively were deleted. Likewise, 8, 1, and 9 windows from Duran Aguilar et al. (2017), Meredith et al. (2013), and Oliveira et al. (2019), respectively, were deleted during the conversion. These markers and windows were deleted due to one of the following reasons: previously reported genomic coordinates did not align well with the new bovine assembly or partially aligned, but not enough to convert, or the region was split up into various parts. After conversion, an additional 8 markers were removed from the dataset as they did not have chromosome number (7 from Cole et al. (2011) and 1 from Strillacci et al. (2014)) (Supplementary Material 2; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022). Twenty-nine markers were removed as they had negative positions (2 from Cole et al. (2011), 26 from Strillacci et al. (2014) and 1 from Wang et al. (2015). Moreover, change in chromosome number was reported for 38 markers (5 articles) and 5 windows (2 articles) during conversion. For 6 windows, start and end positions were reversed as start position was greater than end position. Finally, out of 40 studies qualified for further quantitative analysis, 37 studies resulted in 3,949 markers (3,708 unique markers) and 7 studies resulted in 499 windows (275 unique windows), with 4 studies (Abdel-Shafy et al. (2014), Fang et al. (2018), Kurz et al. (2019), Meredith et al. (2013)) reporting both markers and windows.

4.4.2.2. Annotation of markers and windows

A total of 19,403 positional candidate genes (10,624 genes from markers and 8,779 genes from windows) were annotated for markers and windows, resulting in a total of 9,125 unique positional candidate genes within 0.1 Mb interval from 39 studies with 1 gene in a sex chromosome (Supplementary Material 3; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022). One SNP from Freebern et al. (2020) was not annotated for either genes or QTLs. Many annotated genes were in BTA 7 (1,264 genes) followed by BTA 19 (626 genes) and 18 (469 genes). In general, the proportion of genes shared among the 39 studies was low (mean ± SD = 0.02 ± 0.07) (Supplementary Material 4; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022). Fang et al. (2017a) shared a high proportion of genes with Fang et al. (2018) and Cai et al. (2018) (1 and 0.62, respectively). Likewise, Wijga et al. (2012) shared 0.75 of genes with Mulder et al. (2013). Moreover, genes *GC*, ENSBTAG00000049290 and *NPFFR2* were reported by 9, 8, and 7 studies, whereas the remaining genes were reported by ≤ 6 studies.

In terms of traits associated with resistance to mastitis, 52 types of traits were derived from 39 studies. These traits were mainly grouped into CM and SCC-related traits. Many annotated genes (8,664; unique genes per trait) were related with SCC-related traits, whereas fewer genes (930; unique genes per trait) were associated with CM-related traits (Figure 4-2a). However, 469 genes were shared between those 2 traits. Among 13 studies that conducted CM-related GWAS, in general, gene overlap among studies was low (mean ± SD = 0.02 ± 0.10) (Figure 4-2b). However, among those studies, Sahana et al. (2013) and Fang et al. (2018) shared 39 genes. Similarly, gene overlap across studies which performed GWAS on SCC-related traits was in general also low (mean ± SD =0.01 ± 0.06) (Supplementary Material 5; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022). Interestingly, Duran Aguilar et al.

(2017) who used CNV of SCS to perform GWAS shared a high number of genes with Cole et al. (2011) and Oliveira et al. (2019). In terms of breeds used for GWAS analysis of resistance to mastitis, there were 5,843, 2,311 and 1,915 genes annotated for Holstein, Jersey and Ayrshire, respectively (the 3 main breeds) (Figure 4-3). Furthermore, 74 common genes were identified among the 3 main breeds in the present study.

QTL annotation yielded 132,140 QTLs (87,737 QTLs from markers and 44,403 QTLs from windows) which resulted in 43,646 unique QTLs from 39 studies. A high number of QTLs of 8,513, 6,450 and 2,780 were located in BTA 14, 6 and 20, respectively. A high proportion of 64% of annotated QTLs were associated with milk QTL class, whereas only 7.5% was associated with health QTL class (Figure 4-4). Moreover, within health QTL class, 2.6 and 2.2% of QTLs were associated with CM and SCC-related traits (SCC_SCS) (Figure 4-5). Interestingly, a low percentage of QTLs was also related with susceptibility to other diseases such as ketosis, bovine tuberculosis, and bovine respiratory disease. QTL enrichment analysis demonstrated that a high number of enriched QTLs was associated with SCS in BTA 5, 6, 16, and 20 (Figure 4-6). Additionally, QTL enrichment analysis for CM resulted in fewer significant QTLs on BTA 8, 10, 18, and 25.

4.4.3. Gene Prioritization Analysis

Thirty studies (Supplementary Material 3; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022) with ≥ 400 samples were used for the prioritization analysis, which resulted in 7,280 unique positional candidate genes (5,188 with unique gene symbol). A total of 1,689 genes from 7,280 unique positional candidate genes did not have a gene symbol. Therefore, the human gene symbol was retrieved for 197 unique genes from ensembl if the query gene was ≥ 70% identical to the human gene. Thirty-six genes of the 197 genes were common with 5,188 genes. After removing the duplicate genes, a total of 5,349 test genes with gene symbol (5,188 + 197-36) and the top 100 trained genes (Supplementary Material 3; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022) were used as input in the ToppGene software for gene prioritization analysis. However, only 5,008 test genes (Supplementary Material 3; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022) were used for the analyses, due to unavailability of some of the genes in the ToppGene software. Gene prioritization analysis resulted in 427 significant genes (Supplementary Material 3; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022) after multiple test correction (5% FDR) from 26 studies. Most prioritized genes were located in BTA 19 (40 genes) and 7 (36 genes). Most prioritized genes (197; 46%) were from the cytokine superfamily (e.g., chemokines, interleukins, transforming growth factors, and tumor necrosis factor). Additionally, among 427 prioritized genes, *ARC, GNRHR, MC3R* and *PRLR* genes were reported in 4 studies, whereas 13 and 54 genes were reported in 3 and 2 studies, respectively, whereas the remainder were reported in only 1 study, which were all different studies. In addition, most prioritized genes (397) were associated with SCC-related traits, which included top 5 ranked genes *CCR2*, *CCL2*, *CCR5*, *TLR4* and *GRIN2A*. Only 54 genes were directly associated with CM. Twenty-four genes (*ABCC9*, *ACHE*, *ADCYAP1*, *ARC*, *BCL2L1*, *CDKN1A*, *EPO*, *GABBR2*, *GDNF*, *GNRHR*, *IKBKE*, *JAG1*, *KCNJ8*, *KCNQ1*, *LIFR*, *MC3R*, *MYOZ3*, *NFKB1*, *OSMR*, *PPP3CA*, *PRLR*, *SHARPIN*, *SLC1A3* and *TNFRSF25*) were reported for both SCC and CM-related traits. Gene *KCNQ1* was among the top 50 ranked genes (Figure 4-7). The greatest number of prioritized genes were reported for Holsteins (243), and 139 and 102 genes were reported for Jersey and Ayrshire, respectively, while 5 genes (*ARC*, *IL7*, *ITPR2*, *MAP2K1* and *MC3R*) were shared among all 3 breeds.

4.4.4. Gene Ontology, Metabolic pathway, and Gene Network Analyses

Enriched GO terms for BP (100), CC (90) and MF (101), and KEGG (100) pathways for prioritized genes were derived from WebGestalt (Supplementary Material 3; https://doi.org/10.7910/DVN/HNKBJS; Narayana et al., 2022). Biological process GO terms were mainly associated with positive regulation of intracellular signal transduction, response to biotic and external stimulus, immune system, locomotion, secretion, response to and production of cytokine, inflammatory response, and cell proliferation and differentiation (FDR < $1e^{-15}$). Enriched CC (FDR < $1e^{-15}$) were cell surface, plasma membrane protein complex and receptor complex. The top 2 MF (FDR < $1e^{-15}$) were receptor regulator activity and cytokine receptor binding. Top 8 significant KEGG pathways (FDR < $1e^{-15}$; Table 4-4) were pathways in cancer, PI3K-Akt signaling pathway, cytokine-cytokine receptor interaction, neuroactive ligand-receptor interaction, MAPK signaling pathway, Kaposi sarcoma-associated herpesvirus infection, JAK-STAT signaling pathway and Proteoglycans in cancer.

A co-expression gene network included 48 genes (nodes) with 92 interactions (edges) with network density, clustering coefficient and protein-protein interactions (PPI) enrichment P-value of 0.279, 0.236, and < $1e^{-16}$, respectively (Figure 4-8). Based on degree ≥ 8, Bc ≥ 0.28 and Cc ≥ 0.62, 6 key genes (*CSF1R*, *CD3E*, *CXCL10*, *STAT1*, *CD3D* and *IRF1*) were identified. Gene *CSF1R* had highest the degree of 12 (Bc=0.93, Cc=1.00) and was the most interconnected and powerful gene in the co-expression network.

4.5. Discussion

This systematic review identified 52 unique GWAS on resistance to mastitis, of which 30 (sample size ≥ 400) were used for further gene prioritization analysis. To our knowledge, this is the most comprehensive review of this topic to date and provides valuable insight into the gene architecture of resistance to bovine mastitis. Heterogeneity was high among the 52 included manuscripts, in terms of population, breed, phenotype, number of animals used, genotyping, GWAS quality control and GWAS approach. Most studies were from Europe (34 studies), with fewer from North America (11 studies). Holstein was the most common breed, as it is the most prevalent breed in dairy farming, particularly in North America and Europe. Many studies (38) used SCC-related traits for GWAS analysis, as SCC is recorded regularly, readily available, and cost-effective compared to obtaining the CM phenotype. The number of animals included varied vastly (34 to 293,467) depending on the trait used. For instance, Siebert et al. (2018) (number of samples=34) used time to SCC 'cure' post *S. uberis* challenge, whereas Jiang et al. (2019) (number of samples=293,467) used SCS of first lactation cows for GWAS. As identification of significant genomic variants in a GWAS differs depending on the traits, hence genes associated with mastitis resistance-related traits will also differ. Therefore, we did not impose any restriction on number of samples, GWAS approach, software and population/breed used for the GWAS during the annotation process to capture the breadth of published QTLs and genes. Heterogeneity among studies was considered 1 reason for the low overlap of genes among studies. Since many (31 out of 40 studies) studies used SCC-related traits for GWAS analysis, a high number of annotated genes were related with SCC-related traits compared to CM-related ones. However, in this study, 469 genes were common to both SCC-related traits and CM, indicating that even though SCM and CM were controlled largely by separate genes, some genes regulated both SCM and CM.

Annotation analysis of QTL resulted in a high percentage (64%) of QTLs in the “milk” QTL class and low percentage (7.5%) in the “health” QTL class, perhaps due to publication bias. Milk-related traits are more researched and published, resulting in more QTLs in the “milk” QTL class database. To account for this bias, QTL enrichment analysis was performed, where a hypergeometric test was conducted using the number of annotated QTLs within the candidate regions and the total number of the same QTL in the QTL database using the GALLO (Fonseca et al., 2020). Moreover, within the "health" QTL class, annotated QTLs were also associated with other health traits. This provided evidence for genetic correlation for susceptibility between mastitis and other diseases such as ketosis and bovine tuberculosis (Bermingham et al., 2010; Koeck et al., 2012) and pleiotropic effects of these regions over various health traits.

Gene prioritization analysis conducted in this study used several sources of functional information, as follows: GO terms for MF, BP, and CC; human and mouse phenotypes; metabolic pathways; PubMed publications; co-expression pattern; and diseases to prioritize genes, which promoted identifying interesting functional candidate genes. Many prioritized genes (197; 46%) were from the cytokine superfamily, such as chemokines, interleukins, transforming growth factors, and tumor necrosis factor genes. Inflammatory mediators such as chemokines and other cytokines produced by mammary epithelial cells, drive leucocytes from the blood stream to mammary gland to confer protection against invading foreign pathogens that cause mastitis (Rainard and Riollet, 2006).

Chemokines and their receptors have vital roles in development and homeostasis of the immune system, especially in maintaining innate and adaptive immunity (Gangur et al., 2002; Raman et al., 2011; Hughes and Nibbs, 2018). Chemokines belong to a family of chemoattractant cytokines, small structurally related proteins secreted by cells (Matsushima, 2000; Poeta et al., 2019). Eighteen chemokines and their receptors were identified among prioritized genes. The first ranked chemokine receptor *CCR2* (C-C Motif Chemokine Receptor 2) gene located on BTA 22 was identified by Qanbari et al. (2014) for the Fleckvieh breed and was associated with SCC-related traits. Interestingly, Leyva-Baca et al. (2007) demonstrated that the *CCR2* gene had a significant (at the comparison-wise level) allele substitution effect (-0.04 ± 0.02, $P = 0.05$) on SCS in Canadian Holstein bulls. However, in the present study, *CCR2* was not annotated for the Holstein breed. Perhaps manuscripts included in this study that used Holstein breed for GWAS did not identify genomic coordinate (52945500 bp) that harbour the *CCR2* gene. Some of the interesting BP non-redundant GO terms related with *CCR2* were immune response related, regulation of response to external stimulus, cell activation and adhesion secretion, and locomotion (FDR < $1e^{-15}$). Molecular process GO terms associated with *CCR2* were cytokine receptor binding, G protein-coupled receptor binding, cytokine receptor activity, and cytokine binding (FDR ≤ $2.69e^{-7}$). Moreover, CCR2 was involved in Cytokine-cytokine receptor interaction, and Chemokine signaling KEGG pathways (FDR ≤ $2.90e^{-14}$).

Other interesting chemokine and receptor genes were *CCL2*, *CCR5*, *CXCL10, CCR6, CXCL9, CXCL12*, *CCR10, CCL3, CXCR5, CCRL2, CCL11*, *CCL4, CXCL11*, *CCL28, CXCL8, CCL24* and *CCL26*. Genes *CCL2* and *CXCL10*, which are involved in initial recruitment of leucocytes, were more highly expressed in bovine mammary epithelial cells with mastitis induced by *Escherichia coli* than with mastitis induced by *Staphylococcus aureus* (Gilbert et al., 2013). Genes *CCL2, CXCL10* and *CCL3* were significantly upregulated after an intramammary infusion of lipopolysaccharide containing *E. coli* in a mouse model (Zheng et al., 2006). Chemokine receptor CCR5 is the receptor for leukocidins secreted by *S. aureus* such as γ-haemolysin (HlgAB) and leukotoxin ED (LukED) during bovine mastitis (Alonzo III et al., 2013; Vrieling et al., 2016). C-C Motif Chemokine Receptor 6 (*CCR6*) only binds with *CCL20* chemokine (Schutyser et al., 2003), expressed in udder and mammary epithelial cells during *E. coli*-induced mastitis (Günther et al., 2009; Günther et al., 2010), and the ligand-receptor pair is involved in a humoral immune response (Schukken et al., 2011). Recently, *CXCL9* was reported to be a potential molecular marker of mastitis caused by infection with *S. aureus* in Chinese Holstein dairy cattle (Wang et al., 2020). Gene *CXCL12* was highly differentially expressed between control quarters from cows infected with *S. aureus* than by *E. coli* (Jensen et al., 2013). Expression of *CCRL2* was reported in the mammary gland and mammary epithelial cells (Suzuki et al., 2015). Additionally, *CCL4* and *CXCL8* genes were identified in mammary tissue of cattle as hub genes with important roles in inflammation and immune response (Gorji et al., 2019).

Approximately 17 interleukin and receptor-related genes were among 427 prioritized genes. Protein encoded by Interleukin 6 Cytokine Family Signal Transducer (*IL6ST*) on BTA20, associated with SCC-related traits, was ranked 11th among prioritized genes. The protein encoded by *IL6ST* binds with pro-inflammatory cytokines such as *IL6* which has an important role in immune response during mastitis. Concentrations of *IL6* were higher in milk and blood of cows with SCM infected with coagulase-negative staphylococci (432.09 and 254.32 pg/mL) compared to healthy cows (164.47 and 13.02 pg/mL, respectively) (Bochniarz et al., 2017). In the present study, *IL6* was not annotated; however, only IL6ST was identified from Meredith et al. (2013) in Holsteins. Perhaps genomic coordinates that harbor IL6 were not reported in studies that included annotation analysis. Other interleukin genes (e.g., those coding for *IL15, IL5, IL10, IL3, IL7, IL19, IL20, IL24*) and its receptors (e.g., *IL12RB1, IL23A, IL7R, IL10RB, IL31RA, IL10RA, IL17B* and *IL21R*) were some of the interesting prioritized genes. Interleukins 15 gene was differentially expressed in blood mononuclear cells and milk somatic cells of cows with chronic S. *aureus* mastitis versus healthy controls in Canadian Holsteins cows (Tao and Mallard (2007)). Interleukins 10 gene was highly expressed in Holstein and Gyr cows with CM compared to healthy cows (Fonseca et al., 2009). A similar result was also reported by Faaz and Abdullah (2022), where authors stated that the expression of IL10 was higher ($P < 0.001$) in cows with CM than in healthy animals. Interleukins 19 gene was upregulated in the breast milk somatic cells of women with mastitis as a response to probiotic (De Andres et al., 2018).

Genes *ARC*, *GNRHR*, *MC3R* and *PRLR* were identified in 4 studies, whereas other prioritized genes were identified in < 4 studies. Gene *ARC* located on BTA 14 was identified by Oliveira et al. (2019b), Oliveira et al. (2019a), Tiezzi et al. (2015), and Wang et al. (2015). Additionally, *ARC* was ranked 402nd among prioritized genes and some BP GO terms related with *ARC* genes were positive regulation of transport, regulation of signaling receptor activity, ion transport and transmembrane transport (FDR < $1e^{-15}$). Interestingly, *ARC* was associated with both CM and SCC-related traits and identified in both Holstein, Jersey and Ayrshire cows.

Prolactin Receptor (*PRLR*) (142nd ranked prioritized gene) located on BTA20 is a protein coding gene that belongs to the type I cytokine receptor family; it was identified by Sahana et al. (2013), Jiang et al. (2019), Oliveira et al. (2019a), and Oliveira et al. (2019b). Gene *PRLR* was associated with both CM and SCC-related traits and identified in both Holsteins and Jerseys. Prolactin Receptor (*PRLR*) transmits signal from prolactin to milk protein genes (Brym et al., 2005). Many studies reported associations of *PRLR* with milk performance traits (Brym et al., 2005; Vitala et al., 2006; Zhang et al., 2008). Fontanesi et al. (2014) reported that gene *PRLR* was associated with milk, fat yield, protein yield, protein percentage, SCC, longevity index, fertility index and productivity, functionality, and type index, expect for fat percentage. Authors also stated that favourable alleles for yield traits were unfavourable for SCC and longevity and fertility index. In the present study, the *PRLR* gene was involved in cytokine receptor activity MF GO terms and PI3K-Akt signaling pathway, Cytokine-cytokine receptor interaction, neuroactive ligand-receptor interaction and JAK-STAT signaling pathway KEGG pathway (FDR < $1e^{-15}$).

Gonadotropin Releasing Hormone Receptor (*GNRHR*; 255th) on BTA6 was identified in Abdel-Shafy et al. (2014), Sahana et al. (2013), Cole et al. (2011) and Jiang et al. (2019). Gene *GNRHR* was associated with both CM and SCC-related traits and was identified only in Holstein in the current study. Gene *GNRHR* was reported to be associated with SCC in seven different population of Holsteins (Chen et al., 2015) and mastitis resistance in Canadian Holsteins (Grossi et al., 2014). Melanocortin 3 Receptor (*MC3R*; 346th) gene on BTA13 was reported by Oliveira et al. (2019b), Cole et al. (2011), Fang et al. (2018) and Sahana et al. (2013) as being associated with SCC-related traits and CM in Ayrshire, Holsteins and Jersey breeds, respectively. However, *MC3R* gene has not been explored widely related to mastitis.

Most prioritized genes (397) were associated with SCC-related traits, as many studies used SCC-related traits for their GWAS. In that regard, obtaining the SCC related phenotype is easier and cheaper compared to obtaining CM phenotype. In the present study, 24 genes (*ABCC9, ACHE, ADCYAP1, ARC, BCL2L1, CDKN1A, EPO, GABBR2, GDNF, GNRHR, IKBKE, JAG1, KCNJ8, KCNQ1, LIFR, MC3R, MYOZ3, NFKB1, OSMR, PPP3CA, PRLR, SHARPIN, SLC1A3* and *TNFRSF25*) were identified as associated with both CM and SCC-related traits. From these genes, *KCNQ1* was among the top 50 prioritized genes (38th). This gene is located on BTA29 and was identified by Cole et al. (2011) and Marete et al. (2018b) in the Holstein breed as associated with CM and SCC-related traits. This gene is a protein coding gene and encodes a voltage-gated potassium channel required for repolarization phase of the cardiac action potential (Stelzer et al., 2016). In human mammary epithelium, *KCNQ1* gene plays an important role in regulating cell volume and potentially mediating transepithelial K^+ secretion (VanTol et al., 2007). Moreover, most of the top 50 prioritized genes were from Oliveira et al., 2019a and b. These authors used the same population of three main breeds (Holsteins, Jersey and Ayrshire) for both analyses, and as a result shared more genes.

Most annotated genes (243) were associated in Holsteins, the most prevalent dairy breed. In the present study, genes *ARC*, *IL7, ITPR2, MAP2K1* and *MC3R* were common for Holsteins, Jerseys, and Ayrshires. Gene *IL7* is located on BTA14 and ranked 186th among prioritized genes. Interleukin 7 is a cytokine protein coding gene, and its protein is important for development of both B and T cells (Stelzer et al., 2016). Similar to the current study, *IL7* gene was identified to be associated with SCC in 7 different Holstein population (Chen et al., 2015). In the present study, *IL7* gene was reported to be involved in pathway in cancer, PI3K-Akt signaling pathway, cytokine-cytokine receptor interactions, and JAK-STAT signaling pathway (FDR < $1e^{-15}$) KEGG pathways and positive regulation of intracellular signal transduction, immune system development, positive regulation of immune system process, response to cytokine, and regulation of cell activation, adhesion and differentiation BP GO terms (FDR < $1e^{-15}$). Gene *ITPR2* on BTA5 (342nd) encodes a protein that mediates a rise in cytoplasmic calcium in response to receptor activated production of inositol triphosphate (Stelzer et al., 2016). Gene *ITPR2* was reported to be associated with SCM in Norwegian Red cattle (Kirsanova et al., 2020). Gene *MAP2K1* on BTA10 encodes kinase protein, which is involved in many cellular processes, such as proliferation, differentiation, transcription regulation and development (Stelzer et al., 2016). Gene *MAP2K1* was identified in dairy cows with mastitis caused by *E. coli* (Li et al., 2021) and *Staph. aureus* (Wang et al., 2021). In the current study, gene *MAP2K1* was involved in pathways in cancer, PI3K-Akt signaling pathway, MAPK signaling pathway, Kaposi sarcoma-associated herpesvirus infection and proteoglycans in cancer KEGG pathways (FDR < $1e^{-15}$). However, LD phase across breeds needs to be considered when utilizing these genes for genomic predictions across breeds (De Roos et al., 2008).

Gene network analysis is important in understanding the shared biological processes among prioritized candidate genes. Gene network analysis based on co-expression as interaction source provides insights into genes that are expressed in similar magnitude for similar stimulus and are controlled by the same pathway or transcriptional regulatory program and are functionally related (Weirauch, 2011). Of 427 prioritized genes in this study, 48 were identified as co-expressed during mastitis and were primarily involved in various immune-related pathways: pathways in cancer, cytokine-cytokine receptor interaction, PI3K-Akt signaling pathway, MAPK signaling KEGG pathways primary immunodeficiency, chemokine signaling, T cell receptor signaling, toll-like receptor signaling and Th17 cell differentiation (FDR $\leq 1.74e^{-5}$). Six key genes (*CSF1R*, *CD3E*, *CXCL10*, *STAT1*, *CD3D* and *IRF1*) identified in co-expression network analysis were more informative, as they had high interactions with other co-expressed genes. Colony stimulating factor 1 receptor (*CSF1R*;153rd) (degree=12, Bc=0.93, Cc=1.00) is a cytokine receptor that encodes protein that serves as receptor for colony stimulating factor 1 (*CSF1*), with important roles in production, differentiation, and function of macrophages (Stelzer et al., 2016; Sayers et al., 2021). Thus, the *CSF1R* gene has an important part in the process of innate immunity and inflammation. Gene *CSF1R* has been associated with mastitis in cows, sheep, and swine (Sajjanar et al., 2019; Hu et al., 2020; Wathes et al., 2021). The protein encoded by *CD3E* (53rd) plays an essential role in T-cell development and fault in this gene can cause immunodeficiency ((Stelzer et al., 2016). Gene *CD3E* has also been identified as one of the most interconnected hub gene related with bovine mastitis (Bakhtiarizadeh et al., 2020). Signal Transducer And Activator Of Transcription 1 (*STAT1*; 16th) is a member of STAT protein family. The protein encoded by *STAT1* gene has an important part in immune responses to viral, fungal, and mycobacterial pathogens (Stelzer et al., 2016). The role of *STAT1* gene in mastitis resistance and milk production had been reported in several studies (Cobanoglu et al., 2006; Bakhtiarizadeh et al., 2020; Ghahramani et al., 2021). Protein encoded by *CD3D* gene (163rd) is part of T-cell receptor and defects in this gene will lead to severe immunodeficiency issues (Stelzer et al., 2016). Gene *CD3D* has been reported by several studies as an important gene involved in immune response and mastitis resistance in Holstein cows (Bakhtiarizadeh et al., 2020) and sheep (Bonnefont et al., 2011). Gene *IRF1* (332nd) produces a protein which serves as an activator of genes involved in both innate and acquired immune responses (Stelzer et al., 2016). The *IRFI* gene has been reported as a potential marker for the detection of mastitis in cows (Huma et al., 2020).

4.6. Limitations

The main limitation of this study was that our systematic review was confined to GWAS, a common high-throughput technology to identify genetic variants associated with traits of interest. Studies that had used other technologies including molecular technology, validation of makers associated with resistance to mastitis, and other omics technologies (e.g., high-throughput transcriptome profiling technologies that uses microarray and RNA sequencing), were not included. Hence, the present study may not have included genomic coordinates of all significant markers or windows associated with resistance to mastitis. Therefore, all genes associated with resistance to mastitis may not have been annotated and used in further analyses. Additionally, during conversion of genome assembly to new bovine assembly (ARS-UCD1.2) some studies were lost due to an inability to convert. Another limitation is that 1,689 genes did not have a gene name required to be included in prioritization analyses, so their names had to be retrieved from ensembl using human orthologs. During this process, names of only 374 genes (197 unique gene symbols) were retrieved, with loss of 1,315 genes. Hence, some genes without a name may not be included in the prioritization analysis.

The GUILDify software, which was used to obtain the trained gene list used for the prioritization analysis, uses three algorithms (NetScore, NetZcore, and NetShort) available to prioritize genes potentially involved in diseases using a priori disease-gene associations and protein-protein interactions (PPIs). This might introduce some level of bias of trained gene list related to diseases in the initial stage of prioritization analysis. However, as the ToppGene software ranks the positional candidate genes (test genes) based on the functional similarity between each gene from the list of positional candidate genes and functional profile of the trained list, impact of above-mentioned bias on the prioritization analysis should not be significant. As the software GUILDify does not have a database for cattle, human (*Homo sapiens*) protein interaction network was used to prioritize the genes. However, the use of the human database instead of cattle database is not expected to have substantial impact on the results as cattle share 80% of human orthologs (Weitzman, 2000; Bovine Genome Sequencing and Analysis Consortium, 2009). Moreover, the conservation between protein structures is expected to be higher than to gene sequence. Consequently, the interactions between proteins tend to be more conserved (Kwon et al., 2018; Pérez-Bercoff et al., 2013).

4.7. Implications

Many studies have conducted GWAS to identify genetic variants associated with resistance to mastitis-related traits and they had reported numerous genetic variants. These genetic variants differed based on population, breed, traits, genotyping, and approach used for GWAS. In the present study, heterogeneity among these articles is highlighted in Tables 4-1 to 4-3. This heterogeneity among articles emphasized the importance of and need for a systematic review and reinforced the need to confirm previously identified genetic variants associated with resistance to mastitis.

This systematic review identified studies that performed GWAS on resistance to mastitis-related traits and summarized them into one comprehensive document, extracted significant variants reported on those studies, and annotated to genes and QTLs. Moreover, in the present study gene prioritization analyses and other functional analyses were performed to identify interesting functional candidate genes. This study also mapped previously reported genomic regions and will inform evidence-based makers for genetic selection for mastitis resistance. However, these candidate genes and their markers should be validated before being used in the SNP chip panels for genetic selection for resistance to mastitis. This study opens the possibility to further evaluate in depth the key functional candidate regions identified. For example, it is possible to perform target sequencing of those regions or use whole genome sequences to identify causal mutations and evaluate the potential impact using tools like Variant Effect Predictor (McLaren et al., 2016).

Moreover, this study greatly enhances the current knowledge of the underlying genetic architecture of resistance to mastitis, thereby facilitating development of better genomic selection methods to improve mastitis resistance in dairy cattle.

4.8. Conclusions

In this first systematic review of GWAS for mastitis resistance-related traits in dairy cattle, we identified 52 relevant manuscripts, of which 30 qualified for gene prioritization analysis. There was a high heterogeneity in populations, phenotypes, genotypes, and methods used for GWAS, resulting in the annotation of 9,125 and 43,646 unique genes and QTLs, respectively. The high heterogeneity among GWAS led to lower overlap of genes across studies. Subsequent gene prioritization analysis resulted in 427 prioritized genes, among which the *CCR2* gene ranked first. Most prioritized genes were from the cytokine family and were related to the immune response. Future studies are needed to validate and identify SNPs in these key candidate genes, which may enhance genomic prediction. Moreover, this study confirmed and fine-mapped previously identified genomic regions and candidate genes associated with mastitis resistance, providing additional insights into the genetic architecture of mastitis resistance in dairy cattle. Consequently, the implementation of results from this study is expected to reduce mastitis incidence and prevalence in dairy cattle through genetic selection in the medium and long term.

4.9 References

Abdel-Shafy, H., R. H. Bortfeldt, J. Tetens, and G. A. Brockmann. 2014. Single nucleotide polymorphism and haplotype effects associated with somatic cell score in German Holstein cattle. Genet. Sel. Evol. 46:35.

Aghamohammadi, M., D. Haine, D. F. Kelton, H. W. Barkema, H. Hogeveen, G. P. Keefe, and S. Dufour. 2018. Herd-level mastitis-associated costs on Canadian dairy farms. Front. Vet. Sci. 5:100.

Aguilar I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Proc. 10th World Congr. Genet. Appl. Livest. Prod. Accessed February 2, 2022. https://hal.inrae.fr/hal-02743809.

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

Aguirre-Plans, J., J. Piñero, F. Sanz, L. I. Furlong, N. Fernandez-Fuentes, B. Oliva, and E. Guney. 2019. GUILDify v2.0: A tool to identify molecular networks underlying human diseases, their comorbidities and their druggable targets. J. Mol. Biol. 431:2477-2484.

Alonzo III. F., L. Kozhaya, S. A. Rawlings, T. Reyes-Robles, A. L. DuMont, D. G. Myszka, N. R. Landau, D. Unutmaz, and V. J. Torres. 2013. CCR5 is a receptor for *Staphylococcus aureus* leukotoxin ED. Nature 493:51-55.

Amin, N., C. M. van Duijn, and Y. S. Aulchenko. 2007. A genomic background based method for association analysis in related individuals. PLoS One 2:e1274.

Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn. 2007. GenABEL: An R library for genome-wide association analysis. Bioinformatics 23:1294–1296.

Bagnato, A., F. Schiavini, A. Rossoni, C. Maltecca, M. Dolezal, I. Medugorac, J. Sölkner, V. Russo, L. Fontanesi, A. Friedmann, M. Soller, and E. Lipkin. 2008. Quantitative trait loci affecting milk yield and protein percentage in a three-country Brown Swiss population. J. Dairy Sci. 91:767–783.

Bakhtiarizadeh, M. R., S. Mirzaei, M. Norouzi, N. Sheybani, and M. S. V. Sadi. 2020. Identification of gene modules and hub genes involved in mastitis development using a systems biology approach. Front. Genet. 11:722.

Bermingham, M. L., S. J. More, M. Good, A. R. Cromie, I. M. Higgins, and D. P. Berry. 2010. Genetic correlations between measures of *Mycobacterium bovis* infection and economically important traits in Irish Holstein-Friesian dairy cows. J. Dairy Sci. 93:5413-5422.

Berry, D. P., L. Shalloo, A. R. Cromie, R. F. Veerkamp, D. Dillon, P. R. Amer, J. F. Kearney, R. D. Evans, and B. Wickham. 2007. The economic breeding index: A generation on. Pages 29-34 in Tech. Rep. to the Irish Cattle Breeding Federation. 29-34. Accessed February 2, 2022. http://www.icbf.com/publications/files/economic_breeding_index.pdf.

Bochniarz, M., B. Zdzisińska, W. Wawron, M. Szczubiał, and R. Dąbrowski. 2017. Milk and serum IL-4, IL-6, IL-10, and amyloid A concentrations in cows with subclinical mastitis caused by coagulase-negative staphylococci. J. Dairy Sci. 100:9674-9680.

Bonnefont, C., M. Toufeer, C. Caubet, E. Foulon, C. Tasca, M. R. Aurel, D. Bergonier, S. Boullier, C. Robert-Granié, G. Foucras and R. Rupp. 2011. Transcriptomic analysis of milk somatic cells in mastitis resistant and susceptible sheep upon challenge with *Staphylococcus epidermidis* and *Staphylococcus aureus*. BMC Genom. 12:1-16.

Bovine Genome Sequencing and Analysis Consortium, C. G. Elsik, R. L. Tellam, and K. C. Worley. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324:522–528.

Brym, P., S. Kaminski, and E. Wójcik. 2005. Polymorphism within the bovine prolactin receptor gene (PRLR). Anim. Sci. Pap. Rep. 23:61-66.

Cai, Z. X., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2019. Weighting sequence variants based on their annotation increases the power of genome-wide association studies in dairy cattle. Genet. Sel. Evol. 51:12.

Cai, Z., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2018. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. BMC Genom. 19:656.

Calus, M. P. L., T. H. E. Meuwissen, A. P. W. de Roos, and R. F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561.

Canadian Dairy Network. 2019. 10 Years of genomic selection: What's next? Accessed February 2, 2022. https://www.cdn.ca/document.php?id=530.

Cecchinato, A., N. P. P. Macciotta, M. Mele, F. Tagliapietra, S. Schiavon, G. Bittante, and S. Pegolo. 2019. Genetic and genomic analyses of latent variables related to the milk fatty acid profile, milk composition, and udder health in dairy cattle. J. Dairy Sci. 102:5254–5265.

Chen, J., E. E. Bardes, B. J. Aronow, and A. G. Jegga. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 37:W305–W311.

Chen, X., Z. Cheng, S. Zhang, D. Werling, and D. C. Wathes. 2015. Combining genome wide association studies and differential gene expression data analyses identifies candidate genes affecting mastitis caused by two different pathogens in the dairy cow. Open J. Anim. Sci. 5:358.

Christensen, O. F. and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2.

Cobanoglu, O., I. Zaitoun, Y. M. Chang, G. E. Shook, and H. Khatib. 2006. Effects of the signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. J. Dairy Sci. 89:4433–4437.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, B. A. Crooker, C. P. Van Tassell, J. Yang, S. W. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. BMC Genom. 12:408.

De Andres, J., E. Jiménez, I. Espinosa-Martos, J. M. Rodríguez, and M. T. García-Conesa. 2018. An exploratory search for potential molecular targets responsive to the probiotic *Lactobacillus salivarius* PS2 in women with mastitis: gene expression profiling vs. interindividual variability. Front. Microbiol. 9: 2166.

De Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics. 179:1503–1512.

Duran Aguilar, M., S. I. Roman Ponce, F. J. Ruiz Lopez, E. Gonzalez Padilla, C. G. Vasquez Pelaez, A. Bagnato, and M. G. Strillacci. 2017. Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. J. Anim. Breed. Genet. 134:49–59.

Fang, L. Z., G. Sahana, G. S. Su, Y. Yu, S. L. Zhang, M. S. Lund, and P. Sorensen. 2017a. integrating sequence-based GWAS and RNA-seq provides novel insights into the genetic basis of mastitis and milk production in dairy cattle. Sci. Rep. 7:16.

Fang, L. Z., G. Sahana, P. P. Ma, G. S. Su, Y. Yu, S. L. Zhang, M. S. Lund, and P. Sorensen. 2017b. Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. BMC Genom. 18:12.

Fang, L. Z., P. Sorensen, G. Sahana, F. Panitz, G. S. Su, S. L. Zhang, Y. Yu, B. J. Li, L. Ma, G. Liu, M. S. Lund, and B. Thomsen. 2018. MicroRNA-guided prioritization of genome-wide association signals reveals the importance of microRNA-target gene networks for complex traits in cattle. Sci. Rep. 8:14.

Fang, Z. H. and H. Pausch. 2019. Multi-trait meta-analyses reveal 25 quantitative trait loci for economically important traits in Brown Swiss cattle. BMC Genom. 20:695.

Faaz, R. A., and F. A. Abdullah. 2022. Expressions of IL4, IL10, and IFNγ cytokines genes during bacterial mastitis. J. Adv. Vet. Anim. Res. 9:42.

Fonseca, I., P. V. Silva, C. C. Lange, M. F. M. Guimarães, M. M. D. C. A. Weller, K. R. S. Sousa, P. S. Lopes, J. D. Guimarães, and S. E. F Guimarães. 2009. Expression profile of genes associated with mastitis in dairy cattle. Genet. Mol. Biol. 32: 776-781.

Fonseca, P. A. S., A. Suárez-Vega, G. Marras, and A. Cánovas. 2020. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience. 9:149.

Freebern, E., D. J. A. Santos, L. Fang, J. Jiang, K. L. Parker Gaddis, G. E. Liu, P. M. VanRaden, C. Maltecca, J. B. Cole, and L. Ma. 2020. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. BMC Genom. 21:41.

Gangur, V., N. P. Birmingham, and S. Thanesvorakul. 2002. Chemokines in health and disease. Vet. Immunol. Immunopathol. 86:127–136.

Gao, J., H. W. Barkema, L. Zhang, G. Liu, Z. Deng, L. Cai, R. Shan, S. Zhang, J. Zou, J. P. Kastelic, and B. Han. 2017. Incidence of clinical mastitis and distribution of pathogens on large Chinese dairy farms. J. Dairy Sci. 100:4797-4806.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55.

Ghahramani, N., J. Shodja, S. A. Rafat, B. Panahi, and K. Hasanpur. 2021. Integrative systems biology analysis elucidates mastitis disease underlying functional modules in dairy cattle. Front. Genet. 12:712306.

Gianola, D. and D. Sorensen. 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. Genetics 167:1407–1424.

Gilbert, F. B., P. Cunha, K. Jensen, E. J. Glass, G. Foucras, C. Robert-Granie, R. Rupp, and P. Rainard. 2013. Differential response of bovine mammary epithelial cells to *Staphylococcus aureus* or *Escherichia coli* agonists of the innate immune system. Vet. Res. 44:40.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml User Guide Release 1.0. VSN International Ltd., Hemel Hempstead, UK.

Gilmour, A. R., B. J. Gogel, B. R.Cullis, and R. Thompson. 2009. ASReml User Guide. Release 3.0. VSN International Ltd., Hemel Hempstead, UK. Accessed February 2, 2022. https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/2018/02/ASReml-3-User-Guide.pdf

Gilmour, A. R., B. R. Cullis, S. J. Welhan, and R. Thompson. 2000. ASREML manual. New South Wales Dept. Agric., Orange, Australia.

Gorji, A. E., Z. Roudbari, B. Sadeghi, A. Javadmanesh, and T. Sadkowski. 2019. Transcriptomic analysis on the promoter regions discover gene networks involving mastitis in cattle. Microb. Pathog. 137:103801.

Grossi, D. A., M. K. Abo-Ismail, A. Koeck, S. P. Miller, P. Stothard, G. Plastow, F. Miglior, S. S. Moore, M. Sargolzaei, and F. Schenkel. 2014. Genome-wide association analyses for mastitis in Canadians Holsteins. Proc. 10th World Congr. Genet. Applied Livest. Prod.. Wageningen.

Guney, E., J. Garcia-Garcia, and B. Oliva. 2014. Systems biology GUILDify: a web server for phenotypic characterization of genes through biological data integration and network-based prioritization algorithms. Bioinformatics. 30:1789–1790.

Günther, J., S. Liu, K. Esch, H. J. Schuberth, and H. M. Seyfert. 2010. Stimulated expression of TNF-[alpha] and IL-8, but not of lingual antimicrobial peptide reflects the concentration of pathogens contacting bovine mammary epithelial cells. Vet. Immunol. Immunopathol. 135:152-157.

Günther, J., D. Koczan, W. Yang, G. Nürnberg, D. Repsilber, H. J. Schuberth, Z. Park, N. Maqbool, A. Molenaar, H. M. Seyfert. 2009. Assessment of the immune capacity of mammary epithelial cells: comparison with mammary tissue after challenge with *Escherichia coli*. Vet. Res. 40:31-39.

Guo, J., H. Jorjani, and Ö. Carlborg. 2012. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. BMC Genet. 13:82.

Guo, P., B. Zhu, H. Niu, Z. Wang, Y. Liang, Y. Chen, L. Zhang, H. Ni, Y. Guo, E. H. A. Hay, X. Gao, H. Gao, X. Wu, L. Xu, and J. Li. 2018. Fast genomic prediction of breeding values using parallel Markov chain Monte Carlo with convergence diagnosis. BMC Bioinformatics 19:3.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12:12.

Halasa, T. 2007. Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29:18-31.

Hu, H., Y. Yang, X. Lan, Q. Zhang, and C. Pan. 2020. Relationships between novel nucleotide variants within the colony-stimulating factor 1 receptor (CSF1R) gene and mastitis indicators in sheep. Anim. Biotechnol.:1-8. @@this reference has errors. Please correct

Hughes, C. E. and R. J. B. Nibbs. 2018. A guide to chemokines and their receptors. FEBS J. 285:2944–2971.

Huma, Z. I., N. Sharma, S. Kour, S. Tandon, P. K. Guttula, S. Kour, A. K. Singh, R. Singh and M. K. Gupta. 2020. Putative biomarkers for early detection of mastitis in cattle. Anim. Prod. Sci. 60:1721-1736.

Iung, L. H. S., J. Petrini, J. Ramírez-Díaz, M. Salvian, G. A. Rovadoscki, F. Pilonetto, B. D. Dauria, P. F. Machado, L. L. Coutinho, G. R. Wiggans, and G. B. Mourão. 2019. Genome-wide association study for milk production traits in a Brazilian Holstein population. J. Dairy Sci. 102:5305–5314.

Jensen, K. J., J. Günther, R. Talbot, W. Petzl, H. Zerbe, H. J. Schuberth, H. S. Seyfert, and E. J. Glass. 2013. *Escherichia coli* and *Staphylococcus aureus* induced mastitis differentially modulate transcriptional responses in neighbouring uninfected bovine mammary gland quarters. BMC Genom. 14:1-19.

Jiang, J. C., L. Ma, D. Prakapenka, P. M. VanRaden, J. B. Cole, Y. Da. 2019. A large-scale genome-wide association study in US Holstein cattle. Front. Genet. 10:21.

Jiang, L., J. F. Liu, D. X. Sun, P. P. Ma, X. D. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5:12.

Johansson, K., S. Eriksson, J. Pösö, U. S. Nielsen, and G. P. Aamand. 2007. Predictive ability of different models for clinical mastitis in joint genetic evaluation for Sweden, Denmark and Finland. Proc. 58th EAAP, Dublin, Ireland.

Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S. Kong, N. B. Freimer, C. Sabatti, and E. Eskin. 2010. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42:348–354.

Kirsanova, E., B. Heringstad, A. Lewandowska-Sabat, and I. Olsaker. 2019. Alternative subclinical mastitis traits for genetic evaluation in dairy cattle. J. Dairy Sci. 102:5323–5329.

Kirsanova, E., B. Heringstad, A. Lewandowska-Sabat, and I. Olsaker. 2020. Identification of candidate genes affecting chronic subclinical mastitis in Norwegian Red cattle: Combining genome-wide association study, topologically associated domains and pathway enrichment analysis. Anim. Genet. 51:22–31.

Kirsanova, E., P. Boysen, G. M. Johansen, B. Heringstad, A. Lewandowska-Sabat, and I. Olsaker. 2020. Expression analysis of candidate genes for chronic subclinical mastitis in Norwegian Red cattle. J. Dairy Sci. 103:9142-9149.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012. Health recording in Canadian Holsteins: Data and genetic parameters. J. Dairy Sci. 95:4099–4108.

Kurz, J. P., Z. Yang, R. B. Weiss, D. J. Wilson, K. A. Rood, G. E. Liu, and Z. D. Wang. 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.

Levison, L. J., E. K. Miller-Cushon, A. L. Tucker, R. Bergeron, K. E. Leslie, H. W. Barkema, and T. J. DeVries. 2016. Incidence rate of pathogen-specific clinical mastitis on conventional and organic Canadian dairy farms. J. Dairy Sci. 99:1341-1350.

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

Liao, Y., J. Wang, E. J. Jaehnig, Z. Shi, and B. Zhang. 2019. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 47:199–205.

Li, T., C. Lin, Y. Zhu, H. Xu, Y. Yin, C. Wang, X. Tang, T. Song, A. Guo, Y. Chen and C. Hu 2021. Transcriptome profiling of m6A mRNA modification in bovine mammary epithelial cells treated with *Escherichia coli*. Int. J. Mol. Sci. 22:6254.

Liu, X., M. Huang, B. Fan, E. S. Buckler, and Z. Zhang. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 12:e1005767.

Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. J. Dairy Sci. 87:1896–1907.

Lu, H., Y. Wang, and H. Bovenhuis. 2020. Genome-wide association study for genotype by lactation stage interaction of milk production traits in dairy cattle. J. Dairy Sci. 103:5234–5245.

Ma, L., H. B. Runesha, D. Dvorkin, J. R. Garbe, and Y. Da. 2008. Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies. BMC Bioinformatics 9:315.

Ma, P. P., M. S. Lund, G. P. Aamand, and G. S. Su. 2019. Use of a Bayesian model including QTL markers increases prediction reliability when tests animals are distant from the reference population. J. Dairy Sci. 102:7237–7247.

Macciotta, N. R. P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. J. Dairy Sci. 98:8175–8185.

Madsen, P. and Jensen, J. 2010. DMU: A user's guide. A package for analysing multivariate mixed models. Version 6, release 5.0. University of Aarhus, Tjele, Denmark.

Madsen, P. and J. Jensen. 2012. DMU - A user's guide. A package for analysing multivariate mixed models. Version 6, release 5.1. University of Aarhus, Tjele, Denmark. Accessed February 2, 2022. https://dmu.ghpc.au.dk/dmu/DMU/Doc/Previous/dmuv6_guide.5.1.pdf

Madsen, P. and Jensen, J. 2013. A user's guide to DMU. A package for analysing multivariate mixed models. Version 6, Release 5.2. University of Aarhus, Tjele, Denmark. Accessed February 2, 2022. https://dmu.ghpc.au.dk/dmu/DMU/Doc/Current/dmuv6_guide.5.2.pdf

Marete, A., M. S. Lund, D. Boichard, and Y. Ramayo-Caldas. 2018a. A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. PLoS One 13:e0199931.

Marete, A., G. Sahana, S. Fritz, R. Lefebvre, A. Barbat, M. S. Lund, B. Guldbrandtsen, and D. Boichard. 2018b. Genome-wide association study for milking speed in French Holstein cows. J. Dairy Sci. 101:6205–6219.

Matsushima, K. 2000. Chemokines: Introduction. Springer Semin. Immunopathol. 22:321–328.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek and F. Cunningham.2016. The Ensembl Variant Effect Predictor. Genome Biol. 17:122.

Medrano, J. F. 2017. The new bovine reference assembly and its value for genomic research. Proc. Assoc. Advmt. Anim. Breed. Genet. 22:161–166.

Meier, S., D. Arends, P. Korkuc, G. B. Neumann, and G. A. Brockmann. 2020. A genome-wide association study for clinical mastitis in the dual-purpose German Black Pied cattle breed. J. Dairy Sci. 103:10289–10298.

Mele, M., N. P. P. Macciotta, A. Cecchinato, G. Conte, S. Schiavon, and G. Bittante. 2016. Multivariate factor analysis of detailed milk fatty acid profile: Effects of dairy system, feeding, herd, parity, and stage of lactation. J. Dairy Sci. 99:9820–9833.

Meredith, B. K., D. P. Berry, F. Kearney, E. K. Finlay, A. G. Fahey, D. G. Bradley, and D. J. Lynn. 2013. A genome-wide association study for somatic cell score using the Illumina high-density bovine beadchip identifies several novel QTL potentially related to mastitis susceptibility. Front. Genet. 4:229.

Meredith, B. K., F. J. Kearney, E. K. Finlay, D. G. Bradley, A. G. Fahey, D. P. Berry, and D. J. Lynn. 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet. 13:21.

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Meyer, K. 2007. WOMBAT: A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J. Zhejiang Univ. Sci. B 8:815–821.

Meyer, K. and B. Tier. 2012. “SNP Snappy”: A strategy for fast genome-wide association studies fitting a full mixed model. Genetics 190:275–277.

Miles, A. M. and H. J. Huson. 2020. Time-and population-dependent genetic patterns underlie bovine milk somatic cell count. J. Dairy Sci.103:8292-8304.

Misztal, I., S. E. Aggrey, and W. M. Muir. 2013. Experiences with a single-step genome evaluation. Poult. Sci. 92:2530–2534.

Momen, M., A. Ayatollahi Mehrgardi, M. Amiri Roudbar, A. Kranis, R. Mercuri Pinto, B. D. Valente, G. Morota, G. J. M. Rosa, and D. Gianola. 2018. Including phenotypic causal networks in genome-wide association studies using mixed effects structural equation models. Front. Genet. 9:455.

Momen, M., M. T. Campbell, H. Walia, and G. Morota. 2019. Harnessing phenotypic networks and structural equation models to improve genome-wide association analysis. Plant. Methods. 15:107.

Mulder, H. A., R. E. Crump, M. P. L. Calus, and R. F. Veerkamp. 2013. Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms. J. Dairy Sci. 96:7306–7317.

Naderi, S., M. Bohlouli, T. Yin, and S. Konig. 2018. Genomic breeding values, SNP effects and gene identification for disease traits in cow training sets. Anim. Genet. 49:178–192.

Nani, J. P., M. A. Raschia, M. A. Poli, L. F. Calvinho, and A. F. Amadio. 2015. Genome-wide association study for somatic cell score in Argentinean dairy cattle. Livest. Sci. 175:1–9.

Narayana, S. G., E. de Jong, F. S. Schenkel, P. A. S. Fonseca, T. C. S. Chud, D. Powel, G. Wachoski-Dark, P. E. Ronksley, F. Miglior, K. Orsel, and H. W. Barkema. 2022. Supplementary materials: Underlying genetic architecture of resistance to mastitis in dairy cattle: A systematic review and gene prioritization analysis of genome-wide association studies. Harvard Dataverse, V1. https://doi.org/10.7910/DVN/HNKBJS

Narayana, S. G., F. Miglior, S. A. Naqvi, F. Malchiodi, P. Martin, and H. W. Barkema. 2018. Genetic analysis of subclinical mastitis in early lactation of heifers using both linear and threshold models. J. Dairy Sci. 101:11120–11131.

Narayana, S. G., F. Schenkel, F. Miglior, T. Chud, E. A. Abdalla, S. A. Naqvi, F. Malchiodi, and H. W. Barkema. 2021. Genetic analysis of pathogen-specific intramammary infection in dairy cows. J. Dairy Sci. 104:1982-1992.

O'Connell, J.R. 2015. MMAP User Guide. Accessed February 2, 2022. http://edn.som.umaryland.edu/ mmap/index.php.

Oliveira, H. R., J. P. Cant, L. F. Brito, F. L. B. Feitosa, T. C. S. Chud, P. A. S. Fonseca, J. Jamrozik, F. F. Silva, D. A. L. Lourenco, and F. S. Schenkel. 2019a. Genome-wide association for milk production traits and somatic cell score in different lactation stages of Ayrshire, Holstein, and Jersey dairy cattle. J. Dairy Sci. 102:8159–8174.

Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, J. P. Cant, and F. S. Schenkel. 2019b. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. J. Dairy Sci. 102:9995–10011.

Oliver, S. P., R. A. Almeida, B. E. Gillespie, S. J. Ivey, H. Moorehead, P. Lunn, H. H. Dowlen, D. L. Johnson, and K. C. Lamar. 2003. Efficacy of extended pirlimycin therapy for treatment of experimentally induced *Streptococcus uberis* intramammary infections in lactating dairy cattle. Vet. Ther. 4299-4308.

Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 372:n71.

Pegolo, S., M. Momen, G. Morota, G. J. M. Rosa, D. Gianola, G. Bittante, and A. Cecchinato. 2020. Structural equation modeling for investigating multi-trait genetic architecture of udder health in dairy cattle. Sci. Rep.10:7751.

Piepers, S., L. De Meulemeester, A. de Kruif, G. Opsomer, H. W. Barkema, and S. De Vliegher. 2007. Prevalence and distribution of mastitis pathogens in subclinically infected dairy cows in Flanders, Belgium. J. Dairy Res. 74:478-483.

Poeta, V. M., M. Massara, A. Capucetti, and R. Bonecchi. 2019. Chemokines and chemokine receptors: New targets for cancer immunotherapy. Front. Immunol. 10:1–10.

Prinsen, R., A. Rossoni, B. Gredler, A. Bieber, A. Bagnato, and M. G. Strillacci. 2017. A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. Livest. Sci. 202:7–12.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

Pyörälä, S. 2002. New strategies to prevent mastitis. Reprod. Domest. Anim. 37:211-216.

Qanbari, S., H. Pausch, S. Jansen, M. Somel, T. M. Strom, R. Fries, R. Nielsen, and H. Simianer. 2014. Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet. 10:e1004148.

R core Team. 2020. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Accessed February 2, 2022. http://www.R-project.org/.

Rainard, P. and C. Riollet. 2006. Innate immunity of the bovine mammary gland. Vet. Res. 37:369–400.

Raman, D., T. Sobolik-Delmaire1, and A. Richmond. 2014. Chemokines in health and disease. Exp. Cell Res. 10:575–589.

Rincón Flórez, J. C., A. López Herrera, and J. J. Echeverri Zuluaga. 2018. Genome-wide association study using the Bayes C method for important traits in dairy yield in Colombian Holstein cattle. Acta Sci. Anim. Sci. 40:39015–39015.

Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDaneld, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M Phillippy, J. B. Cole, C. P. V. Tassell, G. Liu, T. P. L. Smith, and J. F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience 9: giaa021.

Rupp, R. and D. Boichard. 2003. Genetics of resistance to mastitis in dairy cattle. Vet. Res. 34:671-688.

Sahana, G., B. Guldbrandtsen, B. Thomsen, and M. S. Lund. 2013. Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. Anim. Genet. 44:620–626.

Sahana, G., B. Guldbrandtsen, B. Thomsen, L. E. Holm, F. Panitz, R. F. Brondum, C. Bendixen, and M. S. Lund. 2014. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. J. Dairy Sci. 97:7258–7275.

Sajjanar, B., N. Trakooljul, K. Wimmers, and S. Ponsuksili. 2019. DNA methylation analysis of porcine mammary epithelial cells reveals differentially methylated loci associated with immune response against *Escherichia coli* challenge. BMC Genom. 20:1-15.

Santman-Berends, I.M.G.A., T.J.G.M. Lam, J. Keurentjes, and G. van Schaik. 2015. An estimation of the clinical mastitis incidence per 100 cows per year based on routinely collected herd data. J. Dairy Sci. 98:6965-6977.

Sayers, E. W., E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, and S. T. Sherry. Database resources of the national center for biotechnology information. Nucleic Acids Res. 49:D10-D17.

Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and B. J. Van Doormaal. 2000. Experience with a test-day model. J. Dairy Sci. 83:1135–1144.

Schukken, Y. H., J. Günther, J. Fitzpatrick, Michael C. Fontaine, L. Goetze, O. Holst, J. Leigh, W. Petzl, H.-J. Schuberth, A. Sipka, D. G. E. Smith, R. Quesnell, J. Watts, R. Yancey, H. Zerbe, A. Gurjar, R. N. Zadoks, and H.-M. Seyfert. 2011. Host-response patterns of intramammary infections in dairy cows. Vet. Immunol. Immunopathol. 144:270-289.

Schutyser, E., S. Struyf, and J. Van Damme. 2003. The CC chemokine CCL20 and its receptor CCR6. Cytokine Growth Factor Rev. 14:409-426.

Segura, V., B. J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, and M. Nordborg. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44:825–830.

Seroussi, E., S. E. Blum, O. Krifucks, A. Shirak, S. Jacoby, and G. Leitner. 2020. Basal levels of CD18 antigen presenting cells in cow milk associate with copy number variation of Fc gamma receptors. Genes (Basel) 11:952.

Shannon P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B Schwikowski, and T. Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res.13:2498–2504

Siebert, L., M. E. Staton, S. Headrick, M. Lewis, B. Gillespie, C. Young, R. A. Almeida, S. P. Oliver, and G. M. Pighetti. 2018. Genome-wide association study identifies loci associated with milk leukocyte phenotypes following experimental challenge with *Streptococcus uberis*. Immunogenetics 70:553–562.

Silva, A. A., D. A. Silva, F. F. Silva, C. N. Costa, H. T. Silva, P. S. Lopes, R. Veroneze, G. Thompson, and J. Carvalheira. 2020. GWAS and gene networks for milk-related traits from test-day multiple lactations in Portuguese Holstein cattle. J. Appl. Genet. 61:465-476.

Sodeland, M., M. P. Kent, H. G. Olsen, M. A. Opsal, M. Svendsen, E. Sehested, B. J. Hayes, and S. Lien. 2011. Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. Anim. Genet. 42:457–465.

Stelzer, G., N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, No. Rappaport, M. Safran, and D. Lancet. 2016. The GeneCards Suite: From gene data mining to disease genome sequence analysis. Curr. Prot. Bioinformatics. 54:1.30.1 - 1.30.33.

Strillacci, M. G., E. Frigo, F. Canavesi, Y. Ungar, F. Schiavini, L. Zaniboni, L. Reghenzani, M. C. Cozzi, A. B. Samoré, Y. Kashi, E. Shimoni, R. Tal-Stein, M. Soller, E. Lipkin, and A. Bagnato. 2014. Quantitative trait loci mapping for conjugated linoleic acid, vaccenic acid and Δ(9) -

desaturase in Italian Brown Swiss dairy cattle using selective DNA pooling. Anim. Genet. 45:485–499.

Strillacci, M. G., E. Frigo, F. Schiavini, A. B. Samore, F. Canavesi, M. Vevey, M. C. Cozzi, M. Soller, E. Lipkin, and A. Bagnato. 2014. Genome-wide association study for somatic cell score in Valdostana Red Pied cattle breed using pooled DNA. BMC Genet. 15:106.

Strucken, E. M., R. H. Bortfeldt, D. J. de Koning, and G. A. Brockmann. 2012. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375–382.

Su, G., O. F. Christensen, L. Janss, and M. S. Lund. 2014. Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. J. Dairy Sci. 97:6547–6559.

Suzuki, Y., S. Haga, D. Katoh, K. So, K. Choi, U. Jung, H. Lee, K. Katoh, and S. Roh. 2015. Chemerin is a novel regulator of lactogenesis in bovine mammary epithelial cells. Biochemical and biophysical research communications. Biochem. Biophys. Res. Comm. 466:283-288.

Svishcheva, G. R., T. I. Axenovich, N. M. Belonogova, C. M. van Duijn, and Y. S. Aulchenko. 2012. Rapid variance components-based method for whole-genome association analysis. Nat. Genet. 44:1166–1170.

Szklarczyk D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, and K. P. Tsafou. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res:43(Database issue):D447–D452.

Tao, W. and B. Mallard. 2007. Differentially expressed genes associated with *Staphylococcus aureus* mastitis of Canadian Holstein cows. Vet. Immunol. Immunopathol. 120:201-211.

Tiezzi, F., K. L. Parker-Gaddis, J. B. Cole, J. S. Clay, and C. Maltecca. 2015. A Genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PLoS One 10:0114919.

Tribout, T., P. Croiseau, R. Lefebvre, A. Barbat, M. Boussaha, S. Fritz, D. Boichard, C. Hoze, and M. P. Sanchez. 2020. Confirmed effects of candidate variants for milk production, udder health, and udder morphology in dairy cattle. Genet. Sel. Evol. 52:55.

Tsuruta, S. and Misztal, I. 2006. THRGIBBS1F90 for estimation of variance components with threshold linear models. Pages 27-31 in Proc. 8th World Congr. Genet. Appl. Livest. Prod. Belo Horizonte, Brazil.

VanRaden, P. M. and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737–2746.

VanTol, B.L., S. Missan, J. Crack, S. Moser, W. H. Baldridge, P. Linsdell and E. A. Cowley. 2007. Contribution of KCNQ1 to the regulatory volume decrease in the human mammary epithelial cell line MCF-7. Am. J. Physiol. Cell. Physiol. 293:C1010-C1019.

Veerkamp, R. F., A. C. Bouwman, C. Schrooten, and M. P. L. Calus. 2016. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. Genet. Sel. Evol. 48:95.

Verbyla, K. L., M. P. L. Calus, H. A. Mulder, Y. de Haas, and R. F. Veerkamp. 2010. Predicting energy balance for dairy cows using high-density single nucleotide polymorphism information. J. Dairy Sci. 93:2757–2764.

Viitala, S., J. Szyda, S. Blott, N. Schulman, M. Lidauer, A. Mäki-Tanila, M. Georges, and J. Vilkki. 2006. The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. Genetics 173:2151-2164.

Vrieling, M., E. M. Boerhout, G. F. Van Wigcheren, K. J. Koymans, T. G. Mols-Vorstermans, C. J. C De Haas, P. C. Aerts, I.J.J.M. Daemen, K. P. M. van Kessel, A. P. Koets, V.P.M.G. Rutten, P. J. M. Nuijten, J. A. G. van Strijp, and L. Benedictus. 2016. LukMF′ is the major secreted leukocidin of bovine *Staphylococcus aureus* and is produced in vivo during bovine mastitis. Sci. Rep. 6:37759.

Walker, M. G., W. Volkmuth, E. Sprinzak, D. Hodgson, and T. Klingler. 1999. Prediction of gene function by genome-scale expression analysis: Prostate cancer-associated genes. Genome Res. 9:1198–1203.

Wang, D., L. Liu, S. M. A. Augustino, T. Duan, T. J. Hall, D. E. MacHugh, J. Dou, Y. Zhang, Y. Wang, and Y. Yu. 2020. Identification of novel molecular markers of mastitis caused by *Staphylococcus aureus* using gene expression profiling in two consecutive generations of Chinese Holstein dairy cattle. J. Anim. Sci. Biotechnol. 11:1-17.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. 94:73–83.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W. M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front. Genet. 5:134.

Wang, X., P. P. Ma, J. F. Liu, Q. Zhang, Y. Zhang, X. D. Ding, L. Jiang, Y. C. Wang, Y. Zhang, D. X. Sun, S. L. Zhang, G. S. Su, and Y. Yu. 2015. Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. BMC Genet. 16:9.

Wang, X., Y. Fan, Y. He, Z. Han, Z. Gong, Y. Peng, Y. Meng, Y. Mao, Z. Yang, and Y. Yang. 2021. Integrative analysis of miRNA and mRNA expression profiles in mammary glands of Holstein cows artificially infected with *Staphylococcus aureus*. Pathogens 10:506.

Wathes, D. C., F. Becker, L. Buggiotti, M. A. Crowe, C. Ferris, L. Foldager, C. Grelet, M. Hostens, K. L. Ingvartsen, C. Marchitelli, and M. Salavati. 2021. Associations between circulating IGF-1 concentrations, disease status and the leukocyte transcriptome in early lactation dairy cows. Ruminants 1:147-177.

Weeks, N. T., G. R. Luecke, B. M. Groth, M. Kraeva, L. Ma, L. M. Kramer, J. E. Koltes, and J. M. Reecy. 2016. High-performance epistasis detection in quantitative trait GWAS. Int. J. High Perform. Comput. Appl. 32:321–336.

Weirauch, M. T. 2011. Gene coexpression networks for the analysis of DNA microarray data. Appl. Stat. Netw. Biol. Meth. Systems Biol. 1:215-250.

Weitzman, J. B. 2000. Comparing cows with humans. Genome Biol. 1, spotlight-20001005-04.

Welderufael, B. G., P. Løvendahl, D.J. de Koning, L. L. G. Janss, and W. F. Fikse. 2018. Genome-wide association study for susceptibility to and recoverability from mastitis in Danish Holstein cows. Front. Genet. 9:141.

Wijga, S., J. W. M. Bastiaansen, E. Wall, E. Strandberg, Y. de Haas, L. Giblin, and H. Bovenhuis. 2012. Genomic associations with somatic cell score in first-lactation Holstein cows. J. Dairy Sci. 95:899–908.

Wilmink, J. B. M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. Livest. Prod. Sci. 16:335–348.

Yang, F., F. H. Chen, L. L. Li, L. Yan, T. Badri, C. L. Lv, D. L. Yu, M. L. Zhang, X. J. Jang, J. Li, L. Yuan, G. L. Wang, H. Li, J. Li, and Y. F. Cai. 2019. Three novel players: PTK2B, SYK, and TNFRSF21 were identified to be involved in the regulation of bovine mastitis susceptibility via GWAS and post-transcriptional analysis. Front. Immunol. 10:20.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88:76–82.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208.

Zhang, J., L. Zan, P. Fang, F. Zhang, G. Shen, and W. Tian. 2008. Genetic variation of PRLR gene and association with milk performance traits in dairy cattle. Can. J. Anim. Sci. 88:33-39.

Zheng, J., A. D. Watson, and D. E. Kerr. 2006. Genome-wide expression analysis of lipopolysaccharide-induced mastitis in a mouse model. Infect. Immun. 74:1907–1915.

Zhou, J., L. Liu, C. J. Chen, M. Zhang, X. Lu, Z. Zhang, X. Huang, and Y. Shi. 2019. Genome-wide association study of milk and reproductive traits in dual-purpose Xinjiang Brown cattle. BMC Genomics 20:827.

Zhou, X. and M. Stephens. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat. Methods 11:407–409.

Zhou, Y., E. E. Connor, G. R. Wiggans, Y. Lu, R. J. Tempelman, S. G. Schroeder, H. Chen, and G. E. Liu. 2018. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. BMC Genom. 19:314.

Table 4-1. Origin of data, breed and phenotype for 52 articles used for the analyses

Manuscript	Country	Breed	Phenotype
Abdel-Shafy et al., 2014	Germany	Holstein	DYD of SCS, calculated across first 3 lactations (Lui et al., 2004)
Cai et al., 2018	Nordic countries	Holstein	DEBV of CM resistance, calculated as a weighted score over first 3 lactations
Cai et al., 2019	Nordic countries	Holstein	DEBV of udder health (combination of CM occurrence, SCC and udder conformation), calculated as a weighted score over first 3 lactations
Cecchinato et al., 2019	Italy	Brown Swiss	Udder health, a combination of lactose content and SCS (Mele et al., 2016)
Cole et al., 2011	United States	Holstein	PTA of SCS
Duran et al., 2017	Mexico	Holstein	EBV of SCS, from phenotypic extreme animals only (2 SD above and below mean)
Fang et al., 2017a	Nordic countries	Holstein, Jersey, Norwegian Red	DEBV of CM resistance, calculated as a weighted score over first 3 lactations
Fang et al., 2017b	Nordic countries	Holstein, Jersey	DEBV of CM resistance, calculated as a weighted score over first 3 lactations

Fang et al., 2018	Nordic countries	Holstein, Jersey, Norwegian Red	DEBV of CM resistance, calculated as a weighted score over first 3 lactations
Fang et al., 2019	Austria, Germany, Switzerland	Brown Swiss	EBV of SCS
Freebern et al., 2020	United States	Holstein	DPTA of CM, calculated according to Garrick et al. (2009)
Guo et al., 2012	Austria, Switzerland, Germany, France, Italy, Slovenia, United States	Brown Swiss	DEBV of SCS
Iung et al., 2019	Brazil	Holstein	SCS of cows ranging from 1st to 6th parity
Jiang et al., 2019	United States	Holstein	SCS of first lactation cows
Kirsanova et al., 2020	Norway	Norwegian Red	DYDs of lactation average SCS and 9 SCM traits as defined by Kirsanova et al. (2019), calculated across first 3 lactations
Kurz et al., 2019	United States	Holstein	Mastitis resistant (absence of CM and SCM) and susceptible (at least 4 CM cases), cows were monitored for 8 months
Lu et al., 2020	Netherlands	Holstein	SCS of first lactation cows
Ma et al., 2019	Germany, France, the Netherlands, Nordic	Holstein	DEBV of mastitis

Macciotta et al., 2015	Italy	Simmental	SCS, indicated by 2 variables per lactation (via principal component analysis) to reflect lactation pattern
Marete et al., 2018a	France	Holstein (SCS only), Normande, Montbeliarde	YDs of SCS and CM, calculated over the first 3 lactations
Marete et al., 2018b	France	Holstein	YDs of SCS and CM, calculated over the first 3 lactations
Meier et al., 2020	Germany	German Black Pied	CM present or absent, during first 3 lactations
Meredith et al., 2012	Ireland	Holstein	DYD and YDs of SCS, calculated across first 5 lactations (Berry et al., 2007)
Meredith et al., 2013	Ireland	Holstein	DYD of SCS, calculated across first 5 lactations (Berry et al., 2007)
Miles et al., 2020	United States	Holstein	Udder inflammation status (average, healthy, chronic), determined using SCS and CMT results
Mulder et al., 2013	Ireland, Scotland, Netherlands, Sweden	Holstein	EBV of SCS of first-lactation heifers
Naderi et al., 2018	Germany	Holstein	CM present or absent, during first lactation

Nani et al., 2015	Argentina	Holstein and Holstein x Jersey	SCS for 1 lactation, indicated by 3 variables (arithmetic mean, maximum value, arithmetic mean of the top 3 values)
Oliveira et al., 2019a	Canada	Ayrshire, Holstein, Jersey	EBV of SCS, calculated across first 3 lactations (Schaeffer et al., 2000).
Oliveira et al., 2019b	Canada	Ayrshire, Holstein, Jersey	EBV of SCS, calculated across first 3 lactations (Schaeffer et al., 2000).
Pegolo et al., 2020	Italy	Brown Swiss	SCS, ranging from 1st to 5th parity
Prinsen et al., 2017	Italy	Brown Swiss	DEBV of SCS
Qanbari et al., 2014	Germany	Fleckvieh	EBV of SCS
Rincon Florez et al., 2018	Colombia	Holstein	EBV of SCS
Sahana et al., 2013	Nordic countries	Holstein	EBVs of CM resistance and SCS, each calculated as a weighted score over first 3 lactations (Johansson et al., 2007)
Sahana et al., 2014	Nordic countries	Holstein, Jersey (CM only), Norwegian Red (CM only)	EBVs of CM resistance and SCS, each calculated as a weighted score over first 3 lactations (Johansson et al., 2007)

Seroussi et al., 2020	Israel	Holstein	Udder inflammation status (healthy, subclinical, clinical) determined using SCC, CMT and bacterial testing, as well as differential SCC, based on flow cytometry of CD antigens in milk. First-lactation heifers only
Siebert et al., 2018	United States	Holstein	Time to SCC 'cure' post *S. uberis* challenge, SCC 0-7 days post challenge and SCC 0-28 days post challenge
Silva et al., 2020	Portugal	Holstein	EBV of SCS, calculated across first 3 lactations
Sodeland et al., 2011	Norway	Norwegian Red	DYDs of lactation average SCS and 7 CM traits, calculated across first 3 lactations
Strillacci et al., 2014	Italy	Valdostana Red Pied	EBV of SCS of first lactation cows, from phenotypic extreme animals only (top and bottom 20%)
Strucken et al., 2012	Germany	Holstein	Lactation curve parameter (Wilmink curve; Wilmink 1987) and YD of SCS, calculated over the first 4 lactations
Su et al., 2014	Nordic countries	Holstein	DEBVs of mastitis
Tiezzi et al., 2015	United States	Holstein	PTA of CM incidence rate, first lactation only

Tribout et al., 2020	France	Holstein, Montbéliarde, Normande	DYDs and YDs of SCS and CM, calculated across lactations
Veerkamp et al., 2016	Netherlands	Holstein	DEBV of SCS
Wang et al., 2015	China	Holstein	EBV of SCS
Welderufael et al., 2018	Denmark	Holstein	Two health events: transition from healthy to clinical mastitis and vice versa, during first 3 lactations
Wijga et al., 2012	Ireland, Scotland, Netherlands, Sweden	Holstein	SCS of first-lactation heifers
Yang et al., 2019	China	Holstein	Mastitis resistant (no mastitis across 3 lactations and low SCC) and susceptible (CM in each of 3 lactations)
Zhou et al., 2018	United States	Holstein	DPTA of SCS, calculated according to VanRaden and Wiggans (1991)
Zhou et al., 2019	China	Xinjiang Brown	EBV of SCS

[1] Nordic = Denmark, Sweden, Finland

[2] SCS = somatic cell score, SCC = somatic cell count, CM = clinical mastitis, DYD = daughter-yield deviations, YD = yield deviations, EBV = estimated breeding values, DEBV = deregressed EBVs, PTA = predicted transmission ability, and DPTA = de-regressed predicted transmitting ability.

Table 4-2. Genotyping and quality control summary of 52 included manuscripts

Manuscript	SNP-chip	Assembly	Call rate (SNPs)	MAF	Hardy-Weinberg threshold
Abdel-Shafy et al., 2014	Illumina BovineSNP50 BeadChip v1	UMD 3.1	< 0.90	< 0.01	$P < 0.001$
Cai et al., 2018	Illumina BovineSNP50 BeadChip (54K) v1 or v2	UMD 3.1.1	-	< 0.01	$P < 10^{-6}$
Cai et al., 2019	Illumina BovineSNP50 BeadChip (54K) v1 or v2	UMD 3.1.1	-	< 0.01	$P < 10^{-6}$
Cecchinato et al., 2019	Illumina BovineSNP50 BeadChip v2	UMD 3.1	< 0.95	< 0.01	$P \leq 0.001$
Cole et al., 2011	Illumina BovineSNP50 BeadChip[1]	UMD 3.0	-	≤ 0.02	-
Duran et al., 2017	Illumina BovineHD BeadChip	UMD 3.1	< 0.98	0.02	-
Fang et al., 2017a	Illumina HiSeq 2000 system	UMD 3.1	-	≤ 0.01	$P \leq 10^{-6}$
Fang et al., 2017b	Illumina BovineHD BeadChip, Illumina BovineSNP50 BeadChip[1]	UMD 3.1	-	< 0.01	$P < 10^{-6}$
Fang et al., 2018	Illumina BovineHD BeadChip, Illumina Bovine BeadChip (50K) [1]	UMD 3.1	-	≤ 0.01	$P \leq 10^{-6}$
Fang et al., 2019	Illumina BovineHD BeadChip, Illumina BovineSNP50 BeadChip v1 and v2	UMD 3.1	< 0.80	< 0.01	$P < 10^{-5}$
Freebern et al., 2020	BovineHD BeadChip (777K)	UMD 3.1	-	< 0.01	$P \leq 10^{-6}$
Guo et al., 2012	Illumina BovineSNP50 BeadChip[1]	-	< 0.90	< 0.02	$P < 0.01$
Iung et al., 2019	Illumina BovineLD BeadChip, GeneSeek Genomic Profiler Bovine (50k)	-	< 0.90	< 0.02	-
Jiang et al., 2019	-	UMD 3.1	-	< 0.05	-
Kirsanova et al., 2020	Affymetrix SNP array (25K), Illumina BovineSNP50 BeadChip (54K) [1], Illumina BovineHD BeadChip (777K)	UMD 3.1.1	< 0.95 (minimal) and < 0.85 (individual)	< 0.01	-
Kurz et al., 2019	Illumina BovineHD BeadChip	UMD 3.1.1/bosTau8	< 0.95	< 0.05	-
Lu et al., 2020	Customized Illumina SNP chip (CRV, 50K) and Illumina Infinium assay	Btau 4.0	< 0.80	-	-
Ma et al., 2019	Illumina BovineSNP50 BeadChip (54K) [1]	-		≤ 0.005	$P \leq 10^{-7}$

Macciotta et al., 2015	Illumina BeadChip (7K)	UMD 3.1	≤ 0.99	≤ 0.01	-
Marete et al., 2018a	Illumina BovineSNP50 BeadChip (50K) [1], Illumina BovineLD BeadChip v.2	UMD 3.1	≤ 0.99	≤ 0.02	$P < 10^{-4}$
Marete et al., 2018b	Illumina BovineSNP50 Beadchip (50k) [1], customized EuroGenomics LD SNP chip	UMD 3.1	< 0.95	< 0.005	$P < 10^{-4}$
Meier et al., 2020	Illumina BovineSNP50 BeadChip v.3	UMD 3.1	< 0.90	< 0.01	-
Meredith et al., 2012	Illumina BovineSNP50 BeadChip[1]	Btau 4.0	< 0.95	≤ 0.05	-
Meredith et al., 2013	Illumina BovineHD BeadChip	UMD 3.1	≤ 0.90	< 0.05	$P < 10^{-6}$
Miles et al., 2020	Illumina BovineHD BeadChip (777K)	ARS-UCD 1.2	< 0.90	< 0.05	-
Mulder et al., 2013	Illumina BovineSNP50 BeadChip[1]	UMD 3.0	≤ 0.95	≤ 0.01	$x^2 < 600$; $P = 1.7 \times 10^{-132}$
Naderi et al., 2018	Illumina Bovine Eurogenomics SNP chip (10K), Illumina Bovine SNP BeadChip (50K) [1]	-	< 0.95	< 0.01	$P < 10^{-6}$
Nani et al., 2015	Illumina BovineSNP50 BeadChip v.2	UMD 3.1	≤ 0.95	< 0.05	$P < 0.001$
Oliveira et al., 2019a	Illumina BovineSNP50 BeadChip (50K) [1]	UMD 3.1	< 0.95	< 0.01	-
Oliveira et al., 2019b	Illumina BovineSNP50 BeadChip (50K) [1]	UMD 3.1	< 0.95	< 0.01	-
Pegolo et al., 2020	Illumina BovineSNP50 BeadChip v.2	UMD 3.1	< 0.95	< 0.01	$P < 0.001$
Prinsen et al., 2017	Illumina BovineHD BeadChip (777K)	UMD 3.1	-	-	-
Qanbari et al., 2014	Illumina Bovine SNP array (50K [1]and 700K)	NCBI 6.1	≤ 0.95	≤ 0.005	$P \leq 10^{-6}$
Rincon Florez et al., 2018	Illumina BovineLD BeadChip	UMD 3.1	< 0.001	< 0.03	$P \leq 0.05$
Sahana et al., 2013	Illumina BovineSNP50 BeadChip[1]	UMD 3.1	-	< 0.05	-
Sahana et al., 2014	Illumina BovineSNP50 BeadChip[1]	UMD 3.1	< 0.95	< 0.05	$P < 10^{-5}$
Seroussi et al., 2020	Illumina BovineSNP BeadChip[1]	ARS-UCD 1.2	-	< 0.05	$P < 10^{-4}$
Siebert et al., 2018	Illumina BovineSNP50 BeadChip[1]	UMD 3.1.1	-	0	-

Silva et al., 2020	Neogen GGP LD v.1, Illumina BovineSNP50 BeadChip v.1 and v.2, Illumina 57K, Illumina 77K, Neogen GGP HD v.3, Affymetrix Ax58K panel	-	< 0.95	< 0.02	$P < 10^{-6}$
Sodeland et al., 2011	Affymetrix 25K MIP array	-	≤ 0.75	≤ 0.025	-
Strillacci et al., 2014	Illumina BovineHD BeadChip (777K)	UMD 3.1 and Btau 4.6.1	-	-	-
Strucken et al., 2012	Illumina BovineSNP50 BeadChip[1]		< 0.95	< 0.0164	-
Su et al., 2014	Illumina BovineSNP50 BeadChip[1]	UMD 3.2	-	< 0.01	-
Tiezzi et al., 2015	Illumina BovineSNP50 BeadChip[1]	-	< 0.99	< 0.05	-
Tribout et al., 2020	Illumina Bovine SNP50 BeadChip (50K) [1], Illumina BovineHD Beadchip (777K)	UMD 3.1	≤ 0.90	< 0.01	$P \leq 10^{-4}$
Veerkamp et al., 2016	Illumina BovineHD BeadChip	-	-	-	-
Wang et al., 2015	Illumina BovineSNP50 BeadChip[1]	UMD 3.1	< 0.90	≤ 0.03	$P \leq 10^{-6}$
Welderufael et al., 2018	Illumina BovineSNP50 BeadChip v.2	UMD 3.1	< 0.90	< 0.01	-
Wijga et al., 2012	Illumina BovineSNP50 BeadChip[1]	Baylor 4.0/bosTau4	< 0.95	≤ 0.01	$X^2 < 600$
Yang et al., 2019	Illumina Hiseq Xten	UMD 3.1.1	-	-	-
Zhou et al., 2018	Illumina BovineHD BeadChip	UMD 3.1	≤ 0.90	≤ 0.05	$P \leq 10^{-6}$
Zhou et al., 2019	llumina Bovine BeadChip (150K)	UMD 3.1.1	≤ 0.90	-	-

[1]Did not mention the version of the Illumina BovineSNP50 BeadChip.

Table 4-3. Genome-wide association analysis summary of 52 full-text manuscripts

Manuscript	GWAS approach	# Animals used for GWAS	# SNPs	Multiple correction	# SNPs/Windows associated
Abdel-Shafy et al., 2014	Linear regression in PLINK (Purcell et al., 2007)	2,354 sires	37,424	Bonferroni	16 SNPs, 9 QTL windows
Cai et al., 2018	GCTA was used (Yang et al., 2011). Specific method not specified	5,147 sires	15,552,968	Bonferroni	22 SNPs
Cai et al., 2019	GCTA was used (Yang et al., 2011). Specific method not specified	4,921 sires	16,503,508	Bonferroni	-
Cecchinato et al., 2019	Mixed model using GenABEL R package (Aulchenko et al., 2007) followed by regression-genomic control approach (GRAMMAR-GC; Amin et al., 2007)	1,011 cows	37,568	Bonferroni	17 SNPs
Cole et al., 2011	Single-locus model	1,654 cows	45,878	Bonferroni	613 SNPs
Duran Aguilar et al., 2017	Two methods: linear regression in SVS 8.3.1 for CNVs detected by CNAM algorithm, and linear regression with CNVRULER for CNVs detected using PennCNV's Hidden Markov Model	220 cows for SVS analysis, 123 cows for PennCNV analysis	-	FDR	24 CNVR:SVS and 47 CNVR:PennCNV
Fang et al., 2017a	Variance based 2-step method with linear models and generalized linear regression models using EMMAX (Kang et al., 2010)	5,056 Holstein, 1,231 Jersey, 4,310 Nordic Red cattle	15,355,382 (Holstein), 13,403,916 (Jersey), 15,243,827 (Nordic Red)	Bonferroni	5 SNPs (Holstein), 0 SNPs (Jersey), 2 SNPs (Nordic Red)
Fang et al., 2017b	Variance based 2-step method with linear models and generalized linear regression models using EMMAX (Kang et al., 2010)	5,056 Holstein, 1,231 Jersey cattle	15,355,382 (Holstein), 13,403,916 (Jersey)	Bonferroni	-

Fang et al., 2018	Variance based 2-step method with linear models and generalized linear regression models using EMMAX (Kang et al., 2010)	5,056 Holstein, 1,231 Jersey, 4,310 Nordic Red cattle	15,355,382 (Holstein), 13,403,916 (Jersey), 15,243,827 (Nordic Red)	Bonferroni	6 QTL windows
Fang et al., 2019	Variance based 2-step method with linear models and generalized linear regression models using EMMAX (Kang et al., 2010)	4,578 sires	598,016	Bonferroni	24 SNPs
Freebern et al., 2020	Mixed model using MMAP (O'Connell et al., 2015)	14,382 sires	3,148,506	Bonferroni	1 SNPs
Guo et al., 2012	Mixed model using EMMAX (Kang et al., 2010)	4,803 sires	-	Bonferroni	1 SNPs
Iung et al., 2019	Weighted single-step genomic BLUP (ssGBLUP) method (Wang et al., 2012)	1,067 cows	56,256	-	3 windows of 10 SNPs each
Jiang et al., 2019	Two methods: approximate generalized least squares (AGLS) analysis using EPSNPmpi (Ma et al., 2008; Weeks et al., 2016), and a Bayesian mixed model using BOLT-LMM	293,467 cows	57,067	Bonferroni	140 SNPs
Kirsanova et al., 2020	Variance based two-step method (Svischeva et al., 2012) using GRAMMA-Gamma function in GenABEL R package (Aulchenko et al., 2007)	3,795 sires	613,908	-	93 SNPs
Kurz et al., 2019	Mixed model (Segura et al., 2012) using SVS	43 cows	585,949	Bonferroni	117 SNPs, 27 QTL windows
Lu et al., 2020	Mixed model using ASREML (Gilmour et al., 2006)	1,800 cows	30,348	FDR	-
Ma et al., 2019	EMMAX (Kang et al., 2010). Specific method not specified	3,114 sires	15,388,916	Bonferroni	-
Macciotta et al., 2015	Mixed model using ASREML (Gilmour et al., 2000)	337 cows	6,891	Bonferroni	1 SNPs

Marete et al., 2018a	Mixed model using GCTA (Yang et al., 2011)	CM phenotype: 13,879 Monbéliarde, 9,013 Normande, 32,491 Holstein. SCS phenotype: 20,141 Monbéliarde, 11,965 Normande, 46,732 Holstein	38,827 (Montbéliarde), 38,109 (Normande), 40,810 (Holstein)	Bonferroni	CM: 14 (Montbéliarde), 12 (Nordic Red), 336 (Holstein), SCC: 11 (Holstein)
Marete et al., 2018b	Mixed model using GCTA (Yang et al., 2011)	32,491 cows	49,835	Bonferroni	19 SNPs (CM), 28 SNPs (SCS)
Meier et al., 2020	Mixed model using GEMMA (Zhou and Stephens, 2014)	1,062 cows	38,224	Bonferroni	5 SNPs
Meredith et al., 2012	Two methods: mixed model and modified 'BayesB' Bayesian method (Meuwissen et al., 2001)	773 sires, 493 cows	40,668 (sires), 39,143 (cows)	FDR	10 SNPs (sires)
Meredith et al., 2013	Mixed model using DMU (Madsen and Jensen, 2010)	702 sires	578,181	FDR	138 SNPs, 28 QTL windows
Miles et al., 2020	Mixed model using EMMAX (Kang et al., 2010)	458 cows	581,663	Bonferroni & FDR	166 SNPs
Mulder et al., 2013	Univariate Bayesian SSVS approach (Calus et al., 2008; Verbyla et al., 2010)	1,563 cows	37,590	-	54 SNPs
Naderi et al., 2018	Mixed model using GCTA (Yang et al., 2011)	6,737 cows	43,939	FDR	4 SNPs
Nani et al., 2015	General linear model in PLINK	540 cows	38,872	FDR	10 SNPs
Oliveira et al., 2019a		1,141 Ayrshire sires and cows, 7,329 Holstein sires and cows, 524 Jersey sires and cows	41,517 (Ayrshire), 43,556 (Holstein), 38,194 (Jersey)	-	377 windows
Oliveira et al., 2019b	Single-step genomic BLUP method (Wang et al., 2014) using postGSf90 (Aguilar et al., 2014)	634 Ayrshire, 8,517 Holstein, 849 Jersey sires. 1,193 Ayrshire, 12,146 Holstein, 578 Jersey cows	38,096 (Ayrshire), 40,658 (Holstein), 34,500 (Jersey)	-	1921 SNPs

Pegolo et al., 2020	Structural equation model (Gianola and Sorensen, 2004; Momen et al., 2018, 2019) using the "SNP Snappy" strategy (Meyer and Tier, 2012) in WOMBAT (Meyer 2007)	1,011 cows	37,519	Bonferroni	6 SNPs
Prinsen et al., 2017	Correlation-trend test using SVS	938 cattle (sires + cows)	-	FDR	-
Qanbari et al., 2014	Mixed model using EMMAX (Kang et al., 2010)	3,602 cattle	15,182,131	Bonferroni	2 SNPs
Rincon Florez et al., 2018	Bayesian regression method 'Bayes Cπ' (Habier et al., 2011)	150 cattle (37 sires and 113 cows)	6,510	-	5 SNPs
Sahana et al., 2013	Mixed model (Yu et al., 2006) using DMU (Madsen and Jensen, 2012)	CM phenotype: 2,098 sires. SCS phenotype: 1,676 sires	36,387	FDR	130 SNPs (CM), 13 SNPs (SCS)
Sahana et al., 2014	Mixed model (Yu et al., 2006) using DMU (Madsen and Jensen, 2012)	CM phenotype: 4,992 Holstein, 4,442 Nordic Red, 1,126 Jersey sires. SCS phenotype: 4,992 Holstein sires	648,219	Bonferroni	66 SNPs
Seroussi et al., 2020	PLINK. Specific method not specified	125 cows		Bonferroni	15 SNPs
Siebert et al., 2018	Wald test using PLINK (Purcell et al., 2007)	34 cows (time to SCC 'cure' post *S. uberis* challenge), 36 cows (SCC 0-7d post challenge), 35 cows (SCC 0-28d post challenge)	45,997	Arbitrary cut-off level	16 SNPs
Silva et al., 2020	Weighted single-step GWAS test-day model (Wang et al., 2012)	1,338 cattle (785 cows and 553 sires)	38,323	-	-
Sodeland et al., 2011	Mixed model	CM phenotype: 2,086 sires. SCS phenotype: 2,118 sires	17,349	-	29 SNPs (CM), 13 SNPs (SCS)

Strillacci et al., 2014	Multiple marker test using an adapted pipeline in R (Bagnato 2008; Strillacci et al., 2014)	79 sires	655,665	Bonferroni	171 SNPs
Strucken et al., 2012	Mixed model using GenABEL package in R (Aulchenko et al., 2007)	151 cows	44,962	Bonferroni	17 SNPs
Su et al., 2014	Mixed model using DMU (Madsen et al., 2010)	4,191 sires	44,919	-	-
Tiezzi et al., 2015	Single step genomic BLUP method (Aguilar et al., 2010; Christensen and Lund, 2010; Misztal et al., 2013) using PREGSF90-POSTGSF90 (Aguilar et al., 2014) and THRGIBBS1F90 (Tsuruta and Misztal, 2006)	1,361 sires	39,004	-	53 windows
Tribout et al., 2020	Mixed model using GCTA (Yang et al., 2011)	CM phenotype: 1,857 Montbéliarde, 1.427 Normande, 4,959 Holstein sires. SCS phenotype: 2.438 Montbéliarde, 2,203 Normande, 6,318 Holstein sires. Phenotype not specified: 44,832 Montbéliarde, 44,659 Normande, 46,753 Holstein cows	46,753 (Holstein), 44,832 (Montbéliarde), and 44,659 (Normande)	Bonferroni	5 windows
Veerkamp et al., 2016	Mixed model using GCTA (Yang et al., 2011)	3,469 sires	13,789,029	Arbitrary cut-off levels	-
Wang et al., 2015	Mixed model using FORTRAN (Jiang et al., 2010)	2,093 cows	43,885	Bonferroni	48 SNPs
Welderufael et al., 2018	Mixed model in DMU (Madsen and Jensen, 2013)	993 cows	39,378	Arbitrary cut-off level	14 SNPs (from healthy to CM), 15 SNPs (from CM to healthy)

Wijga et al., 2012	Mixed model using ASREML (Gilmour et al., 2009)	1,484 cows	37,590	FDR	2 SNPs (LASCS), 2 SNPs (SCS-SD)
Yang et al., 2019	Logistic regression (Guo et al., 2018) followed by Bayesian analysis	40 cows	10,058	-	42 SNPs (Bayesian), 51 SNPs (logistic regression), both analysis 27 SNPs
Zhou et al., 2018	Linear regression using SVS	396 cows	-	FDR	-
Zhou et al., 2019	Iterative fixed and random models (FarmCPU; Liu et al., 2016)	473 cows	139,376	Bonferroni	3 SNPs

[1] SCS = somatic cell score, SCC = somatic cell count, CM = clinical mastitis, CNVR = copy number variant region, SNP =single nucleotide polymorphism, QTL = quantitative trait loci, LASCS = lactation-average somatic cell score, SCS-SD = standard deviation for test-day SCS.

Table 4-4. Top 8 significant KEGG pathways

Gene set	Description	FDR
BTA05200	Pathways in cancer	< $1e^{-15}$
BTA04151	PI3K-Akt signaling pathway	< $1e^{-15}$
BTA04060	Cytokine-cytokine receptor interaction	< $1e^{-15}$
BTA04080	Neuroactive ligand-receptor interaction	< $1e^{-15}$
BTA04010	MAPK signaling pathway	< $1e^{-15}$
BTA05167	Kaposi sarcoma-associated herpesvirus infection	< $1e^{-15}$
BTA04630	JAK-STAT signaling pathway	< $1e^{-15}$
BTA05205	Proteoglycans in cancer	< $1e^{-15}$

Figure 4-1. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram of search strategy

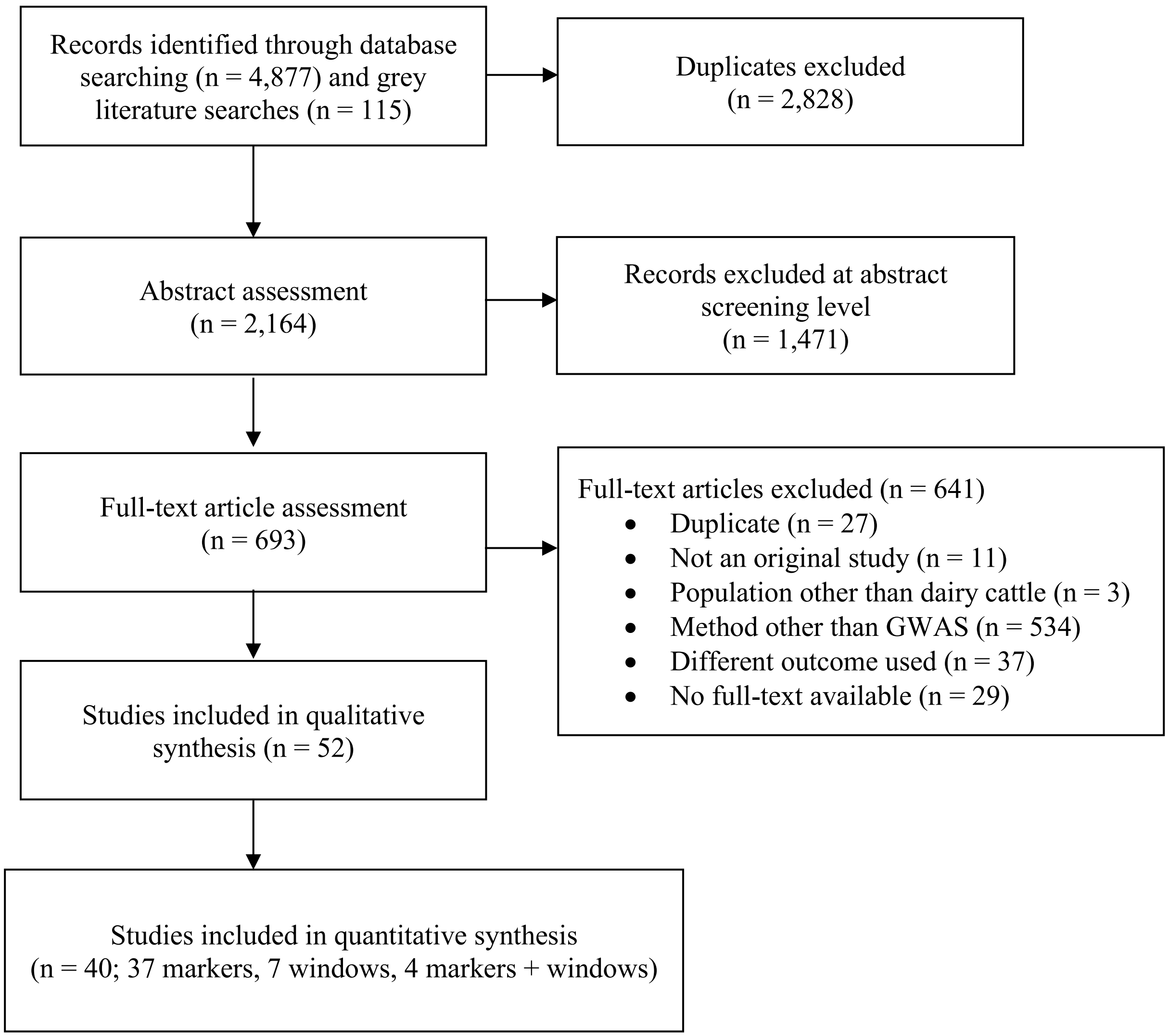


Figure 4-2. Overlapping of genes among 13 clinical mastitis studies. Within the matrix, the intensity of colors corresponds to the proportion of shared genes. Dark red colored diagonals show the proportion of genes (= 1; 100%) for the particular studies

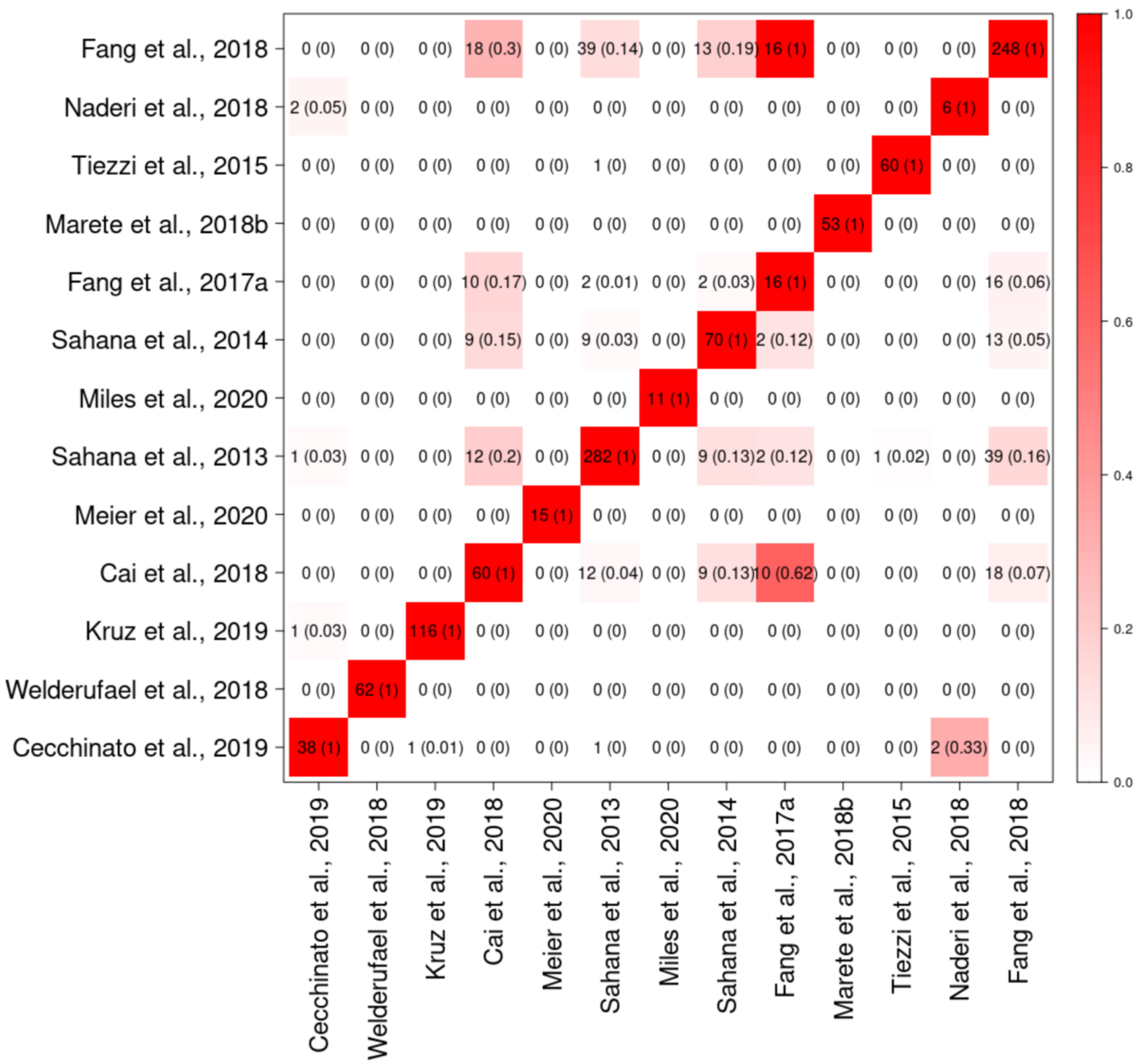


Figure 4-3. Overlap of genes between Holstein, Jersey and Ayrshire breeds; each color in the Venn diagram represents a breed

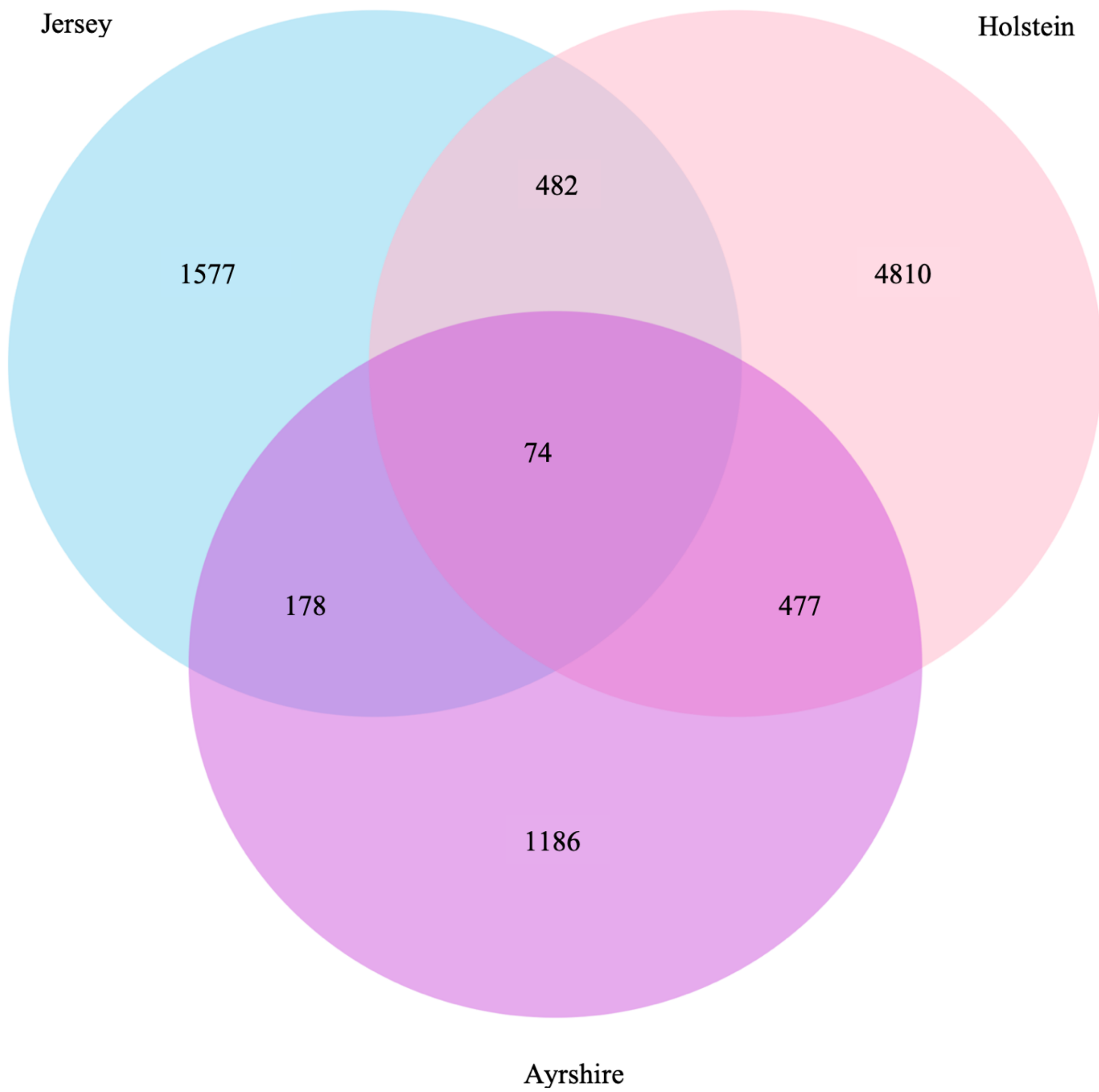


Figure 4-4. Percentage of total annotated QTL (132,140 QTLs) in different QTL classes. The different colors in the pie chart represents different QTL class and the numbers are the corresponding QTL percentages

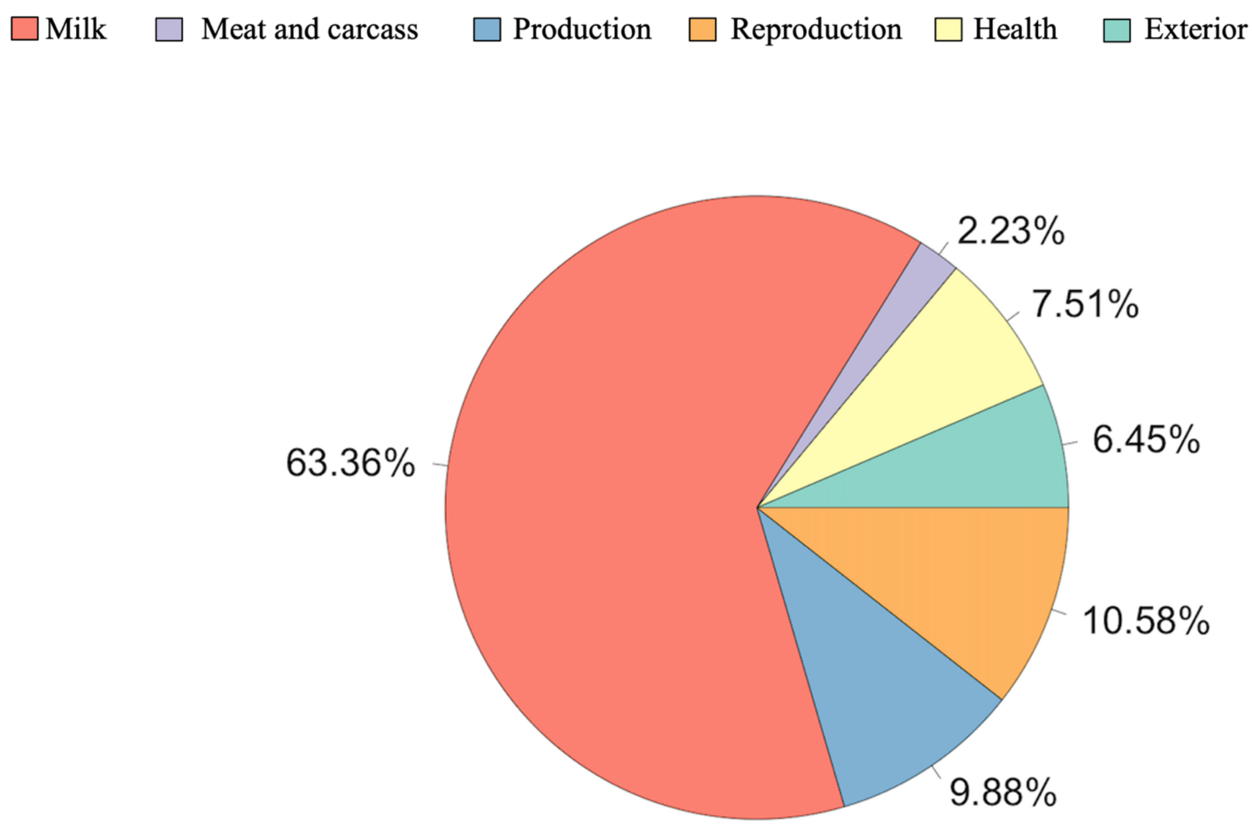


Figure 4-5. Percentage of annotated QTLs associated with “Health” QTL class. Somatic cell count related traits (SCC_SCS)

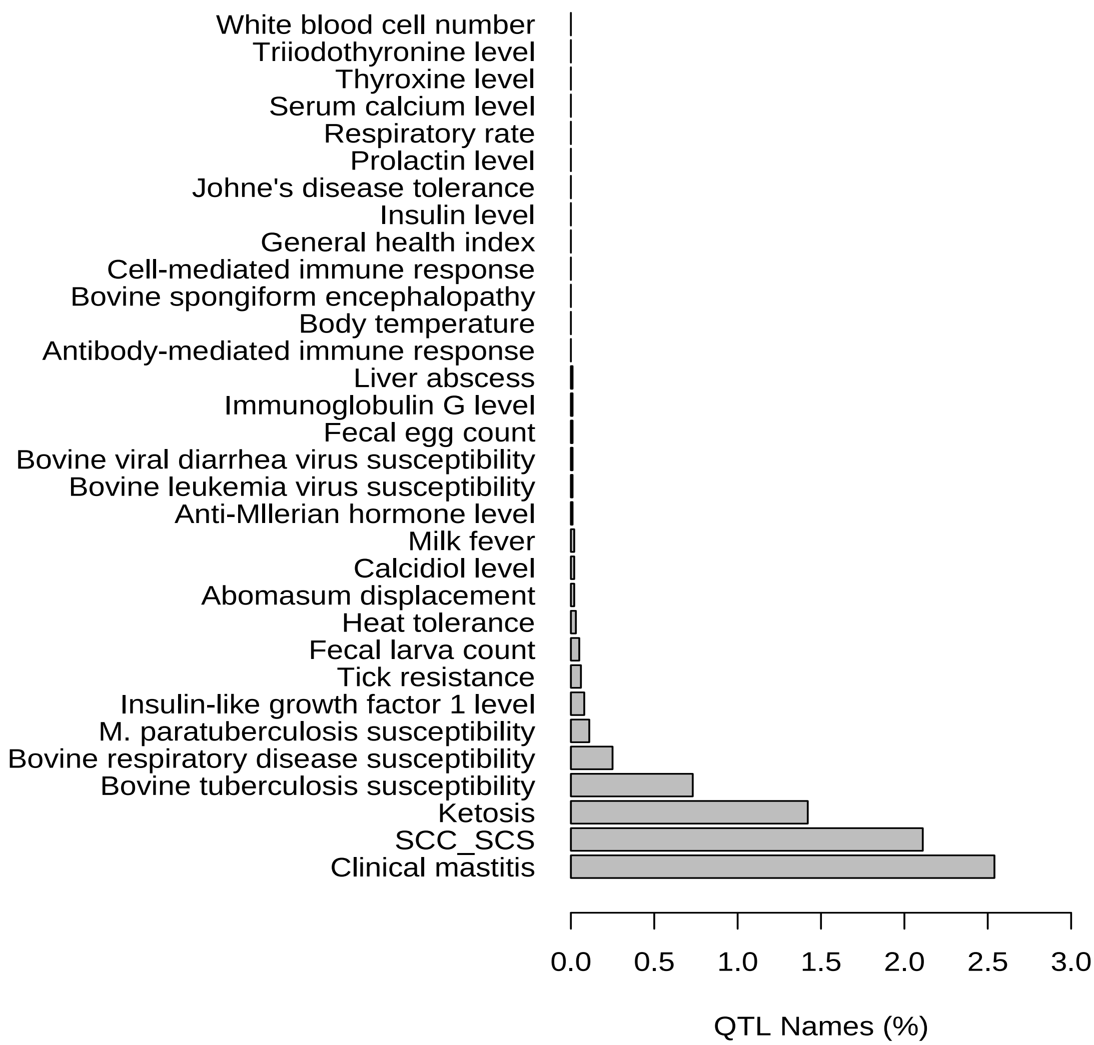


Figure 4-6. Top 25 traits identified in QTLs enrichment analysis. Area of bubbles represents the number of observed QTL for that trait and color of bubbles represents the P-value scale (darker color indicates smaller p-value). P $<10^{-50}$ were truncated to 10^{-50}. Richness factor on the x-axis represents the ratio of number of QTLs and expected umber of that QTL

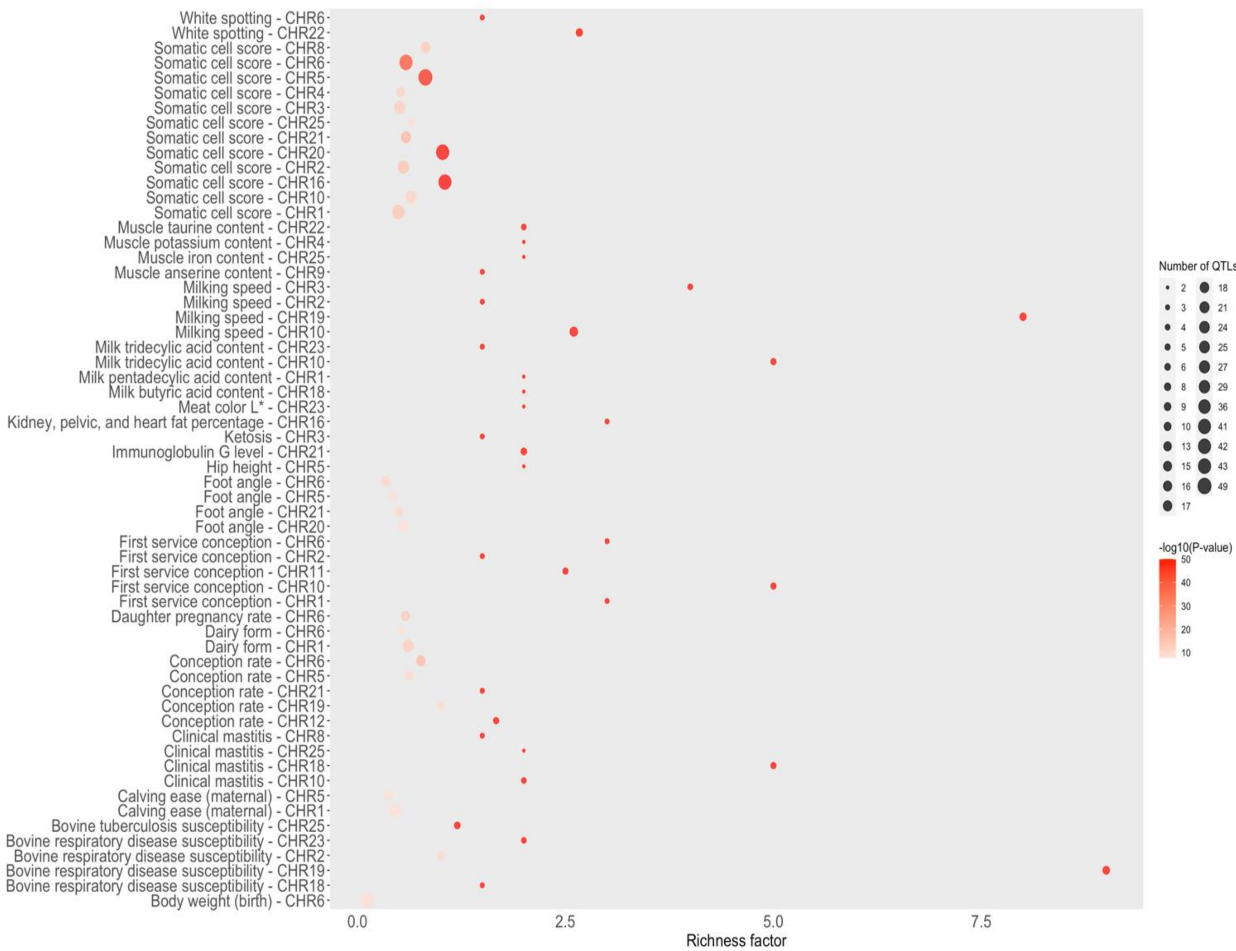


Figure 4-7. Chord plot displaying relationships among top 50 prioritized genes and studies, along with respective traits. Clinical mastitis (CM), SCC-related traits (SCC)

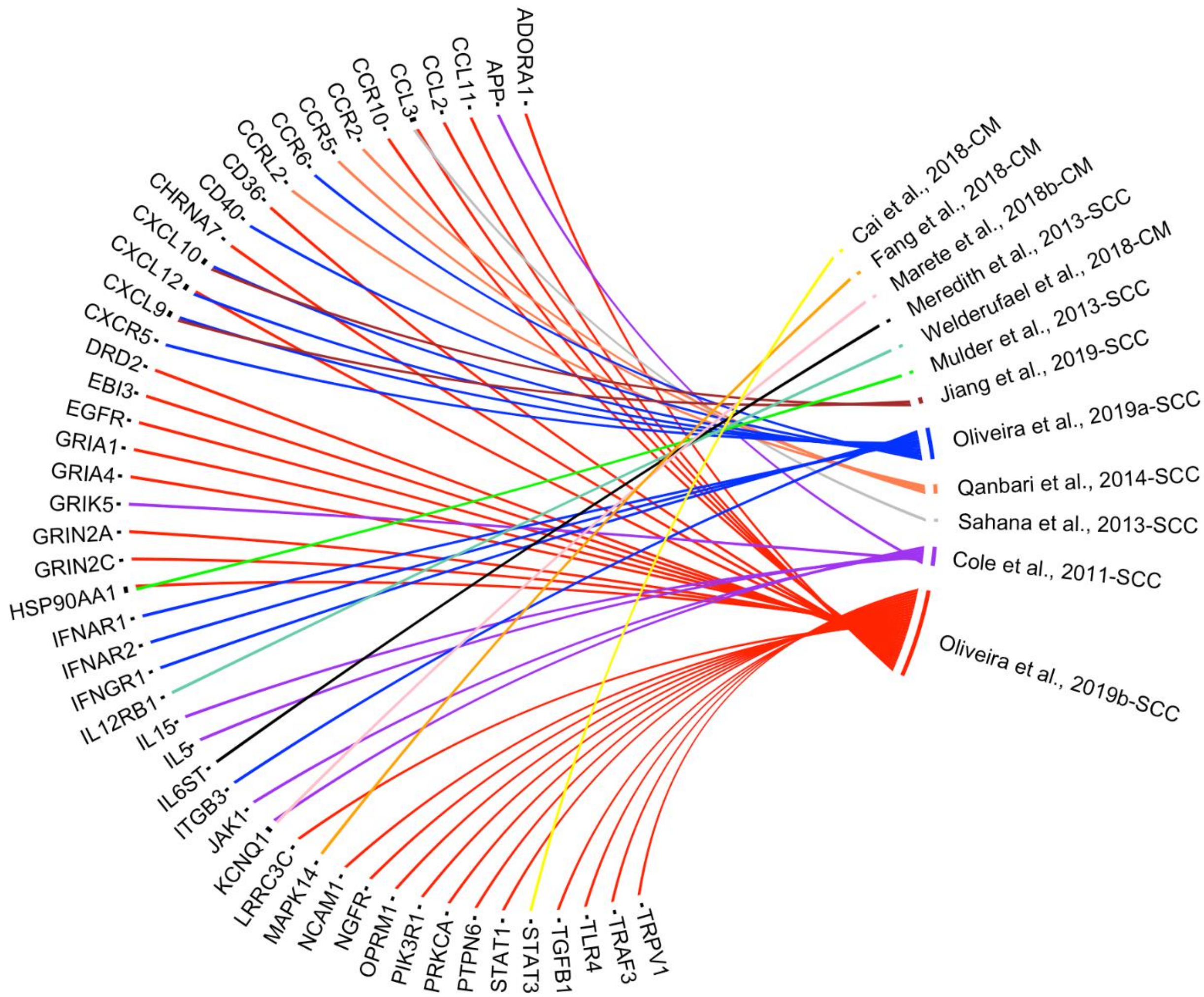
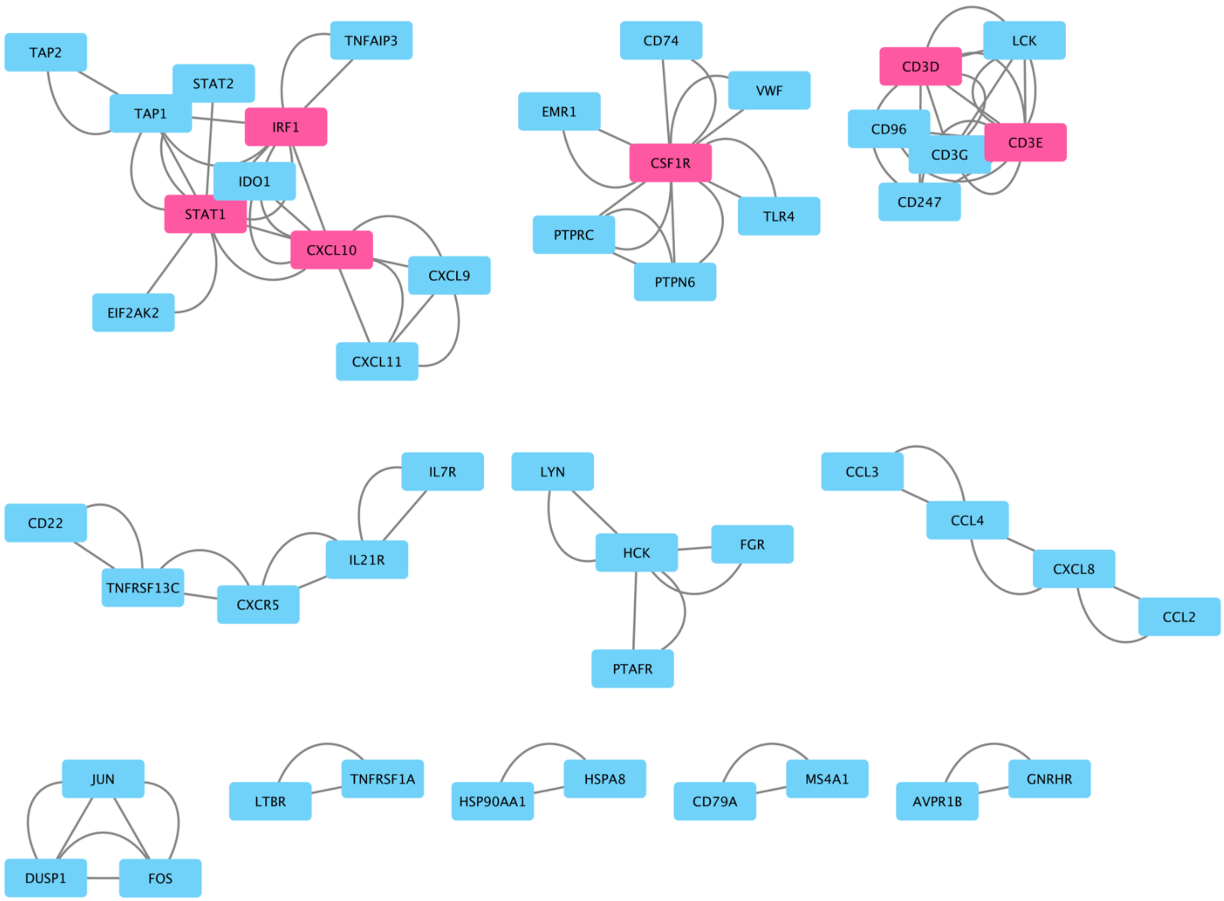


Figure 4-8. Co-expression gene network analysis produced using STRING and Cytoscape, consisting of 48 genes and 92 edges



CHAPTER 5: Association between genomic copy number variants and somatic cell score in Canadian Holstein cattle

5.1. Abstract

Mastitis affects animal health and welfare as well as farm profitability. Understanding the genetic background of mastitis is crucial to performing effective genetic selection for mastitis resistance. Somatic cell count and its derived somatic cell score (SCS) have been used as an indicator trait for mastitis for a long time. Several candidate genes and quantitative trait loci (QTLs) have been identified for SCS using single nucleotide polymorphisms (SNPs). However, copy number variants (CNVs) can potentially reveal additional genomic information that was not previously identified using SNPs. Therefore, this study aimed to perform a genome-wide association study between the de-regressed estimated breeding value (dEBV) for SCS and CNVs. A total of 443 CNV regions (CNVRs) were built using previously identified 2,502 CNVs from 743 animals, which were then used in the association analysis. Association analyses based on linear regressions on a single CNVR at a time, including polygenic effect in the model, were carried out between 44 CNVRs (present in at least 1% of the 718 Holstein cows) with deletions or/and duplications in at least 4 cows and SCS of 400 extreme cows (cows with the 200 highest and 200 lowest dEBVs for SCS). Average length of the identified 443 CNVRs was 362.59 Kb, and they covered 6.45% of the total genome. A high proportion (14.02%) of the bovine genome was covered by CNVRs on *Bos taurus* autosome (BTA) 25. A high number of CNVRs were identified in BTA 1 (n=23) and BTA 4 (n=22), which covered 4.05 and 6.60% of the respective chromosomes. Association analysis resulted in 3 significant CNVRs (duplicates), of which 2 CNVRs (CNVR314 and CNVR68) located on BTA 3 and 12 did not overlap with previously

reported CNVRs for SCS in Holstein cattle. Gene and QTLs annotation resulted in 165 and 337 unique genes and QTLs, respectively. Genes *LCN9, SURF4, SURF6, UBAC1, SNAPC4, LCN8, LCN6, OLFM1 and PMPCA*, which were highly interactive and reported by other studies to be associated with SCS, could be considered as potential functional candidate genes for SCS. Results obtained from this study provide additional value in understanding the genetic background of SCS and should be explored to improve mastitis resistance through genomic selection.

5.2. Introduction

Bovine mastitis, inflammation of the mammary gland, is caused by diverse pathogens, such as *Escherichia coli, Streptococcus uberis, Staphylococcus aureus,* and *Klebsiella* spp. (Schukken et al., 2011). Mastitis affects cow health and welfare and causes severe economic losses to the dairy industry. The cost associated with mastitis in a Canadian dairy farm was estimated at CAD $662 per cow per year (Aghamohammadi et al., 2018). During the infection, somatic cells are released to the udder to provide a defence mechanism (Sharma et al., 2011). An increase in SCC is associated with milk loss. In early lactation (5 to 14 DIM), an increase of one unit of natural log-transformed SCC was associated with a 0.13 Kg/d decrease in milk yield in later lactation (De Vliegher et al., 2005).

One of the potential strategies to combat mastitis is to breed for mastitis resistance through genetic selection, which would result in cumulative, long-lasting, and cost-effective change. Mastitis is a complex polygenic trait, so untangling its genetic architecture is fundamental for identifying potential functional candidate genes. Somatic cell score ($\log_2$(SCC/100,000) + 3 (Shook, 1993)) has been used in breeding programs as an indicator of mastitis resistance, as it has higher heritability (0.18) than clinical mastitis (**CM**; ≤ 0.05) (de Haas et al., 2002; Carlen et al.,

2004) and has a moderate genetic correlation with CM (0.64 to 0.77) (Koeck et al., 2010; CDN, 2014).

To date, genomic variants, such as SNPs, have been used widely in genome-wide association studies (**GWAS**) to identify potential QTLs and functional candidate genes associated with mastitis resistance (Sahana et al., 2013; Welderufael et al., 2018). Copy number variants (CNVs), deletions or duplications of genomic sequences from 50 bp to 5Mb long (Feuk et al., 2006; Mills et al., 2011), have been suggested as genetic markers in the human genome (Feuk et al., 2006; Redon et al., 2006). These structural variants cover large genomic regions, accounting for 18 to 30% of genetic variation in gene expression and are associated with phenotypic variability of disease resistance in humans and mice (Stranger et al., 2007; Henrichsen et al., 2009). Researchers (Xu et al., 2014; Hay et al., 2018) have reported that CNVs could explain part of the missing heritability of complex traits in livestock. Even though SNPs captured 75% of the genomic information provided by CNVs in Holstein cows using Illumina BovineSNP50 arrays, 25% were captured solely by CNVs (Xu et al., 2014), indicating that SNPs alone cannot capture all genomic information. Hence, using CNVs in GWAS can identify additional genomic regions and genes associated with the trait of interest. Due to challenges related to CNV identification and validation, few studies have explored associations of CNV with CM (Szyda et al., 2019; Żukowski et al., 2019) and SCS (Durán Aguilar et al., 2016; Ben Sassi et al., 2016; Prinsen et al., 2017; Zhou et al., 2018; Liu et al., 2019). Moreover, the deletion state of CNVs was reported to be associated with cows susceptible to CM (Szyda et al., 2019).

To our best knowledge, there is no published research on the association of CNVs with SCS in Canadian Holsteins. Copy number variants detected by Butty et al. (2021) provided an excellent opportunity to identify CNV associated with SCS. Therefore, the objective of this study was to examine the association between CNVs and SCS in Canadian Holstein cows.

5.3. Materials and Methods

5.3.1. Data and association analysis

The CNVs used for the GWAS in this study were identified by Butty et al. (2021) using the Holstein cattle breed. In summary, 23,256 CNVs were identified in 5,845 animals using genotype array information aligned to the bovine reference genome ARS-UCD1.2. Average length of the identified CNVs was 168.52 Kb (ranged from 76 bp to 4.17 Mb), and most identified CNVs were deletions (13,724 CNVs). Estimated breeding values of SCS were obtained from the official April 2020 genetic evaluation performed by Lactanet (Guelph, ON, Canada) for 1,560 animals. The EBVs were standardized to a mean of 100 and a standard deviation of 5, where a higher value is more desired (CDN, 2018). The EBVs were de-regressed (dEBV) as described in VanRaden et al. (2009) and used as pseudo-phenotypes for the association analysis. Heritability of 0.18 (CDN, 2014) was used during the de-regression. After the de-regression process, only animals with effective record contributions ≥ 0.1 were kept in the analysis, and CNVs were filtered to be located in autosomal chromosomes and contain at least 3 SNPs. The final dataset comprised 2,502 CNVs from 743 animals (reliability of EBV of SCS ranged from 0.19 to 0.99). Copy number variant regions (CNVRs) were created with 2,502 CNVs using merge command in BEDTools (v2.28.0) (Quinlan and Hall, 2010) which was employed to identify overlapping CNVs and create CNVRs at both individual and population levels. Overlapping was defined as a minimum of one bp to a maximum interval of one Mb (Redon et al., 2006; Butty et al., 2021).

Due to the small number of bulls (29) with pseudo-phenotypes and CNVRs, only 718 cows (dEBV; 101.86 ± 13.17) were considered for further analysis. To increase the power of detection of association analysis, cows were grouped into high and low groups based on the ranking of dEBVs. Top-ranked 200 cows (115.52 ± 8.31) and bottom-ranked 200 cows (87.02 ± 11.48) were selected as high and low pseudo-phenotype groups. High-ranked cows were considered to have low SCC in the milk, and thus to be more preferred (CDN, 2018). The association analysis was carried out between dEBVs of SCS of 400 cows and CNVRs (present in at least 1% of the cows) with deletions and/or duplications in at least 4 cows. The association analysis was performed for one CNVR at a time, including the animal polygenic effect in the model, using a linear mixed regression available in the coxme R package (Therneau, 2015). The model used can be summarized as follow (Lin et al., 2013):

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Xb} + \mathbf{Za} + \mathbf{e},$$

where $\boldsymbol{y}$ is a vector of pseudo-phenotypes (i.e., dEBV), 1 is a vector of ones, μ is the overall mean of the dEBVs, $\mathbf{X}$ is the incidence matrix for the fixed effect of the CNVR considered at a time, $\mathbf{b}$ is the vector of CNV deletion or duplication effects for the CNVs found in the CNVR considered, $\boldsymbol{a}$ is the vector of the random animal additive genetic effects, $\boldsymbol{Z}$ is the incidence matrix for the vector $\boldsymbol{a}$, and $\mathbf{e}$ is the vector of residual effects. The model assumptions were:

$$E[\boldsymbol{y}] = \boldsymbol{Xb} \text{ and } \mathrm{Var}\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{e}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}\sigma_a^2 & 0\\ 0 & \boldsymbol{I}\sigma_e^2\end{bmatrix},$$

where $\boldsymbol{A}$ is the pedigree-based relationship matrix, σ_a^2 is the additive genetic variance, $\boldsymbol{I}$ is an identity matrix, and σ_e^2 is the residual variance. All other terms were previously described.

Copy number variant region was encoded as deletion and duplication depending on the state of CNV (0: deletion of 2 copies, 1: deletion of one copy, 2: normal state, 3: single copy duplication, 4: double copy duplication (Colella et al., 2007)). Thus, if no CNV was predicted (normal 2 copies), both deletion and duplication were encoded as 0. For CNV states of 0 and 1, deletion was encoded as 2 and 1, respectively, whereas duplication was encoded as 0. For CNV states of 3 and 4, duplication was encoded as 1 and 2, respectively, whereas deletion was encoded as 0. More details about the codification can be found in Lin et al. (2013). Deletion and duplication were then fit in the model as covariates. Multiple hypotheses test correction was conducted using chromosome-wide Bonferroni correction. A CNVR was considered significant at chromosome-level if the comparison-wise p-values were ≤ 0.10/N, where N is number of CNVR per chromosome used for the association analysis. The estimated CNV effects were adjusted for their upward bias (Darvasi and Soller, 1994; Falconer and Mackay, 1996), i.e.:

$$b^* = {}^{b}/_{(i)}\,2$$

where b^* is adjusted CNV effect, b is the CNV effect on the original scale (not adjusted), and i is the standardized mean deviation of individuals with values exceeding the threshold (i.e., the selection intensity, 1.201 for 28% of selection on each tail of the distribution). In order to compare identified significant CNVs with other studies, coordinates (ARS-UCD1.2) were converted to respective genome assembly using Lift Genome Annotations. After conversion, only coordinates in the same chromosome were considered.

5.3.2. Functional analysis

Gene and QTL annotation for significant CNVRs were carried out in GALLO R package (Fonseca et al., 2020) using bovine genome assembly ARS-UCD1.2 (Medrano, 2017; Rosen et al., 2018). Positional candidate genes and QTLs were mapped within CNVR. Gene network analysis was conducted for annotated genes using the STRING database (Szklarczyk et al., 2019) for *Bos taurus* species using medium confidence (0.40) interaction score. Protein-protein interaction (PPI) information was derived from text mining, experiments, database, co-expression, neighbourhood, gene fusion and co-occurrence sources. The generated network was brought into Cytoscape 3.9.0 (Shannon et al., 2003), and network analyses were conducted using the Cytoscape network analyzer plugin with undirected network criterion Key genes (nodes) were identified based on criteria such as proteins with a degree of at least 8, betweenness centrality (Bc) ≥ 0.006, and closeness centrality (Cc) ≥ 0.34. Functional information of identified genes was derived from GeneCards (Human Gene Database) (Safran et al., 2022).

5.4. Results

5.4.1. Copy number variant region

A total of 443 CNVRs (Figure 5-1; Table 5-1) distributed across the 29 autosome chromosomes were identified in 743 animals (29 bulls and 718 cows). Of the 443 CNVR samples, 433 were from cows, and the remaining 10 were from bulls. A high number of CNVRs was identified on BTA 1 (23), whereas a low number of CNVRs was identified on BTA 27 (7). Average length of CNVRs was 362.59 Kb ranging from 218 bp to 5.14 Mb. In total, the 443 CNVRs identified covered 6.45% of the total bovine genome (ARS-UCD1.2) (Table 5-1). The genome's highest proportion (14.02%) covered by CNVRs was on BTA 25, whereas the lowest was on BTA 8 (1.77%). Most CNVRs (238) were singletons (consisting of one CNV), whereas 72 CNVRs comprised at least five CNVs. Most CNVRs had low frequency in the population where 32% were identified only in one animal. The most frequent CNVR (CNVR394) identified in the population was present in 18.84% of the animals (140 animals). This CNVR was located on BTA 7 (8,782,634 to 11,008,026 bp) and comprised 147 CNVs. Only 44 CNVRs identified in at least 1% of the cow population and with deletion and/or duplication found in at least 4 cows in the analyzed data file of 400 extreme cows were used for the association analysis.

5.4.2. Association and functional analysis

A total of 3 significant CNVRs were associated with de-regressed EBVs for SCS on BTAs 3, 11 and 12 (Table 5-2; Figure 5-2). All associated CNVRs were duplication, with an average length of 1.65 Mb (ranging from 0.63 to 3.51 Mb). The frequency of associated significant CNVRs in the population ranged from 1.39 to 12.67% and comprised 10 to 108 CNVs. Adjusted CNV effect ranged from -4.76 (CNVR58; BTA 11) to 15.27 (CNVR68; BTA12). Gene annotation resulted in 165 unique positional candidate genes for the three significantly associated CNVRs. There were 22, 142 and 1 genes annotated for CNVR314 (BTA 3), CNVR58 (BTA11), and CNVR68 (BTA12), respectively. Within annotated 128 genes were protein-coding genes. Annotation for QTL resulted in 337 QTLs. Most QTLs were associated with milk yield (254 QTLs), whereas 8 QTLs were associated with health. However, no QTL were annotated for CM or SCC. Gene network analysis consisted of 77 interactive genes (nodes) with 145 interactions (edges) with PPI enrichment *P*-value < 1.0e-16 (Figure 5-3). Based on degree ≥8, betweenness centrality ≥0.006, and closeness centrality ≥0.34, 9 key genes (*LCN9*, *SURF4*, *SURF6*, *UBAC1*, *SNAPC4*, *LCN8*, *LCN6*, *OLFM1* and *PMPCA*) were identified. They were all annotated for CNVR58 on BTA 11. Gene *LCN9*, *SURF4*, *SURF6* and *UBAC1*, had highest the degree of 12 (Bc = 0.22, Cc = 0.38) and were the most interconnected and influential genes in the gene network.

5.5. Discussion

The present study identified 443 CNVRs in 743 Canadian Holstein cattle. The CNVRs identified in at least 1% of all the cows evaluated and with deletions and/or duplications found in at least 4 cows in the analyzed data file of 400 extreme cows based on their dEBV for SCS were used for the association analysis. Genome coverage of all the identified CNVRs was 6.45%. However, Butty et al., 2021 reported 9.43% of genome coverage with CNVRs identified from 23, 256 CNV in 5,845 animals. The lower percentage of genome coverage in the present study was likely due to the lower number of animals (743) used for determining the CNVRs. However, most studies have reported genome coverage by CNVR ranging from 0.68 to 4.7% (Fadista et al., 2010; Liu et al., 2010; Hou et al., 2011; 2012). Identification of a CNVR highly depends on the number of animals, density of SNP panel, whole genome sequence availability, definition of CNVR and the parameters used to create the CNVRs (Durán et al., 2016; Ben Sassi et al., 2016; Butty et al., 2020).

Some studies have performed GWAS for SCC/SCS (Durán et al. (2016); Ben Sassi et al. (2016); Prinsen et al. (2017); Zhou et al. (2018); Liu et al. (2019)), CM (Szyda et al. (2019); Żukowski et al. (2019); Lee et al. (2021)) with CNV in different Holstein populations (Mexico, Spain, Poland, USA and the Netherlands) and Brown Swiss. The present study identified 1 overlapping and 2 novel significant CNVRs associated with SCS in the Canadian Holstein population. For instance, CNVR58 identified in the current study on BTA 11 overlapped with CNVR_125_SVS identified by Durán et al. (2016) for SCC in Holstein cows using the SVS 8.3.1 software, where authors used a similar approach to present study (i.e., selecting 2 extreme groups). Gene annotation of CNVR58 in the present study also resulted in the similar genes (i.e., *LHX3*, *QSOX2*, *GPSM1*, *DNLZ*, *CARD9*, *SNAPC4*, *PMPCA*, *INPP5E*, *SEC16A*, *NOTCH1*, *EGFL7*, *bta-mir-126*, and *AGPAT2*) reported by Durán et al. (2016) for CNVR_125_SVS.

The 2 significant CNVRs (CNVR314 and CNVR68) identified on BTA 3 and 12 in this study did not overlap with CNVRs previously reported for SCS in Holstein, thus, emphasizing the need for further validation in a larger population with sequence information or quantitative polymerase chain reaction (qPCR). Reasons for differences in identified CNVRs associated with SCS may be due to difference in the populations, the number of individuals and the methodology used for association analysis. For instance, 2 different software programs (SVS 8.3.1 and PENNCNV-CNVRuler) were used for CNVR creation and association analysis for the same population with a different number of individuals (80 to 140 for SVS 8.3.1 and from 7 to 103 for PENNCNV-CNVRULER, respectively) in Durán et al. (2016). These authors reported that 85 CNVs (24 CNVR in ≥ 2% population) and 47 CNVRs related to SCS were identified by SVS 8.3.1 and PENNCNV-CNVRuler, respectively, with an overlap of only 6 CNVRs. In addition, CNVRs associated with SCS identified in the present study did not overlap with Zhou et al. (2018) who identified 4 CNVRs associated with SCS, and Liu et al. (2019) who identified 2 CNVRs for SCS and 2 CNVRs for CM in Holstein cows as well as with Szyda et al. (2019) who reported CNVR associated with CM. Prinsen et al. (2017) and Ben Sassi et al. (2016) did not identify any significant CNVRs associated with SCS in Brown Swiss and Holstein dairy cattle, respectively. However, CNVR27 on BTA 10 (22965315 to 23448437 bp) identified in the present study overlapped with a CNVR (22.11-25.51 Mb) identified in Prinsen et al. (2017) before FDR correction in Brown Swiss Cattle. This illustrates the possible effect of the population, methodology, and number of individuals used to define the CNVR on the CNVRs, thus having a potential impact on the association analysis between CNVRs and the trait of interest. Regarding the adjusted CNV effect, two CNVRs exhibited a negative impact, with 2 of them being duplications (CNVR314 on BTA 3 and CNVR58 on BTA 11). This suggests that the presence of either CNVR314 and CNVR58 or both, or the absence of CNVR383, could lead to a decrease in dEBV, resulting in an undesirable increase in SCS.

Gene annotation resulted in a total of 165 genes in the present study where 9 genes (*LCN9*, *SURF4*, *SURF6*, *UBAC1*, *SNAPC4*, *LCN8*, *LCN6*, *OLFM1* and *PMPCA*) annotated for CNVR58 on BTA 11 were highly interactive. Interestingly, 7 of those 9 genes were already reported by other studies to be associated with SCS. The top interactive gene *LCN9* was reported to be associated with SCS in Brown Swiss cattle by Fang et al. (2019). Lipocalin 9 is a protein-coding gene, which binds small hydrophobic ligands and transports them to specific cells (Suzuki et al., 2004). Gene *LCN9* is associated with tick infestation (Safran et al., 2022), and interestingly, CNVR58 region was also annotated for 2 tick resistance QTLs in the current study. Genes *SURF4*, *SURF6* and *UBAC1* were reported to be associated with SCS by Oliveira et al., 2019b in Ayrshire and Jersey. *Surfeit 4* is a protein coding gene, and the encoded protein is a conserved integral membrane protein that interacts with endoplasmic reticulum-Golgi intermediate compartment proteins (Safran et al., 2022). Gene *SURF4* is involved in innate immune system (Safran et al., 2022). Protein encoded by gene *SURF6* function as a nucleolar-matrix protein with nucleic acid-binding properties (Safran et al., 2022). Gene *UBAC1* predicted to be involved in protein ubiquitination and involved in innate immune system (Safran et al., 2022). Genes *SNAPC4*, *LCN8*, and *PMPCA* were reported to be associated with SCS in Holsteins by Duran et al. (2017), where authors used a similar approach to the present study of selecting 2 extreme group animals for the association analysis. The encoded protein of *SNAPC4* possesses a Myb DNA-binding domain which plays a crucial role in facilitating the transcription of RNA polymerase II and III polymerase from promoters of small nuclear RNA (Safran et al., 2022). The protein encoded *PMPCA* is located within the mitochondrion, serving as the alpha subunit in a proteolytic heterodimer. This heterodimer plays a key role in removing the transit peptide from nuclear-encoded mitochondrial proteins (Safran et al., 2022).

A limitation of this study is the relatively low number of animals used for the association analysis. To increase the power of detecting a true associated genetic variant, cows were grouped into two extreme groups based on the dEBV for SCS, similar to the strategy performed by Durán et al. (2016). The estimated CNV effect would be biased upward; hence an adjustment was applied to account for the selection intensity used to create the extreme groups. Results obtained from this study add value to the genetic information of mastitis resistance, and this information can be used to improve mastitis resistance through genomic selection. Including CNV together with SNPs in genomic prediction may improve genomic prediction accuracy (Hay et al., 2018), thus resulting in a more accurate selection of animals for mastitis resistance using SCS. However, results from this study need to be validated using a larger number of animals and other omics technologies, such as transcriptomics. In this study, genes mapped within CNVRs lay the foundation to investigate further those genes for their functional relevance to SCS.

5.6. Conclusions

This study outlines CNVRs associated with SCS and offers insights into the functional interpretation of the related CNVRs. A total of 443 CNVRs identified in this study covered 6.45% of the bovine genome. The association analysis resulted in three significant CNVRs, where two CNVRs (CNVR314 and CNVR68) were not previously reported as associated with SCS for Holstein cows. Gene and QTL annotation resulted in 165 and 337 unique genes and QTLs. Nine highly interactive genes *LCN9*, *SURF4*, *SURF6*, *UBAC1*, *SNAPC4*, *LCN8*, *LCN6*, *OLFM1* and *PMPCA* were annotated for CNVR58 on BTA 11 and could be considered potential functional candidate genes. This study lays a good foundation for understanding the genetic background of SCS using CNVs, and this information can be incorporated into the genomic selection to breed for improved mastitis resistance using SCS. However, further information will be required to reinforce the results of this analysis.

5.7. References

Aghamohammadi, M., D. Haine, D. F. Kelton, H. W. Barkema, H. Hogeveen, G. P. Keefe, and S. Dufour. 2018. Herd-level mastitis-associated costs on Canadian dairy farms. Front. Vet. Sci. 5:100.

Ben Sassi, N., Ó. González-Recio, R. de Paz-Del Río, S. T. Rodríguez- Ramilo, and A. I. Fernández. 2016. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. J. Dairy Sci. 99:6371–6380.

Butty, A. M., T. C. S Chud, D. F. Cardoso, L. S. F. Lopes, F. Miglior, F. S. Schenkel, A. Cánovas, I. M. Häfliger, C. Drögemüller, P. Stothard, and F. Malchiodi. 2021. Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle.

J. Dairy Sci. 104:8050-8061.

Butty, A. M., T. Chud, F. Miglior, F. Schenkel, A. Kommadath, K. Krivushin, J. Grant, I. Häfliger, C. Drögemüller, A. Canovas, P. Stothard, and C. F. Baes. 2020. High confidence copy number variants identified in Holstein dairy cattle from whole genome sequence and genotype array data. Sci. Rep. 10:8044.

Carlen, E., E. Strandberg, and A. Roth. 2004. Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish Holstein cows. J. Dairy Sci. 87:3062-3070.

CDN. 2014. Mastitis resistance selection: now a reality! Accessed November 6, 2023. https://www.cdn.ca/document.php?id=356

CDN. 2018. Should somatic cell score evaluations be expressed as relative breeding values?. Accessed November 6, 2023. https://www.cdn.ca/Articles/GEBAPR2018/Possible%20Expression%20of%20SCS%20as%20RBV.pdf

Colella, S., C. Yau, C, J. M. Taylor, G. Mirza, H. Butler, P. Clouston, A. S. Bassett, A. Seller, C. C. Holmes and J. Ragoussis. 2007. QuantiSNP: An Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 35:2013–2025.

Constable, P., K. Hinchcliff, S. Done, W. Grünberg, and O. Radostits. 2017. Veterinary medicine: A textbook of the diseases of cattle, horses, sheep, pigs and goats (Edition 11). Elsevier, St. Louis, MO.

Darvasi, A. and M. Soller. 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics. 138:1365-1373.

De Haas, Y., H. W. Barkema, R. F. Veerkamp. 2002. Genetic parameters of pathogen-specific incidence of clinical mastitis in dairy cows. Anim. Sci. 74: 233-242.

De Vliegher, S., H. W. Barkema, H. Stryhn, G. Opsomer, and A. De Kruif. 2005. Impact of early lactation somatic cell count in heifers on milk yield over the first lactation. J. Dairy Sci. 88:938–947.

Durán Aguilar, M., S. I. Román Ponce, F. J. Ruiz López, E. González Padilla, C. G. Vásquez Peláez, A. Bagnato, and M. G. Strillacci. 2016. Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. J. Anim. Breed. Genet. 134:49–59.

Fadista, J., B. Thomsen, L. E. Holm, and C. Bendixen. 2010. Copy number variation in the bovine genome. BMC Genom. 11:284.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th edn. Longmans Green, Harlow, Essex, UK.

Fang, Z. H. and H. Pausch. 2019. Multi-trait meta-analyses reveal 25 quantitative trait loci for economically important traits in Brown Swiss cattle. BMC Genom. 20:695.

Fonseca, P. A. S., A. Suárez-Vega, G. Marras, and A. Cánovas. 2020. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience 9:149.

Hay, E. H. A., Y. T. Utsunomiya, L. Xu, Y. Zhou, H. H. R. Neves, R. Carvalheiro, D. M. Bickhart, L. Ma, J. F. Garcia, and G. E. Liu. 2018. Genomic predictions combining SNP markers and copy number variations in Nellore cattle. BMC Genom. 19:441.

Henrichsen, C. N., E. Chaignat, and A. Reymond. 2009. Copy number variants, diseases and gene expression. Hum. Mol. Genet. 18:R1–R8.

Hou, Y., D. M. Bickhart, H. Chung, J. L. Hutchison, H. D. Norman, E. E. Connor, and G. E. Liu. 2012. Analysis of copy number variations in Holstein cows identify potential mechanisms contribut- ing to differences in residual feed intake. Funct. Integr. Genom. 12:717–723.

Hou, Y., G. E. Liu, D. M. Bickhart, M. F. Cardone, K. Wang, E-S. Kim, L. K. Matukumalli, M. Ventura, J. Song, P. M. VanRaden, T. S. Sonstegard, and C. P. Van Tassell. 2011. Genomic characteristics of cattle copy number variations. BMC Genom. 12:127.

Koeck, A., B. Heringstad, C. Egger-Danner, C. Fuerst, P. Winter, and B. Fuerst-Waltl. 2010. Genetic analysis of clinical mastitis and somatic cell count traits in Austrian Fleckvieh cows. J. Dairy Sci. 93:5987-5995.

Lee, Y. L., H. Takeda, G. C. M. Moreira, L. Karim, E. Mullaart, W. Coppieters, GplusE consortium. R. Appeltant, R. F. Veerkamp, M. A. M. Groenen, M. Georges, M. Bosse,T. Druet, A. C. Bouwman, and C. Charlier. 2021. A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle. PLoS Genet. 17:e1009331.

Liao, Y., J. Wang, E. J. Jaehnig, Z. Shi, and B. Zhang. 2019. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 47:199–205.

Lin, C. F., A. C. Naj, and L. S. Wang. 2013. Analyzing copy number variation using SNP array data: protocols for calling CNV and association tests. Curr. Protoc. Hum. Genet. 79:1-27.

Liu, G.E., Y. Hou, B. Zhu, M. F.Cardone, L. Jiang, A. Cellamare, A. Mitra, L. J. Alexander, L. L. Coutinho, M. E. Dell'Aquila, and L. C. Gasbarre. 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20:693-703.

Liu, M., L. Fang, S. Liu, M. G. Pan, E. Seroussi, J. B. Cole, L. Ma, H. Chen, and G. E. Liu. 2019. Array CGH-based detection of CNV regions and their potential association with reproduction and other economic traits in Holsteins. BMC Genom. 20:1-10.

Medrano, J. F. 2017. The new bovine reference assembly and its value for genomic research. Proc. Assoc. Advmt. Anim. Breed. Genet. 22:161–166.

Mills, R. E., K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, A. Chinwalla, D. F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L. M. Iakoucheva, Z. Iqbal, S. Kang, J. M. Kidd, M. K. Konkel, J. Korn, E. Khurana, D. Kural, H. Y. K. Lam, J. Leng, R. Li, Y. Li, C. Y. Lin, R. Luo, X. J. Mu, J. Nemesh, H. E. Peckham, T. Rausch, A. Scally, X. Shi, M. P. Stromberg, A. M. Stütz, A. E. Urban, J. A. Walker, J. Wu, Y. Zhang, Z. D. Zhang, M. A. Batzer, L. Ding, G. T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, K. Ye, E. E. Eichler, M. B. Gerstein, M. E. Hurles, C. Lee, S. A. McCarroll, and J. O. Korbel. 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470:59–65.

Narayana, S. G., F. Schenkel, F. Miglior, T. Chud, E. A. Abdalla, S. A. Naqvi, F. Malchiodi, and H. W. Barkema. 2021. Genetic analysis of pathogen-specific intramammary infection in dairy cows. J. Dairy Sci. 104:1982–1992.

Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, J. P. Cant, and F. S. Schenkel. 2019b. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. J. Dairy Sci. 102:9995–10011.

Prinsen, R.T.M.M., A. Rossoni, B. Gredler, A. Bieber, A. Bagnato, and M. G. Strillacci. 2017. A genome wide association study between CNV and quantitative traits in Brown Swiss cattle. Livest. Sci. 202:7–12.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H.

Shapero, A.R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. González, M. Gratacòs, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, J. Zhang, L. Armengol, D.F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, and M. E. Hurles. 2006. Global variation in copy number in the human genome. Nature 444:444–454.

Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, A. Zimin, C. Dreischer, S. Schultheiss, R. Hall, S. G. Schroeder, C. P. Van Tassell, T. P. L. Smith, and J. F. Medrano. 2018. Modernizing the bovine reference genome assembly. Proc. World Congr. Genet. Appl. Livest. Prod., Vol. Mol. Genet. 3:802.

Safran, M.,N. Rosen, M. Twik, R. BarShir, T. Iny Stein, D. Dahary, S. Fishilevich, and D. Lancet D. 2022. The GeneCards Suite. In: Abugessaisa I., Kasukawa T. (eds) Practical Guide to Life Science Databases:27-56. Springer, Singapore.

Sahana, G., B. Guldbrandtsen, B. Thomsen, and M. S. Lund. 2013. Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. Anim. Genet. 44:620-626.

Schukken, Y. H., J. Günther, J. Fitzpatrick, Michael C. Fontaine, L. Goetze, O. Holst, J. Leigh, W. Petzl, H.-J. Schuberth, A. Sipka, D. G. E. Smith, R. Quesnell, J. Watts, R. Yancey, H. Zerbe, A. Gurjar, R. N. Zadoks, and H.-M. Seyfert. 2011. Host-response patterns of intramammary infections in dairy cows. Vet. Immunol. Immunopathol. 144:270-289.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13:2498–2504.

Sharma, N., N. K. Singh, and M. S. Bhadwal. 2011. Relationship of somatic cell count and mastitis: An overview. Asian-Australas. J. Anim. Sci. 24:429-438.

Shook, G. E. 1993. Genetic improvement of mastitis with through selection on somatic cell count. Vet. Clin. North Am. Food Anim. Pract. 9:563–581.

Stranger, B. E., M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazlsy, N. Thorne, R. Redon, C. P. Bird, A. De Grassi, C. Lee, C. Tyler-Smith, N. Carter, S. W. Scherer, S. Tavaré, P. Deloukas, M. E. Hurles, and E. T. Dermitzakis. 2007. Relative impact of nucleotide and copy number variation on gene phenotypes. Science 315:848–853.

Suzuki, K., J. J. Lareyre, D. Sánchez, G. Gutierrez, Y. Araki, R. J. Matusik, and M. C. Orgebin-Crist. 2004. Molecular evolution of epididymal lipocalin genes localized on mouse chromosome 2. Gene 339:49-59.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, and K. P. Tsafou. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43:D447–D452.

Szyda, J., M. Mielczarek, M. Frąszczak, G. Minozzi, J. L. Williams, and K. Wojdak-Maksymiec. 2019. The genetic background of clinical mastitis in Holstein-Friesian cattle. Animal 13:2156-2163.

Therneau, T. Package ‘coxme’. Mixed Effects Cox Models. 2015. R package version, 2.24. URL http://r-forge.r-project.org.

Welderufael, B. G., P. Løvendahl, D. J. de Koning, L. L. G. Janss, and W. F. Fikse. 2018. Genome-wide association study for susceptibility to and recoverability from mastitis in Danish Holstein cows. Front. Genet. 9:141.

Xu, L., J. B. Cole, D. M. Bickhart, Y. Hou, J. Song, P. M. VanRaden, T. S. Sonstegard, C. P. Van Tassell, and G. E. Liu. 2014. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genom. 15:683.

Zhou, Y., E. E. Connor, G. R. Wiggans, Y. Lu, R. J. Tempelman, S. G. Schroeder, H. Chen, and G. E. Liu. 2018. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. BMC Genom. 19:314.

Żukowski, K., P. Topolski, and A. Żarnecki. 2019. Genome-wide patterns of copy number variation in case-control mastitis study of Polish HF cows using high-density Illumina Bovine SNP50 v3 BeadChip array. Translational Res. Vet. Sci. 2:9-22.

Table 5-1. Number of copy number variants regions (CNVRs) identified in each chromosome, length of CNVRs and proportion of autosomal genome covered

Chromosome[1]	No. CNVRs	Length (Mb)	Proportion covered (%)
1	23	3.86	2.43
2	20	5.52	4.06
3	16	7.99	6.60
4	22	8.23	6.86
5	19	8.23	6.86
6	22	11.41	9.69
7	22	8.55	7.72
8	13	2.01	1.77
9	19	5.22	4.95
10	17	4.39	4.25
11	18	7.36	6.88
12	17	11.45	13.13
13	20	4.47	5.36
14	11	6.52	7.91
15	10	3.00	3.53
16	14	5.21	6.43
17	12	4.09	5.59
18	18	6.55	9.94
19	14	3.25	5.12
20	13	3.57	4.96
21	18	6.66	9.53

22	12	4.01	6.59
23	9	2.88	5.48
24	12	3.64	5.84
25	12	5.94	14.02
26	13	5.53	10.63
27	7	1.57	3.44
28	11	2.53	5.51
29	9	7.00	13.70
TOTAL/MEAN	443	160.63	6.45

[1]Chromosome sizes were obtained from the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1/#/st), using the latest estimation based on the ARS-UCD1.2 reference genome.

Table 5-2. Significant copy number variant regions (CNVRs) identified for somatic cell score

CNVR	BTA	chromStart[1]	chromEnd[2]	Length	Type	No. genes annotated	*P*-value[3]	Adjusted CNV effect
CNVR314	3	22,071,533	22,886,248	814,715	Duplication	22	0.003	-14.59
CNVR58	11	103,257,028	106,768,608	3,511,580	Duplication	142	0.047	-4.76
CNVR68	12	59,372,779	60,002,395	629,616	Duplication	1	0.002	15.27

[1]Chromosome starting position

[2]Chromosome ending position

[3]*P*-value before Bonferroni correction

Figure 5-1. Distribution of 443 copy number variant regions (CNVRs, represented by strips) on the bovine autosomal genome. The darker and thicker are the strips, the longer are the CNVRs.

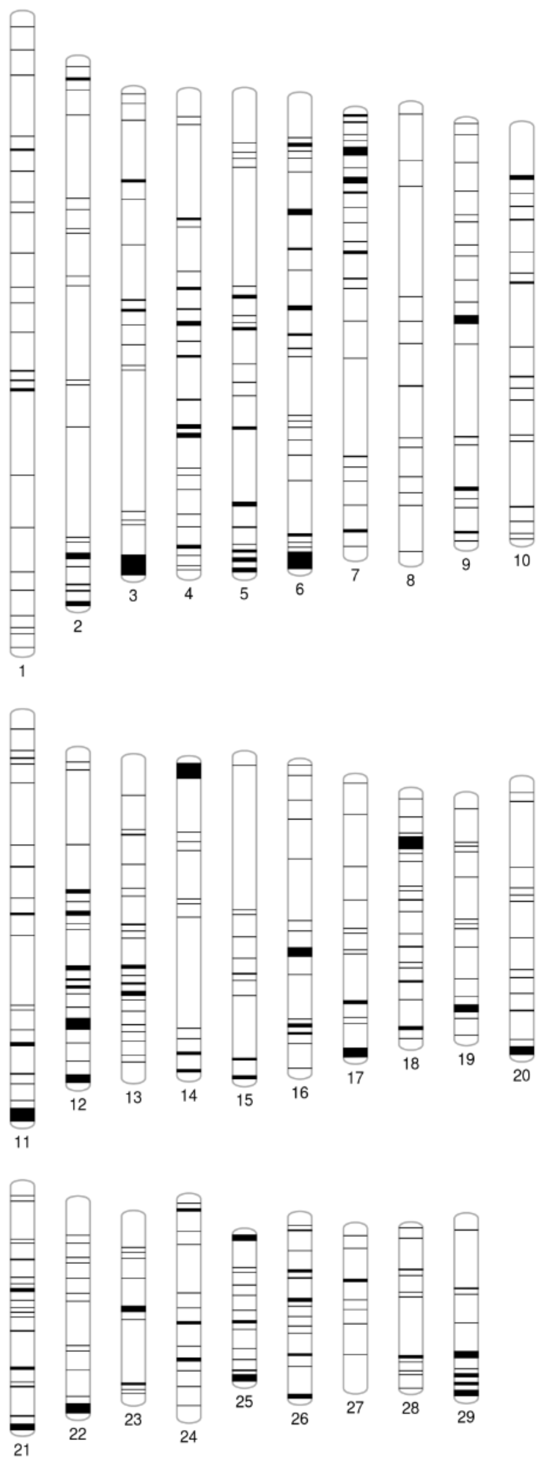


Figure 5-2. Distribution of 3 significant copy number variant regions (CNVRs) on the bovine autosomal genome.

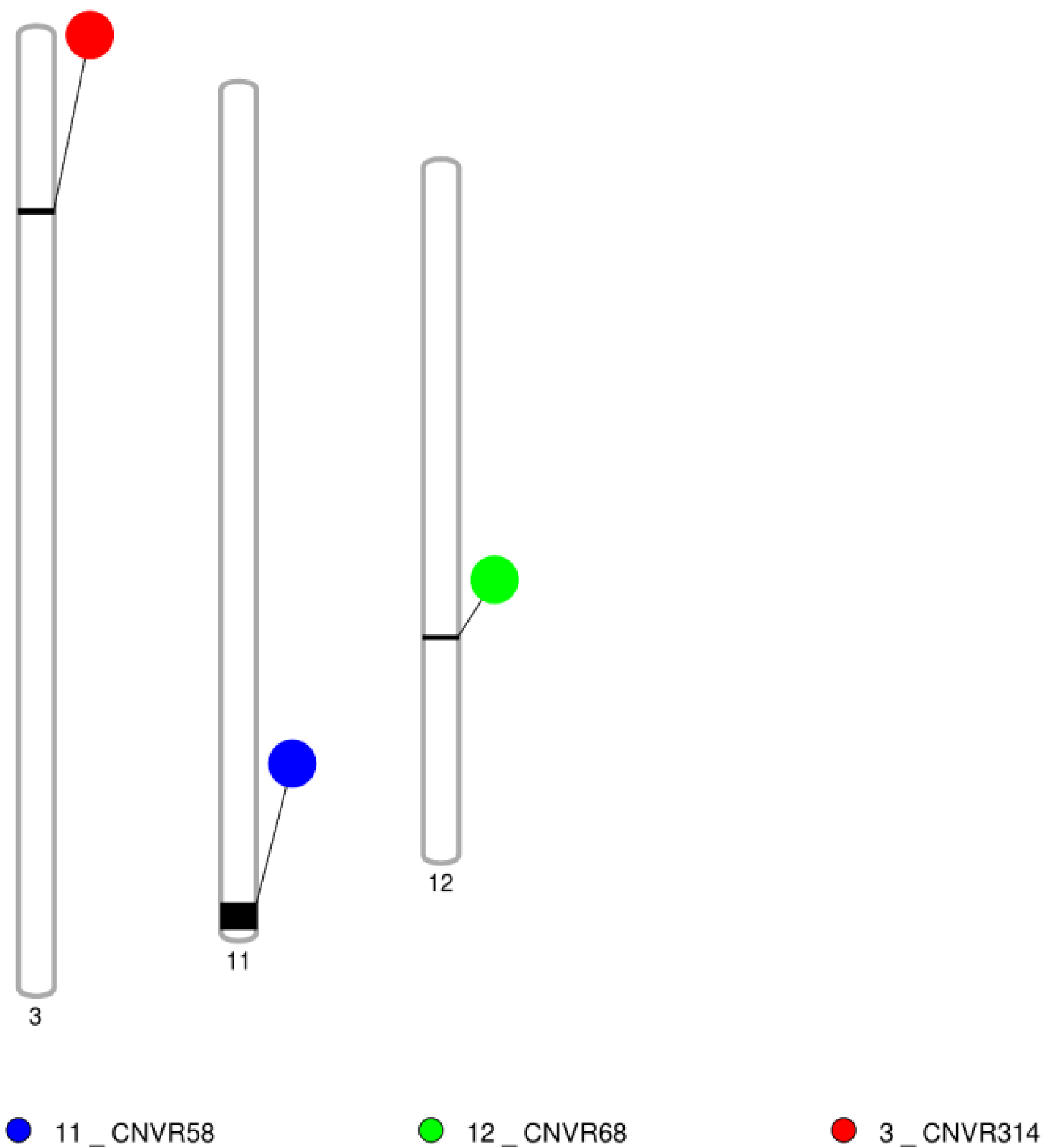
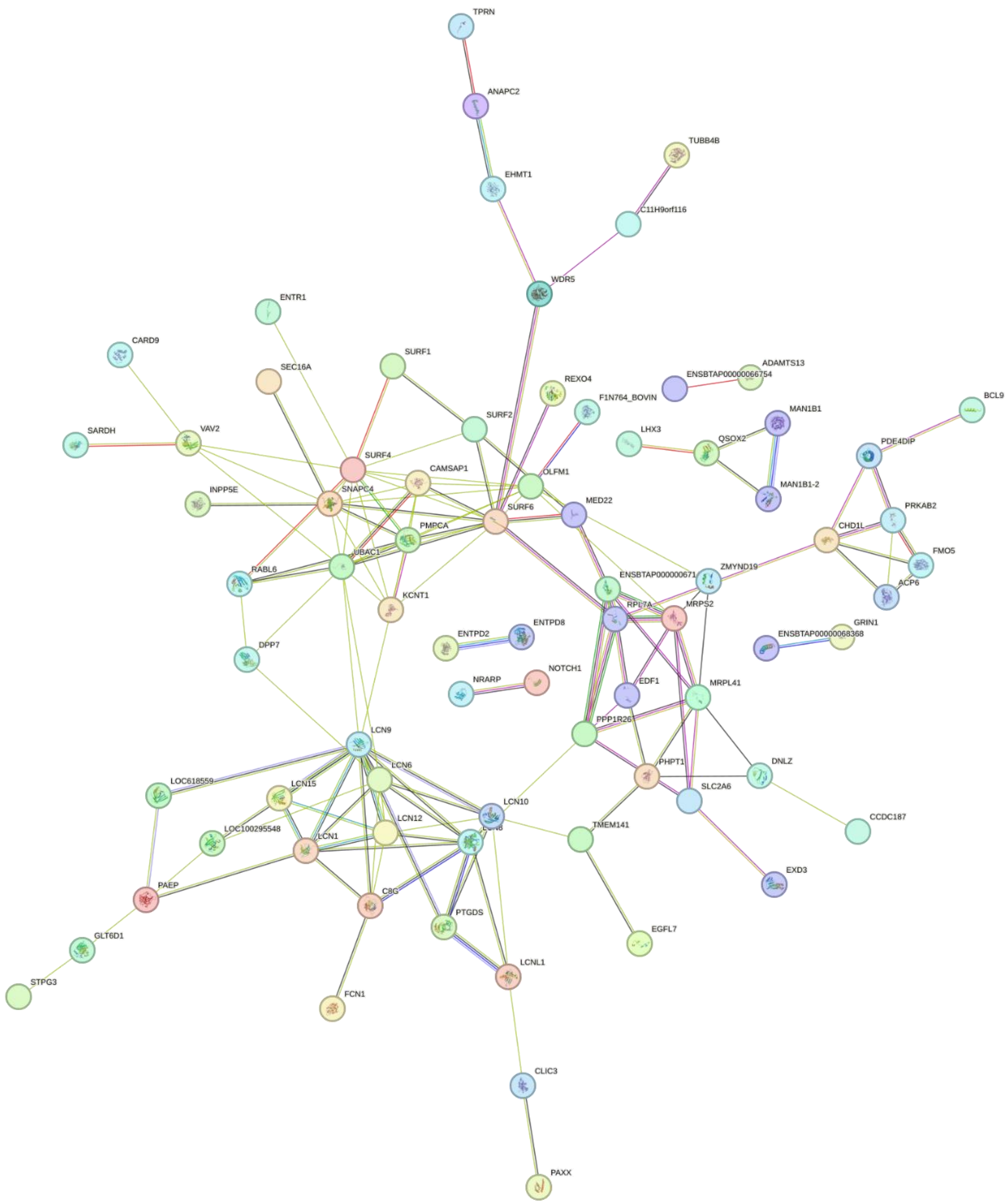


Figure 5-3. Gene network produced using STRING. Genes are represented by spheres. Edges represent the protein-protein interaction. Source of interaction: curated database: light blue, experimentally determined: pink, gene neighbourhood: green, gene fusion: red, gene co-occurrence: dark blue, text mining: yellow, co-expression: black, and protein homology: purple.



CHAPTER 6: Genomic predictions of mastitis-related traits in early lactation of first lactating dairy heifers using single-step genomic approach

6.1. Abstract

Mastitis is the most frequent and economically burdensome disease in the dairy industry. Given its low heritability and the resource-intensive nature of phenotype collection, the training population (i.e., animals with both genotype and phenotype) for mastitis is inherently limited. Consequently, application of the single-step Genomic Best Linear Unbiased Prediction (**ssGBLUP**) methodology emerges as a better alternative to the current multistep genomic approach adopted in Canada. This approach combines phenotypic data and genomic and pedigree information simultaneously through a hybrid relationship matrix (H) to generate genomic estimated breeding values (**GEBV**), offering a more promising strategy to address the challenges posed by this disease. Therefore, in this study, accuracy and bias of GEBV for subclinical mastitis (**SCM**; defined in six alternative ways), and somatic cell score (**SCS**) in early lactation from the ssGBLUP approach were assessed in Canadian Holstein heifers. Two datasets, named full and truncated datasets, were created, which contained records up to 2021 and 2016, respectively. Genetic parameters were estimated using a linear animal model. Validation reliability and prediction bias of GEBV were estimated using ssGBLUP and were compared to the estimated breeding value (**EBV**) derived from traditional BLUP. For constructing the H matrix used in the ssGBLUP, different scaling factors were tested for combining the genomic and pedigree relationships. The incidence of 6 SCM trait definitions within 5-30 DIM ranged from 15.32 to 24.71%. Heritability of 6 SCM traits and SCS were 0.047 to 0.069 and 0.102, respectively. Application of the ssGBLUP model resulted in a substantial increase in validation reliabilities of

GEBV of young animals, exhibiting an average gain of 0.28 (SCM traits) and 0.19 (SCS) points when best combination of scaling factors were used. Furthermore, in comparison to EBVs obtained through the traditional BLUP method, the ssGBLUP model demonstrated slightly reduced bias in GEBV (overall with different scaling factors). Moreover, in terms of average theoretical reliabilities, gains of 0.22 and 0.27, and 0.20 and 0.28 were observed, respectively, for full and truncated datasets of SCM traits and SCS, respectively. These results effectively illustrate that ssGBLUP produces GEBV with increased reliability and less bias for young animals compared to EBV from conventional BLUP approach. Hence, implementation of ssGBLUP in routine evaluation of mastitis needs to be considered within the context of Canadian dairy industry.

6.2. Introduction

Good udder health is important, not only for dairy producers but also to consumers and the dairy industry as a whole. Mastitis is the most frequent and costliest disease impacting cows' performance, longevity, and welfare and, thus, the profitability of farms. The cost associated with a single case of clinical mastitis (**CM**) in a typical Canadian farm was CAD$662/cow per year (Aghamohammadi et al., 2018). Mastitis is caused by intramammary bacterial infection, which can lead to subclinical (**SCM**) and CM. During SCM, cows do not present physical irregularities but have reduced milk yield and undesirable changes in milk composition (Heringstad et al., 2000). Subclinical mastitis is generally defined based on an elevated SCC using a variety of cut-offs (Naqvi et al., 2018; Narayana et al., 2018).

Prevalence of SCM during the early lactation period is high, reported to be 13-24% during 5 to 30 DIM in the first lactation heifers in Canada (Narayana et al., 2018; Naqvi et al., 2018). During early lactation, cows undergo drastic physiological, metabolic, immunological, and nutritional changes, which render cows more susceptible to diseases (Mulligan and Doherty, 2008; Sordillo et al., 2009). Moreover, the cumulative gross profit of primiparous heifers with CM in the transition period (1–21 DIM) and early lactation (22–100 DIM) were CAD$ -577 and CAD$ -890, respectively (Puerto et al., 2021). Considering the impact of mastitis on the animal welfare and farm profitability, prevention and management of mastitis in early lactation of heifers is critical. Hence, the genetic selection of heifers resistant to mastitis in early lactation is desirable, as it is cumulative, permanent, and cost-effective.

As the heritability of mastitis traits, such as SCM and SCS, ranges between low to moderate (0.03 to 0.18; Svendsen and Heringstad, 2006, Narayana et al., 2018; Koeck et al., 2012a), genetic progress with traditional dairy cattle breeding methods is relatively slow. As an alternative, genomic selection provides an excellent opportunity to accelerate the genetic progress for mastitis resistance. Genomic selection involves enhancing the genetics of animals by selecting them based on evaluations performed using genomic information in addition to pedigree and performance data. The impact of genomic selection on mastitis resistance was demonstrated by an increase in genetic gain of 0.46 genetic SD units when comparing the past 5 years to the 5 years before the introduction of genomic selection (Lactanet, 2019). Combining phenotypes, genotypes, and pedigree improves the accuracy of genomic prediction, which combined with decreased generation interval leads to substantially enhanced genetic progress (Schaeffer, 2006). However, the accuracy of genomic estimated breeding value (**GEBV**) from which individuals are selected depends on the amount of information derived from genotyped and phenotyped animals. In many countries, genomic evaluations are currently estimated using the multi-step approach. This approach is complex and consists of many steps: 1) traditional best linear unbiased prediction (BLUP) evaluation using pedigree and phenotypic information; 2) de-regression of estimated breeding value (**EBV**)

(Garrick et al., 2009); 3) prediction of direct genomic value (**DGV**) for genotyped animals; and 4) blending of EBV or parent average (**PA**) with DGV to obtain GEBV (Van Doormaal et al., 2009). A multi-step approach can suffer from biases and errors introduced during de-regression process and double counting of information in the genetic evaluation (Legarra et al., 2014; Oliveira et al., 2019).

To overcome limitations inherent to a multi-step approach, an alternative methodology called single-step GBLUP (ssGBLUP) was introduced (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). The ssGBLUP methodology allows inclusion of genotyped and non-genotyped animals simultaneously in the evaluation through a hybrid relationship matrix (**H**). The **H** matrix combines the pedigree (**A**) and the genomic (**G**) relationship matrices, propagating the genomic relationships between genotyped and non-genotyped animals (Legarra et al., 2009). Moreover, ssGBLUP easily accommodates threshold and multiple trait models (Tiezzi et al., 2015; Oliveira et al., 2019). Compared to the multi-step genomic prediction method, ssGBLUP is simpler and avoids pre-selection bias (Vitezica et al., 2011; VanRaden and Wright, 2013; Legarra et al., 2014). Furthermore, ssGBLUP is beneficial for low-heritability traits and novel traits with limited number of genotyped animals relative to the number of animals with phenotypic information (Guarini et al., 2018). Two studies (Jamrozik et al., 2019; Oliveira et al., 2019) dealt with application ssGBLUP to mastitis-related traits such, i.e., CM and SCS, respectively. However, the ssGBLUP approach has not yet been explored for mastitis resistance in the early lactation of heifers. The objective of this study was, therefore, to assess the accuracy and bias of genomic predictions for alternative SCM trait definitions and SCS in early lactation (5-30 DIM) Canadian Holstein dairy heifers using the ssGBLUP approach.

6.3. Materials and Methods

6.3.1. Phenotypic, pedigree, and genotypic data

Phenotypic data on SCC, pedigree, and genotypic data of Holstein cattle were obtained from the Canadian Dairy Network (a member of Lactanet, Guelph, ON, Canada). The initial SCC data contained 115,346,326 records of 5,347,101 cows from 19,793 herds from the years 1988 to 2021. Somatic cell count was converted to SCS to approximate a normal Gaussian distribution using the following formula: SCS = $\log_2$(SCC/100,000) + 3 (Shook, 1993). Thereafter, outliers (mean SCS ± 3 SD) and missing values were removed. Only first lactation cows with age at calving between 18 and 40 months, and only first test-day records 5 ≤ DIM ≤ 30 were considered for further analyses. In addition, herds with ≤ 5 heifers in a 365-day period were removed. The final dataset contained SCS records of 3,042,014 Holstein first-lactation heifers from 15,671 herds. Subsequently, 6 alternate SCM traits (binary traits: 0: healthy and 1: presence of SCM) were defined based on a SCC threshold, following Narayana et al. (2018). In summary, for the first 3 SCM traits, heifers were considered to have SCM if their SCC was > 150,000, 200,000 and 250,000 cells/mL any time in the period 5 ≤ DIM ≤ 30 (**$\mathbf{SCM_{150}}$**, **$\mathbf{SCM_{200}}$**, and **$\mathbf{SCM_{250}}$**, respectively). For the other 3 SCM traits, heifers were considered to have SCM if SCC was >400,000 cells/mL within 5 ≤ DIM ≤ 10 and SCC was >150,000, 200,000, and 250,000 cells/mL (namely, **$\mathbf{SCM_{400,150}}$**, **$\mathbf{SCM_{400,200}}$**, and **$\mathbf{SCM_{400,250}}$**), respectively, any time within 11 ≤ DIM ≤ 30 to account for the dilution effect. Immediately after calving SCC is high regardless of infection due to low milk production; however, SCC in milk declines rapidly if there is no infection, called dilution effect (Schepers et al., 1997; Narayana et al., 2018). Hence, in this study, traits $SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$ were defined to accommodate for the dilution effect, where these traits incorporate a higher threshold value of 400,000 cells/mL within 5 to 10 DIM and slightly lower values from 11 to 30 DIM (150,000, 200,000, and 250,000 cells/mL). For the SCM traits, any herd × year with no variability in the binary trait was removed. The final 6 SCM trait datasets contained records of 2,516,243 to 2,853,637 heifers from 14,759 to 15,281 herds. However, due to the limitation imposed by freely available BLUPF90+ software (Misztal et al., 2014) and to reduce the computational demand, only a random sample of 20% of the herds for SCS and SCM traits were used for the genomic prediction (Table 6-1). Hence, the final full dataset used for the genomic analysis contained records of 498,216 to 567,837 and 605,388 heifers (herds: 2,952 to 3,056) for SCM traits and SCS, respectively. The respective pedigree was trimmed using the pedigree package (Coster, 2013) available in R version 3.6.3 (R Core Team, 2023). The pedigree file contained 1,221,663 to 1,145,906 and 1,281,771 animals for SCM traits and SCS, respectively, generated tracing back 10 generations. The full data set was used to estimate EBVs and GEBVs using traditional BLUP and ssGBLUP, respectively.

6.3.1.1. Quality control of genotypic data

A total of 11,407 to 11,872 (for the 6 SCM traits), and 11,551 (SCS) animals were genotyped. The positions of the SNPs were based on the bovine assembly ARS-UCD 1.2 (Medrano, 2017; Rosen et al., 2020). The preGSF90 software (Aguilar et al., 2014; Misztal et al., 2014) was used for filtering duplicated genotypes and SNPs with Mendelian conflicts, call rate < 0.90, minor allele frequency < 0.05, and with difference between observed and expected heterozygosity > 0.15 (Wiggans et al., 2009). After quality control, 40,669 to 40,744, and 40,395 autosomal SNPS of 11,374 to 11,828, and 11,475 animals were retained for further analyses for 6 SCM traits and SCS, respectively (Table 6-1).

6.3.1.2. Truncated dataset

A truncated dataset was created from the full dataset of 6 SCM traits and SCS by excluding the last five years of records (Table 6-1). Hence, the truncated dataset for 6 SCM traits and SCS only contained records until 2016. The truncated dataset was used for the prediction of estimated breeding values (EBVs) using traditional BLUP and GEBVs using ssGBLUP (for validation animals). Predicted EBVs and GEBVs from truncated dataset were validated (by assessing validation reliability and genomic prediction bias: described in a later section) with EBVs and GEBVs predicted from full dataset using traditional BLUP and ssGBLUP, respectively.

6.3.1.3. Validation population

A forward validation approach was utilized. Validation bulls were defined as genotyped bulls with ≥ 20 daughters with records in the full dataset, but no daughters in the truncated dataset. Validation cows were defined as genotyped cows with no records in the truncated dataset, but with records in the whole dataset. Both bulls' and cows' theoretical reliability of EBVs estimated using the full dataset were ≥ 0.40 and ≥ 0.50 for SCM traits and SCS, respectively. Validation was performed by combining both bull and cow validation populations (Table 6-1). Any genotyped animal not included as the validation population was retained as part of the training population.

6.3.2. Statistical analyses

6.3.2.1. Variance components estimation

To reduce the computational demand, for the estimation of variance components of 6 SCM traits and SCS a random sample of 5% of the initial number of herds of the final dataset was used (Table 6-2). Variance components were estimated using the Average-Information Restricted Maximum Likelihood (AIREML) procedure using the AIREMLF90 software (Misztal et al., 2014), including only pedigree and phenotypic information. The following univariate linear model was applied to SCM traits and SCS:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_{hy}hy} + \mathbf{Z_a a} + \mathbf{e},$$

where **y** is the vector of observations (0 and 1 for SCM traits or the scores for SCS); **b** is the vector of systematic effects, including fixed effects of age at first calving and year-season of calving; **hy** is a vector of random herd-year of calving effects; **a** is a vector of random animal additive genetic effects; **e** is a vector of random residuals; and **X**, $\mathbf{Z_{hy}}$, and $\mathbf{Z_a}$ are corresponding incidence matrices.

The expectation and (co)variances for the model were:

$$\mathrm{E}[\mathbf{y}] = \mathbf{Xb}$$

and

$$\mathrm{Var}\begin{bmatrix}\mathbf{hy}\\ \mathbf{a}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{I}\sigma^2_{\mathrm{hy}} & 0 & 0\\ 0 & \mathbf{A}\sigma^2_{\mathrm{a}} & 0\\ 0 & 0 & \mathbf{I}\sigma^2_{\mathrm{e}}\end{bmatrix},$$

where $\sigma^2_{\mathrm{hy}}, \sigma^2_{\mathrm{a}},$ and σ^2_{e} are the herd-year of calving, animal additive genetic, and residual variances, respectively. **I** and **A** are an identity matrix and the additive relationship matrix, respectively.

Across herd heritability (h^2) for each trait was calculated from univariate analysis as follows:

$$h^2 = \sigma^2_{\mathrm{a}} / (\sigma^2_{\mathrm{hy}} + \sigma^2_{\mathrm{a}} + \sigma^2_{\mathrm{e}})$$

6.3.2.2. Breeding value estimation

Estimated variance components were used for predictions of EBVs and GEBVs using the BLUPF90+ software (Misztal et al., 2014). The same linear model as above was used for the prediction of EBVs of full and truncated dataset using traditional BLUP for all traits. Likewise, GEBVs were predicted with the same model using the ssGBLUP approach (Aguilar et al., 2010; Christensen and Lund, 2010), based on the full and truncated datasets.

The single-step genomic prediction approach uses information from both genotyped and non-genotyped animals simultaneously. In ssGBLUP, the hybrid **H** matrix was used instead of the **A** matrix, which combines the pedigree and genomic information.

The inverse of **H** matrix was obtained as:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(0.95\boldsymbol{G} + -0.05\boldsymbol{A}_{22})^{-1} - \omega\boldsymbol{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix (in which **G** was constructed as in VanRaden (2008), first method), $\mathbf{A}^{-1}$ is the inverse of the pedigree relationship matrix created using 10 generation of pedigree and accounting for inbreeding, and $\mathbf{A}_{22}^{-1}$ is the inverse numerator relationship matrix of the section of the **A** matrix for genotyped animals. To avoid singularity problems and to account for polygenic effects, a blending factor of 0.95 (α) was used to blend **G** with 0.05 (β) of $\mathbf{A}_{22}$. Inbreeding was considered when both relationship matrices **A** and **G** were constructed. Scaling factors τ and ω were used to make $\mathbf{G}^{-1}$ compatible with $\mathbf{A}_{22}^{-1}$ and $\mathbf{A}^{-1}$. Scaling factors τ and ω account for the reduced genetic variance and different pedigree depth, respectively. Fine-tuning of τ and ω aids in avoiding convergence issues by making **G** compatible with $\mathbf{A}_{22}$ and gives higher reliability and lower biases of GEBVs (Misztal et al., 2013). Different τ (0.5, 1.0 and

1.5) and ω (0.6, 0.7, 0.8, 0.9 and 1.0) options were analyzed and best combination were chosen based on higher validation reliabilities and lower biases.

6.3.2.3. Theoretical reliability

Average theoretical reliability (*Rel*) (Aguilar et al., 2020) was estimated for EBV and GEBV predicted with full dataset ($(G)EBV_{full}$) and truncated dataset ($(G)EBV_{trun}$) using the BLUPF90+ software as follows:

$$Rel_i = 1 - \frac{PEV_i}{(1 + F_i)\sigma_a^2},$$

where Rel_i is the reliability of individual, PEV_i is the prediction error variance (square of the standard error of EBV for animal i), F_i is the inbreeding coefficient of animal i, and σ_a^2 is the additive genetic variance.

6.3.2.4. Validation reliability and bias

Validation reliability (*Val_Rel*) was estimated using the method described in Legarra and Reverter (2018):

$$Val\_Rel_i = \frac{COV[(G)EBV_{i\,full}, (G)EBV_{i\,trun}]}{(1 - \bar{F})\sigma_{a_i}^2},$$

where $COV\big[(G)EBV_{full}, (G)EBV_{trun}\big]$ is the covariance between EBV or GEBV predicted with full ($(G)EBV_{full}$) and truncated datasets ($(G)EBV_{trun}$), for the validation animals for the *i*th trait,

$\bar{F}$ is the mean inbreeding coefficient of validation animals, and $\sigma^2_{a_i}$ is the additive genetic variance for the i^{th} trait. Expected observed reliability for the validation population was estimated by dividing validation reliability of $(G)EBV$ by the respective theoretical reliability estimated using the full dataset ($(G)EBV_{full}$).

Genomic prediction bias (i.e., inflation or deflation of breeding values predicted from the truncated dataset ($(G)EBV_{trun}$) relative to the full dataset ($(G)EBV_{full}$)) of validation animals was estimated as regression coefficient (b_1) of the linear regression of $(G)EBV_{full}$ on $(G)EBV_{trun}$:

$$(G)EBV_{full} = b_0 + b_1(G)EBV_{trun},$$

where $(G)EBV_{full}$ is the EBV/GEBV predicted using full dataset, $(G)EBV_{trun}$ is the EBV/GEBV predicted using truncated dataset, b_0 is the intercept and b_1 is the genomic prediction bias (close to 1 is desirable).

6.4. Results

6.4.1. Descriptive statistics

Prevalence of 6 alternate SCM traits ranged between 16.70 to 24.71% (Table 6-1). Among the first 3 SCM traits (SCM_{150}, SCM_{200}, and SCM_{250}), which were defined based on single threshold values of 150K, 200K and 250K cells/mL from 5 to 30 DIM, SCM_{150} had the highest prevalence of 24.71% followed by SCM_{200} (19.67%) and SCM_{250} (16.70%). Similarly, among the other 3 SCM traits ($SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$), which accounted for the dilution effect, $SCM_{400,150}$ had the highest prevalence of 20.56% followed by $SCM_{400,200}$ (17.45%) and $SCM_{400,250}$ (15.32%). Mean (SD) SCS was 2.53 (1.71).

6.4.2. Genetic parameters

In general, heritability estimates for SCM traits were < 0.07 (Table 6-2). For the first 3 SCM traits (SCM_{150}, SCM_{200}, and SCM_{250}) heritability estimates ranged from 0.051 to 0.069 (± 0.004 to 0.005). Heritability estimates decreased with the increase of threshold of SCC value and decrease of incidence of the respective trait. Likewise, for the other SCM traits ($SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$), heritability estimates ranged from 0.047 to 0.055 (± 0.004 to 0.005). Heritability for SCS was estimated at 0.102 ± 0.005.

6.4.3. Validation reliability and prediction bias

Overall, for all traits, validation reliabilities of GEBVs estimated from ssGBLUP with best combination of scaling factors of τ =0.5 and ω =1.0 were higher than validation reliabilities of EBVs (Table 6-3). On average (for all traits) there was a gain of 0.27 in validation reliability by including genomic information with the ssGBLUP method. For all 6 SCM traits, validation reliabilities of GEBVs (ranging from 0.28 to 0.40) were higher than those from EBVs (ranging from 0.03 to 0.07) with a gain of validation reliabilities ranging from 0.24 to 0.30. For SCS, validation reliabilities gain was 0.19, where validation reliability of EBVs and GEBVs were 0.08 and 0.27, respectively. Expected observed reliability for EBV and GEBV ranged from 0.06 to 0.15 and 0.34 to 0.55, respectively, with a gain ranging from 0.21 to 0.40 with best combination of scaling factors.

The prediction bias (regression coefficient) of EBV estimated using traditional BLUP and GEBV estimated from ssGBLUP deviated from 1 for all traits (Table 6-3). The prediction bias of

GEBV estimated from ssGBLUP with best combination of scaling factors (τ =0.5 and ω =1.0) was lower for SCM_{200}, SCM_{250} and $SCM_{400,150}$ (ranged from 0.77 ± 0.01 to 0.82 ± 0.02) compared to the prediction bias of EBV estimated using traditional BLUP (ranged from 0.64 ± 0.04 to 0.80 ± 0.04). However, for traits SCM_{150}, $SCM_{400,200}$, $SCM_{400,250}$ and SCS, regression coefficients deviated more from 1 for GEBVs estimated from ssGBLUP (ranging from 0.68 ± 0.02 to 0.86 ± 0.02) compared to the prediction bias of EBV estimated using traditional BLUP (ranging from 0.78 ± 0.04 to 0.97 ± 0.05).

Regarding factors τ and ω which were used to scale the $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ matrices, validation reliabilities of GEBVs estimated from ssGBLUP using all scaling factor options were higher than the validation reliabilities of EBVs (Supplementary material Table S1 to S8; https://doi.org/10.7910/DVN/YEBRV4; Narayana et al., 2023). Generally, with constant scaling factor of τ (0.5, 1.0 and 1.5) and increasing ω (0.6 to 1.0), validation reliabilities of GEBV increased for all traits. However, when the scaling factor of τ increased from 0.5 to 1.5, validation reliability of GEBV decreased. Moreover, even using the default scaling factor (τ =1.0 and ω =1.0), validation reliabilities of GEBVs were higher compared to EBVs estimated using traditional BLUP, which resulted in gains of 0.19, 0.20, 0.17, 0.20, 0.15, 0.12 and 0.13 reliability points for SCM_{150}, SCM_{200}, SCM_{250}, $SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$ and SCS, respectively. Regarding the prediction bias, when the ω factor increased from 0.6 to 1.0, regression coefficient deviated more from 1, resulting in more biased of GEBV prediction for all traits. An opposite trend was observed for the τ parameter when it was increased from 0.5 to 1.5, where regression coefficient deviated less from 1 resulting in less biased GEBV prediction for all traits, except $SCM_{400,250}$. Best combination of scaling factors were chosen based on the highest reliabilities and the least biased predictions.

6.4.4. Theoretical Reliabilities

In general, a substantial gain was observed in average theoretical reliabilities of GEBVs estimated from ssGBLUP with best combination of scaling factors (τ = 0.5 and ω = 1.0) compared to EBVs for validation population for full and truncated datasets for all traits (Table 6-4). For the truncated dataset of the 6 SCM traits, average theoretical reliabilities of GEBVs ranged from 0.57 to 0.60, whereas the average theoretical reliabilities of EBVs estimated from traditional BLUP ranged from 0.30 to 0.32 for the 6 SCM traits. Gain in average theoretical reliabilities of GEBVs relative to EBVs ranged from 0.25 to 0.28, with an average of 0.27 for the 6 SCM traits. For the full data, average theoretical reliabilities of GEBVs ranged from 0.71 to 0.74, whereas the average theoretical reliabilities of EBVs estimated from traditional BLUP ranged from 0.50 to 0.52 for the 6 SCM traits. Gain in average theoretical reliabilities of GEBVs relative to EBVs ranged from 0.21 to 0.23, with an average of 0.22 for the 6 SCM traits. For SCS, average theoretical reliabilities full and truncated dataset were 0.60 and 0.80; 0.32 and 0.60, respectively for EBVs and GEBVs. Gain in average theoretical reliabilities of GEBVs relative to EBVs were 0.20 and 0.28 for the full and truncated dataset, respectively. When the scaling factor ω increased with constant τ, theoretical reliabilities of GEBVs increased for all traits. However, theoretical reliabilities of GEBVs decreased for all traits when the τ increased with constant ω. Overall, average theoretical reliabilities of EBV and GEBV of truncated data (young animals without records) estimated from both traditional BLUP and ssGBLUP were lower compared to EBV and GEBV of full data for all traits.

6.5. Discussion

The present study delved into the prospective advantages of utilizing ssGBLUP for traits associated with SCM and SCS in early lactation of heifers. Genomic prediction of both SCM and SCS, employing varying scaling factors within the ssGBLUP framework, particularly during the early stages of lactation, remains an area that has not been extensively explored. In alignment with the literature (De Vliegher et al., 2012; Naqvi et al., 2018), in the present study, estimation of incidence of the 6 SCM traits in early lactation ranged between 15 to 25% with different threshold values of SCC. The main reason for the elevated SCC in early lactation is IMI (Harmon, 1994; Laevens et al., 1997; Barkema et al., 1999). The IMI is often already acquired before calving (Green et al., 2002). The early lactation is a stressful critical phase in the life of dairy cows, when cows go through physiological, metabolic, and immunological changes when they transition from the non-lactating to lactating phase. This period is critical for the health, fertility, and productivity of dairy cows (Leblanc, 2010; Horst et., 2021). Moreover, during early lactation cows undergo a negative energy balance, the imbalance between nutrient supply and demand, which leads to impaired immune functions (Nyman et al., 2008; Wathes et al., 2009; Esposito et al., 2014) that leaves cows at a high risk of developing health disorders, and an IMI can turn into SCM or even clinical mastitis. During the transition period, nearly 30-50% of cows are affected by some form of infectious or metabolic disease (Leblanc, 2010). Hence, breeding for heifers that are resilient during early lactation is paramount.

Heritability estimates of the 6 SCM traits (0.069 ± 0.005 to 0.047 ± 0.004) were in line with previously reported studies (Koivula et al., 2004; Svendsen and Heringstad, 2006; De Haas et al., 2008; Urioste et al., 2012; Narayana et al., 2018), and as expected followed a declining trend with increased threshold of SCC. As heritability estimation using a linear model is prevalence dependent, it decreased with increasing threshold of SCC used to define SCM traits, because the prevalence was lower at higher thresholds. Likewise, the heritability estimate of SCS (0.102 ± 0.05) in the present study aligned with reported values (Koeck et al., 2012b; Oliveira et al., 2019) for heifers. The heritability of mean SCS in both early (5 to 150 DIM) and late lactation (151 to 305 DIM) was previously reported to be 0.11 in Canadian Holsteins (Koeck et al., 2012b). However, it is important to note that in our study only the first test-day records of SCS between 5-30 DIM were considered.

Validation reliabilities of EBV estimated using traditional BLUP were low (≤ 0.08) for all analyzed traits. The main reason for the low reliabilities of EBV estimated using traditional BLUP for all analyzed traits is the low heritability of these traits. In addition, only one record per animal within 5-30 DIM was considered for the analysis in the present study. To the authors' knowledge no literature is published exploring the validation reliability of SCM. High gains (0.24 to 0.32) of validation reliability of GEBV using ssGBLUP compared to EBV provide evidence that ssGBLUP is a better alternative for low heritability traits such as SCM, where GEBV is estimated directly by combining all animals sources of information (phenotypes, genotypes and pedigree) and all available animals with records (genotyped and non-genotyped).

Buaban et al. (2021) reported low parent average (**PA**) validation reliability for SCS of 0.05 (accuracy 0.22) in first-lactation (5-305 DIM) Holsteins using a random regression model, whereas in the current study, validation reliability of EBV of SCS was estimated to be 0.08. Higher validation reliability of EBV of SCS in the current study was likely due to a higher number of animals in the validation population (929; cow and bulls) relative to only 13 young bulls in Buaban et al. (2021). Moreover, these authors also reported validation reliability of 0.12 and 0.15 (validation accuracy 0.35 and 0.39, respectively) for GEBV by including genotypes of bulls only and of bulls and cows in the random regression ssGBLUP model, respectively. Gain in validation reliability of SCS in our study (0.19) was higher than reported by Buaban et al. (2021), who reported 0.07 and 0.10 for including genotypes of bulls only and of bulls and cows in the random regression ssGBLUP model, respectively. Oliveira et al. (2019) reported a higher validation reliability of 0.20 for PA of SCS in first lactation (5-305 DIM) Holsteins using a random regression model. They also reported a higher validation reliability of 0.44 for GEBV estimated with ssGBLUP with best combination of scaling factors (τ =2.0 and ω =0.6), which gave a 1.26-fold gain relative to present study. This might be because Oliveira et al. (2019) used more data per animal (5-305 DIM, 1st lactation), model differences, and high reliability of validation bulls ($\geq$ 0.80). It is important to note that in the current study only first test day records between 5-30 DIM were used and the theoretical reliability of validation animal was on average 0.60.

The unbiasedness of genetic and genomic predictions was evaluated by the coefficient of the linear regression of (G)EBV$_{full}$ on (G)EBV$_{trun}$. Generally, for all traits with different scaling factors (except SCM$_{400,200}$, and SCM$_{400,250}$) regression coefficient of both EBV and GEBV were $<$ 1, which suggests that the genetic and genomic predictions are inflated and would result in up-ward biased estimation of differences in genetic merit of young animals compared to their actual future performances. Several studies have reported inflation in GEBV in cases where young genotyped dairy bulls do not have any daughters with phenotypes for different traits (Aguilar et al., 2010; Tsuruta et al., 2011). Potential reasons for this bias are the use of animals with low reliability of EBV for estimating the bias, incomplete pedigree (Misztal et al., 2017; Tsuruta et al., 2019), incompatibility between **A** and **G** matrix (Tsuruta et al., 2019), small dataset (Legarra and Reverter, 2018), and more likely due to pre-selection of the young genotyped bulls (Vitezica et al., 2011), which is a known cause of inflation of EBV from traditional genetic evaluations.

As scaling factors influence the reliability, bias, and convergence, they need to be wisely defined to optimize the GEBV prediction (Martini et al., 2018; Tsuruta et al., 2011; 2019). Scaling factors are impacted by population, trait, and model (Tsuruta et al., 2011; Tsuruta et al., 2019; Lourenco et al., 2020). Generally, reduction in bias of GEBV is expected when τ increases and ω decrease by decreasing the variance of GEBV (Martini et al., 2018). This was observed in the current study for almost all traits. Tsuruta et al. (2011; 2019) reported that when scaling factor ω is <1.0, inflation of GEBV was reduced and convergence improved for 18 type traits using multiple trait ssGBLUP models in U.S. Holsteins. In the current study, a similar pattern of reduced inflation of GEBV was observed when ω decreased from 1 to 0.6 keeping τ constant; however, on the other hand, reliabilities reduced for all traits. Oliveira et al. (2019) reported similar trend of decreasing inflation of GEBV for SCS with decreasing scaling factor ω in first lactation Holsteins. Regarding validation reliability, Oliveira et al. (2019) reported higher validation reliability when τ >1 for SCS in Holsteins. Conversely, in our study validation reliability reduced for all traits when τ increased maintaining constant ω. This might be because validation reliability is data, model, and population specific (Martini et al., 2018), and influenced by genetic parameters of the trait.

Theoretical reliability is an indicator of selection accuracy. By comparing validation animals' average theoretical reliability of EBV estimated by traditional BLUP evaluation to GEBV estimated by ssGBLUP, one can gain insight into the potential advantages genomic selection could offer. In the current study, a substantial increase in theoretical reliability was observed when genomic information was included through ssGBLUP in addition to phenotype and pedigree information, suggesting major advantage of genomic selection for low heritability traits like SCM. Genomic reliability is influenced by relationship of the selection candidates with reference population (animals with genotypes and phenotypes) (Wiggans and VanRaden, 2010), and heritability of the trait (Misztal et al., 2013; Buaban et al., 2021). In agreement with our findings for the truncated data for SCS, Buaban et al. (2021) reported a gain of 0.28 for GEBV estimated by ssGBLUP compared to PA in first lactation Thai Holsteins. Moreover, Buaban et al. (2021) reported a higher gain for low-heritability traits (e.g., SCS) than high-heritability traits (e.g., production traits) by including genomic information.

6.6. Conclusions

This study unveiled the significant potential for genetic improvement of SCM and SCS in first lactation heifers through genomic selection based on ssGBLUP evaluation, as an alternative to traditional BLUP evaluation. The single-step methodology is expected to be beneficial for low-heritability, as it is the case for mastitis-associated traits. The outcomes of this study revealed a substantial increase in validation and theoretical reliabilities, accompanied by minimized GEBV bias using ssGBLUP compared to traditional BLUP. Notably, these advancements are observed in the context of the first lactation of Holstein cows during their early lactation stage. Consequently, adopting the ssGBLUP methodology for genetic evaluation of mastitis-related traits should be considered by the Canadian dairy industry. Nevertheless, careful consideration must be given to selecting appropriate scaling factors to ensure optimal implementation and success.

6.7. References

Aghamohammadi, M., D. Haine, D. F. Kelton, H. W. Barkema, H. Hogeveen, G. P. Keefe, and S. Dufour. 2018. Herd-level mastitis-associated costs on Canadian dairy farms. Front. Vet. Sci. 5:100.

Aguilar, I., E.N. Fernandez, A. Blasco, O. Ravagnolo, and A. Legarra. 2020. Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. J. Anim. Breed. Genet. 137:356-364.

Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

Aguilar, I., I. Misztal, S. Tsuruta, and A. Legarra. 2014. PREGSF90-POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs Proc. 10th World Congr. Genet. Appl. to Livest. Prod., Vancouver, BC, Canada. Pp.1–3.

Barkema, H. W., H. A. Deluyker, Y. H. Schukken, and T.J.G.M. Lam. 1999. Quarter-milk somatic cell count at calving and at the first six milkings after calving. Prev. Vet. Med. 38:1–9.

Buaban, S., S. Prempree, P. Sumreddee, M. Duangjinda, and Y. Masuda. 2021. Genomic prediction of milk-production traits and somatic cell score using single-step genomic best linear unbiased predictor with random regression test-day model in Thai dairy cattle. J. Dairy Sci. 104:12713–12723.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2.

Coster, A. 2013. pedigree: Pedigree functions. Accessed October 10, 2023. https://cran.r-project.org/web/packages/pedigree/pedigree.pdf.

De Haas, Y., W. Ouweltjes, J. Ten Napel, J. J. Windig, and G. De Jong. 2008. Alternative somatic cell count traits as mastitis indicators for genetic selection. J. Dairy Sci. 91:2501–2511.

De Vliegher, S., L. K. Fox, S. Piepers, S. McDougall, and H. W. Barkema. 2012. Invited review: Mastitis in dairy heifers: Nature of the disease, potential impact, prevention, and control. J. Dairy Sci. 95:1025–1040.

Esposito, G., P. C. Irons, E. C. Webb, and A. Chapwanya. 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. Anim. Reprod. Sci. 144:60–71.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55.

Guarini, A. R. 2018. Combining genotypic, phenotypic and pedigree information to analyze functional traits in dairy cattle. PhD Thesis, University of Guelph, Guelph, ON, Canada.

Harmon, R. J. 1994. Physiology of mastitis and factors affecting somatic cell counts. J. Dairy Sci. 77:2103–2112.

Heringstad, B., G. Klemetsdal, and J. Ruan. 2000. Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. Livest. Prod. Sci. 64:95–106.

Horst, E., S. Kvidera, and L. Baumgard. 2021. Invited review: The influence of immune activation on transition cow health and performance—A critical evaluation of traditional dogmas. J. Dairy Sci. 104:8380–8410.

Jamrozik, J., A. Koeck, F. Miglior, G. Kistemaker, F. Schenkel, D. Kelton, and B. van Doormaal. 2013. Genetic and genomic evaluation of mastitis resistance in Canada. Interbull Bull. 47:43–51.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012a. Health recording in Canadian Holsteins: Data and genetic parameters. J. Dairy Sci. 95:4099–4108.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012b. Alternative somatic cell count traits to improve mastitis resistance in Canadian Holsteins. J. Dairy Sci. 95:432–439.

Koivula, M., E. Negussie, and E. A. Mäntysaari. 2004. Genetic parameters for test-day somatic cell count at different lactation stages of Finnish dairy cattle. Livest. Prod. Sci. 90:145–157.

Lactanet. 2019. 10 Years of genomic selection: What's next? Accessed October 10, 2023. https://www.cdn.ca/document.php?id=530

Laevens, H., H. Deluyker, Y. H. Schukken, L. De Meulemeester, R. Vandermeersch, E. De Muelenaere, and A. De Kruif. 1997. Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. J. Dairy Sci. 80:3219–3226.

LeBlanc, S. 2010. Monitoring metabolic health of dairy cattle in the transition period. J. Reprod. Dev. 56(Suppl.): S29–S35.

Legarra, A. and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet. Sel. Evol. 50:1–18.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single step, a general approach for genomic selection. Livest. Sci. 166:54–65.

Lourenco, D., A. Legarra, S. Tsuruta, Y. Masuda, I. Aguilar, and I. Misztal. 2020. Single-step genomic evaluations from theory to practice: Using snp chips and sequence data in blupf90. Genes 11:790.

Martini, J. W., M. F. Schrauf, C. A. Garcia-Baccino, E. C. Pimen-tel, S. Munilla, A. Rogberg-Muñoz, R. J. C. Cantet, C. Reimer, N. Gao, V. Wimmer, and H. Simianer. 2018. The effect of the H−1 scaling factors τ and ω on the structure of H in the single-step procedure. Genet. Sel. Evol. 50:16.

Medrano, J. F. 2017. The new bovine reference assembly and its value for genomic research. Proc. Assoc. Advmt. Anim. Breed. Genet. 22:161–166.

Misztal, I. 2017. Studies on inflation of GEBV in single-step GBLUP for type. Interbull Bull. 51:38–42.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655.

Misztal, I., S. Tsuruta, D. A. L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. Manual for BLUPF90 family of programs. Accessed October 10, 2023. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all8.pdf

Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra, P. VanRaden, and T. Lawlor. 2013. Methods to approximate reliabilities in single-step genomic evaluation. J. Dairy Sci. 96:647–654.

Mulligan, F. J. and M. L. Doherty. 2008. Production diseases of the transition cow. Vet. J. 176:3–9.

Naqvi, S. A., J. De Buck, S. Dufour, and H. W. Barkema. 2018. Udder health in Canadian dairy heifers during early lactation. J. Dairy Sci. 101:3233–3247.

Narayana, S. G., F. Miglior, S. A. Naqvi, F. Malchiodi, P. Martin, and H. W. Barkema. 2018. Genetic analysis of subclinical mastitis in early lactation of heifers using both linear and threshold models. J. Dairy Sci. 101:11120–11131.

Nyman, A. K., U. Emanuelson, K. Holtenius, K. L. Ingvartsen, K. L. Ingvartsen, and K. P. Waller. 2008. Metabolites and immune variables associated with somatic cell counts of primiparous dairy cows. J. Dairy Sci. 91:2996–3009.

Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, J. L. F. Brito, F. F. Silva, and F. S. Schenkel. 2019. Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. J. Dairy Sci. 102:2365–2377.

Puerto, M. A., E. Shepley, R. I. Cue, D. Warner, J. Dubuc, and E. Vasseur. 2021. The hidden cost of disease: I. Impact of the first incidence of mastitis on production and economic indicators of primiparous dairy cows. J. Dairy Sci. 104:7932–7943.

R Core Team. 2023. R: A language and environment for statistical computing. Accessed October 10, 2023. https://www.r-project.org/.

Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDaneld, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M Phillippy, J. B. Cole, C. P. V. Tassell, G. Liu, T. P. L. Smith, and J. F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience. 9:giaa021.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123, 218–223.

Schepers, A. J., T.J.G.M. Lam, Y. H. Schukken, J. B. M. Wilmink, and W. J. A. Hanekamp. 1997. Estimation of variance components for somatic cell counts to determine thresholds for uninfected quarters. J. Dairy Sci. 80:1833–1840.

Sordillo, L. M., G. A. Contreras, and S. L. Aitken. 2009. Metabolic factors affecting the inflammatory response of periparturient dairy cows. Anim. Health Res. Rev. 10:53–63.

Svendsen, M. and B. Heringstad. 2006. Somatic cell count as an indicator of subclinical mastitis. Genetic parameters and correlations with clinical mastitis. Interbull Bull. 35:12–16.

Tiezzi, F., K. L. Parker-Gaddis, J. B. Cole, J. S. Clay, and C. Maltecca. 2015. A Genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PLoS One 10:e0114919.

Tsuruta, S., D. A. L. Lourenco, Y. Masuda, I. Misztal, and T. J. Lawlor. 2019. Controlling bias in genomic breeding values for young genotyped bulls. J. Dairy Sci. 102:9956–9970.

Tsuruta, S., I. Aguilar, I. Misztal, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198–4204.

Urioste, J. I., J. Franzén, J. Windig, and E. Strandberg. 2012. Genetic relationships among mastitis and alternative somatic cell count traits in the first 3 lactations of Swedish Holsteins. J. Dairy Sci. 95:3428–3434.

Van Doormaal, J., G. Kistemaker, P. G. Sullivan, M. Sargolzaei, and F. S. Schenkel. 2009. Canadian implementation of genomic evaluations. Interbull Bull. 40:214.

VanRaden, P. M. and J. R. Wright. 2013. Measuring genomic preselection in theory and in practice. Interbull Bull. 47:147–150.

Vitezica, Z., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. 93:357–366.

Wathes, D. C., Z. Cheng, W. Chowdhury, M. A. Fenwick, R. Fitzpatrick, D. G. Morris, J. Patton, and J. J. Murphy. 2009. Negative energy balance alters global gene expression and immune responses in the uterus of postpartum dairy cows. Physiol. Genomics. 39:1–13.

Wiggans, G., T. Sonstegard, P. VanRaden, L. Matukumalli, R. Schnabel, J. Taylor, F. Schenkel, and C. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92:3431–3436.

Table 6-1. Number of animals included in the full and reduced datasets, pedigree, number of genotyped animals, and prevalence of alternate subclinical mastitis (SCM) traits and mean (±SD) of somatic cell score (SCS)

Trait[1]	No. animals in full dataset (herds)	No. animals in pedigree	Prevalence in full dataset (%); mean ± SD	No. animals in reduced dataset	No. genotyped animals	No. SNPs after quality control	No. animals in validation population
SCM_{150}	567,837 (3,056)	1,221,663	24.71%	474,693	11,374	40,669	1,192
SCM_{200}	549,144 (3,012)	1,201,989	19.67%	460,241	11,541	40,744	839
SCM_{250}	523,930 (2,970)	1,172,197	16.70%	434,263	11,468	40,733	978
$SCM_{400,150}$	550,268 (3,032)	1,198,554	20.56%	460,651	11,828	40,736	1,095
$SCM_{400,200}$	514,762 (2,989)	1,160,358	17.45%	433,052	11,785	40,729	978
$SCM_{400,250}$	498,216 (2,952)	1,145,906	15.32%	415,040	11,667	40,734	854
SCS	605,388 (3,134)	1,281,771	2.53 ± 1.71	503,769	11,475	40,395	929

[1]SCS=somatic cell score. SCM_{150}, SCM_{200}, and SCM_{250} = presence of SCC >150,000, 200,000, and 250,000 cells/mL at $5 \leq DIM \leq 30$, respectively. $SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$ = presence of SCC >400,000 cells/mL at $5 \leq DIM \leq 10$ and SCC >150,000, 200,000, and 250,000 cells/mL, respectively, at $11 \leq DIM \leq 30$.

Table 6-2. Variance components and heritability (h^2) estimates and SE for alternate subclinical mastitis (SCM) traits and somatic cell score (SCS) in early lactating Canadian dairy heifers

Trait[1]	No. animals	No. herds	Genetic variance (SE)	Herd-year variance (SE)	Residual variance (SE)	h^2 (SE)
SCM_{150}	132,861	764	0.013 (0.001)	0.002 (0.000)	0.171	0.069 (0.005)
SCM_{200}	145,269	753	0.009 (0.001)	0.001 (0.000)	0.147	0.056 (0.004)
SCM_{250}	126,692	742	0.007 (0.001)	0.000 (0.000)	0.133	0.051 (0.005)
$SCM_{400,150}$	131,964	758	0.009 (0.001)	0.001 (0.000)	0.153	0.055 (0.005)
$SCM_{400,200}$	139,839	747	0.008 (0.001)	0.001 (0.000)	0.133	0.053 (0.005)
$SCM_{400,250}$	130,099	738	0.006 (0.001)	0.000 (0.000)	0.123	0.047 (0.004)
SCS	155,905	784	0.301 (0.016)	0.140 (0.005)	2.510	0.102 (0.005)

[1]SCS=somatic cell score. SCM_{150}, SCM_{200}, and SCM_{250} = presence of SCC >150,000, 200,000, and 250,000 cells/mL at $5 \leq DIM \leq 30$, respectively. $SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$ = presence of SCC >400,000 cells/mL at $5 \leq DIM \leq 10$ and SCC >150,000, 200,000, and 250,000 cells/mL, respectively, at $11 \leq DIM \leq 30$.

Table 6-3. Validation (r^2) and expected observed reliabilities ($\mathbf{r_o^2}$), and regression coefficients (b_1±SE) (i.e., prediction bias) for pedigree-based estimated breeding values (EBV) and genomic breeding value (GEBV) with best combination of scaling factors (τ=0.5 and ω=1.0) to scale $\mathbf{A_{22}^{-1}}$ and $\mathbf{G^{-1}}$ for the validation animals

	EBV			***GEBV***				
Trait[1]	r^2 (EBV_{full}-EBV_{trun})	$r^2_{o\_EBV}$	b_1±SE	r^2 ($GEBV_{full}$-$GEBV_{trun}$)	$r^2_{o\_GEBV}$	b_1±SE	Gain in r^2	Gain in r_o^2
SCM$_{150}$	0.07	0.13	0.78± 0.04	0.37	0.50	0.76 ± 0.02	0.30	0.37
SCM$_{200}$	0.08	0.15	0.80 ± 0.04	0.40	0.55	0.82± 0.02	0.32	0.40
SCM$_{250}$	0.03	0.06	0.64 ± 0.04	0.30	0.42	0.82± 0.02	0.27	0.36
SCM$_{400,150}$	0.06	0.12	0.76 ± 0.03	0.36	0.49	0.77 ± 0.01	0.30	0.37
SCM$_{400,200}$	0.04	0.08	0.90 ± 0.06	0.28	0.39	0.86 ± 0.02	0.24	0.31
SCM$_{400,250}$	0.06	0.12	0.97 ± 0.05	0.33	0.46	0.84 ± 0.02	0.27	0.34
SCS	0.08	0.13	0.86 ± 0.05	0.27	0.34	0.77± 0.03	0.19	0.21

[1]SCS=somatic cell score. SCM_{150}, SCM_{200}, and SCM_{250} = SCC >150,000, 200,000, and 250,000 cells/mL at 5 ≤ DIM ≤ 30, respectively. $SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$ = SCC >400,000 cells/mL at 5 ≤ DIM ≤ 10 and SCC >150,000, 200,000, and 250,000 cells/mL, respectively, at 11 ≤ DIM ≤ 30.

Table 6-4. Average theoretical reliability (± SD) for estimated breeding value (EBV) and genomic breeding value (GEBV) for full and reduced data with best combination of scaling factors (τ=0.5 and ω=1.0) to scale $\mathbf{A}_{22}^{-1}$ and $\mathbf{G}^{-1}$ for the validation animals

Trait[1]	EBV (full)	GEBV (full)	Gain (full)	EBV (truncated)	GEBV (truncated)	Gain (truncated)
SCM_{150}	0.52 ± 0.08	0.74 ± 0.05	0.22	0.33 ± 0.07	0.60 ± 0.05	0.27
SCM_{200}	0.52 ± 0.09	0.73 ± 0.05	0.21	0.32 ± 0.07	0.59 ± 0.05	0.27
SCM_{250}	0.50 ± 0.08	0.72 ± 0.05	0.22	0.30 ± 0.07	0.57 ± 0.05	0.27
$SCM_{400,150}$	0.50 ± 0.08	0.73 ± 0.05	0.23	0.32 ± 0.07	0.60 ± 0.05	0.28
$SCM_{400,200}$	0.50 ± 0.08	0.72 ± 0.05	0.22	0.32 ± 0.07	0.58 ± 0.05	0.26
$SCM_{400,250}$	0.50 ± 0.08	0.71 ± 0.05	0.21	0.32 ± 0.06	0.57 ± 0.05	0.25
SCS	0.60 ± 0.09	0.80 ± 0.05	0.20	0.32 ± 0.06	0.60 ± 0.05	0.28

[1]SCS=somatic cell score. SCM_{150}, SCM_{200}, and SCM_{250} = presence of SCC >150,000, 200,000, and 250,000 cells/mL at 5 ≤ DIM ≤ 30, respectively. $SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$ = presence of SCC >400,000 cells/mL at 5 ≤ DIM ≤ 10 and SCC >150,000, 200,000, and 250,000 cells/mL, respectively, at 11 ≤ DIM ≤ 30.

CHAPTER 7: Summarizing discussion

Mastitis, a persistent health concern in the dairy cattle industry over several decades, exerts a substantial impact on the sector. To put this into perspective, for a typical Canadian dairy farm, cost associated with mastitis was CAD$66,178 per 100 cows per year (Aghamohammadi et al., 2018). Multifaceted and intricated mastitis arises as a response to the invasion of primarily bacteria into the udder, a condition referred to as intramammary infection (IMI) (Berry and Meaney, 2006; Andersen et al., 2010; Constable et al., 2017). Mastitis can manifest in two forms: subclinical mastitis (SCM) and clinical mastitis (CM). Bovine mastitis is the result of a complex interplay between the host, environmental factors, and pathogens. The causative agents of mastitis can be categorized as contagious, environmental, or opportunistic based on their epidemiological and pathophysiological characteristics (Constable et al., 2017).

7.1. Prevalence

Worldwide, the prevalence of IMI is major, and it exhibits variations across different stages of lactation and parities. Prevalence of IMI ranges from 20 to 75% for individual udder quarters and from 49 to 97% for cows (Oliver et al., 2003; Sampimon et al., 2009; Piepers et al., 2011). In the case of heifers, approximately 12 to 57% of their udder quarters are affected by IMI at the time of calving (Cook et al., 1992; Parker et al., 2007). In an extensive study conducted in Belgium, of 44,677 cows 41% had at least one udder quarter that was culture-positive (Piepers et al., 2007). **Chapter 3** delved into the prevalence of IMI, specifically at the quarter level, in multiparous cows within the range of 0 to 400 DIM, considering various pathogen groups and across Canadian provinces. In this study, the overall prevalence of IMI was 45.5%. Among the IMI specific to particular pathogens, non-*aureus* staphylococci (**NAS**) had the highest quarter-level prevalence

(30.8%), followed by environmental pathogens (*Escherichia coli, Klebsiella* spp., *Enterobacter* spp., *Nocardia* spp., *Prototheca* spp., *Trueperella pyogenes, Streptococcus uberis, Bacillus* spp., *Streptococcus* spp., *Serratia* spp., *Citrobacter* spp., *Proteus* spp., *Salmonella* spp., *Pseudomonas* spp., and *Pasteurella multocida*; 14.0%) and contagious pathogens (*Staphylococcus aureus, Streptococcus dysgalactiae*, *Streptococcus agalactiae*, and *Corynebacterium* spp.; 7.5%). Minor pathogen-related IMI had a higher prevalence (*Corynebacterium spp*., NAS; 34.0%) at the quarter level in comparison to major pathogen-related IMI (*Staph. aureus, Strep. agalactiae, Strep. dysgalactiae, Streptococcus* spp., *Strep. uberis, E. coli, Klebsiella* spp., *Enterobacter* spp., or *Trueperella pyogenes*; 7.0%). Across provinces, the quarter-level prevalence of overall IMI was highest in Ontario (53.0%) herds, followed by Alberta (49.0%), Québec (45.0%), Nova Scotia (33.0%), and New Brunswick (32.0%), with dairy herds in Prince Edward Island having the lowest overall IMI prevalence (30.0%).

Somatic cells have served as a longstanding indicator of SCM, as they are released in the mammary gland to act as a defence mechanism against IMIs (Sharma et al., 2011). Subclinical mastitis, which is characterized by somatic cell count (SCC) surpassing a defined cutoff value, may not exhibit any apparent physical abnormalities (Saidi et al., 2013). However, it is associated with reduced milk production, adverse alterations in milk composition, and the presence of inflammatory components in the milk (Heringstad et al., 2000). Elevated SCC is a commonly employed method for detecting SCM (Harmon, 1994; Pyörälä, 2003). However, the specific SCC cutoff used to define SCM varies, resulting in inconsistent SCM prevalence across studies. **Chapter 2** investigated the prevalence of SCM using 6 different criteria based on distinct SCC cutoff values within the first 5 to 30 DIM in heifers. For the first 3 criteria, heifers exceeding SCC cutoff values of 150,000, 200,000, and 250,000 cells/mL within the 5 to 30 DIM range were considered to be infected, denoted as SCM_{150}, SCM_{200}, and SCM_{250}, respectively. The latter 3 criteria accounted for the dilution effect caused by the increased milk production, classifying heifers as infected if their SCC was >400,000 cells/mL within the 5 to 10 DIM range and >150,000, 200,000, and 250,000 cells/mL within the 11 to 30 DIM range, referred to as $SCM_{400,150}$, $SCM_{400,200}$, and $SCM_{400,250}$, respectively. Among the first 3 criteria, SCM_{150} had the highest overall SCM prevalence at 24.3%, followed by SCM_{200} (18.7%) and SCM_{250} (15.2%). Similarly, among the other 3 criteria, $SCM_{400,150}$ had the highest overall prevalence at 19.6%, followed by $SCM_{400,200}$ (15.9%) and $SCM_{400,250}$ (13.4%). The prevalence of SCM was highest in Ontario, ranging from 15.0 to 26.8%, and Prince Edward Island, ranging from 14.2 to 27.1%, whereas it was lowest in Alberta, ranging from 10.4 to 19.0%.

7.2. Genetics of mastitis

Despite extensive efforts in implementing management practices to control and prevent mastitis, the incidence of CM has remained persistent over time (Ruud et al., 2010; Richert et al., 2013; Bates and Dohoo, 2016). Consequently, there is a growing emphasis on prioritizing selective breeding for improved mastitis resistance, alongside the adoption of enhanced management practices, as a sustainable approach to mastitis control. The success of genetic selection is contingent on the presence of genetic variation. Although the heritability of mastitis resistance is relatively low, measuring below 0.05 (Koeck et al., 2012; Jamrozik et al., 2013), numerous studies have explored the genetic diversity of CM and have consistently concluded that while the response to selection may be gradual, sufficient genetic variability exists to support genetic selection (Heringstad et al., 2003; Koeck et al., 2012). Regarding genetic variability, **Chapters 2** and **3** explored the genetic variability of SCM and IMI, respectively. Heritability estimates (standard errors) of SCM traits ranged from 0.037 to 0.057 (0.015 to 0.018) from linear models and from

0.040 to 0.051 (0.017 to 0.020) from threshold models in **Chapter 2**. Despite the challenges posed by low heritability, examination of estimated breeding values (EBV) derived from both models revealed the presence of exploitable genetic variation among sires. Using a linear model that considered all 6 SCM traits, the percentage of daughters with SCM ranged from 5 to 13% for the top 10% of 69 sires, and from 19 to 33% for the worst 10% of sires, each having a minimum of 20 daughters across at least 5 herds.

In **Chapter 3**, we determined the posterior mean of quarter heritability estimates (posterior standard deviations) for overall IMI and pathogen-specific IMI traits. These estimates ranged from 0.017 to 0.073 (±0.009 to 0.030). Percentage of daughters with IMI attributed to various pathogen groups spanned from 13 to 80% for the top 10% of sires and from 38 to 94% for the bottom 10% of sires, as categorized by their pathogen-specific IMI group estimated breeding values. Furthermore, the findings indicated that selecting for lower somatic cell score (SCS) can enhance resistance to IMI. However, it is important to note that there were only weak to moderate rank correlations (ranging from 0.04 to 0.47) between pathogen-specific IMI traits and SCS, suggesting that selecting for lower SCS may not uniformly improve resistance to IMI caused by different pathogen-specific IMI groups.

7.3. Impact of genomics on mastitis

Genomic information can be used for genomic prediction (GP) and can help uncover the genetic architecture of mastitis through the identification of genomic variants and genes associated with mastitis, where this knowledge may also help to enhance the genomic selection (GS) accuracy. In comparison to traditional genetic selection methods, GS offers advantages, such as increased prediction accuracy and reduced generation interval, leading to substantial advancements in the genetic selection of traits with low heritability, such as mastitis (Parker Gaddis et al., 2014). In **Chapter 4**, the results of a systematic review were reported, along with a gene prioritization analysis of Genome-Wide Association Studies (GWAS), aimed at identifying potential functional candidate genes linked to resistance against mastitis-related traits in dairy cattle. The gene and QTL annotation process revealed a total of 9,125 unique genes and 43,646 unique QTL from a pool of 39 different studies. A majority of the annotated genes were associated with traits related to SCC, particularly within the Holstein breed. Among all annotated genes, 74 genes were common to Holstein, Jersey, and Ayrshire breeds. Approximately 7.5% of the annotated QTL were categorized under the 'health' class, within which 2.6% and 2.2% of the QTL were specifically linked to CM and somatic cell count-related traits. Prioritization analysis resulted in the identification of 427 significant genes following correction for multiple testing (false discovery rate of 5%) from 26 studies. These prioritized genes were predominantly associated with immune response, the regulation of secretion, locomotion, cell proliferation, and development. This study not only fine-mapped previously identified genomic regions related to mastitis resistance, but also pinpointed crucial functional candidate genes that play a role in resistance against mastitis.

Furthermore, in **Chapter 5**, the results of an investigation to examine the relationship between SCS and copy number variants (CNVs) were reported. A deletion or duplication of genomic sequences ranging from 50 bp to 5 Mb is commonly referred to as CNV (Feuk et al., 2006; Mills et al., 2011). Copy Number Variants represent a quantitative genomic structural variation and stand as a substantial source of genomic diversity (Redon et al., 2006). These CNVs have the potential to unveil additional genomic insights that may not have been previously identified using Single Nucleotide Polymorphisms (SNPs). **Chapter 5** resulted in 443 CNV

Regions (CNVRs), encompassing approximately 6.45% of the bovine genome. The association analysis revealed 3 significant CNVRs linked to SCS in Holstein cattle, with 2 of them being novel discoveries. Annotation analysis provided a list of 165 unique genes and 337 unique QTLs.

In the context of GP, the study reported in **Chapter 6** focused on the application of the single-step genomic best linear unbiased prediction (ssGBLUP) method for the prediction of 6 different SCM traits and SCS within the first 5 to 30 DIM in heifers. The single-step GBLUP approach integrates genomic, pedigree, and phenotypic information into a single model for genetic evaluation. This approach enables the simultaneous inclusion of both genotyped and non-genotyped animals in the evaluation, facilitated through a hybrid relationship matrix (Legarra et al., 2009). The main advantages of ssGBLUP include its ability to address pre-selection bias, accommodate complex models (including multiple traits and threshold models), and enhance the accuracy of genomic estimated breeding values (GEBV) (Aguilar et al., 2010; Christensen et al., 2012; Legarra et al., 2014). **Chapter 6** demonstrated that the application of the ssGBLUP model led to a substantial improvement in the validation reliabilities of genomic estimated breeding value (GEBV) for young animals, resulting in an average gain of 0.28 reliability points for SCM traits and 0.19 points for SCS when best combination of scaling factors were utilized. Furthermore, with respect to average theoretical reliabilities, gains of 0.22 and 0.27, and 0.20 and 0.28 were observed for full and truncated datasets of SCM traits and SCS, respectively.

In summary, the research presented in this thesis comprehensively explored genetic and genomic dimensions of mastitis. **Chapters 2** and **3**, despite the low heritability of the traits, presented compelling evidence for exploitable genetic variability of SCM and IMI traits. **Chapters 4** and **5** reinforced the notion that mastitis is a polygenic trait influenced by multiple genes, each with small effects. Lastly, in **Chapter 6**, it was demonstrated that the ssGBLUP method generates

GEBV with improved reliability and reduced bias for young animals when compared to EBV obtained through the conventional BLUP approach.

7.4. Future directions

Even in the era of genomics phenotypes are critically important. It is, therefore, crucial to enhance the current approach to collecting mastitis phenotype data, particularly in the case of CM. Mastitis is caused by several pathogens that can vary with time and location. To date, there is no pathogen-specific selection program for mastitis resistance due to the limited number of phenotypes available regarding the pathogens. Therefore, an increased focus on microbial testing to identify the pathogens responsible for IMI is necessary to create a reference population for genomic prediction, which comprises animals with both phenotypic and genomic data which substantially contributes to increasing the accuracy of genetic evaluation. Currently, in Canada, the number of proven bulls within the reference population for the currently defined mastitis resistance traits is limited. Hence, there is a pressing need to augment the number of proven bulls and expand the inclusion of cows within the reference population for mastitis resistance. Additionally, as highlighted in **Chapter 6**, the use of ssGBLUP is especially advantageous when working with a small number of animals that are both genotyped and phenotyped for mastitis-related traits, but there are additional phenotyped animals to include in the analyses. Moreover, further research is still required to identify new, biologically sound proxies for mastitis resistance. Finally, the integration of other 'omics' data, such as epigenomics, transcriptomics, proteomics, and metabolomics, may enhance the selection for mastitis resistance and should be investigated.

7.5. References

Aghamohammadi, M., D. Haine, D. F. Kelton, H. W. Barkema, H. Hogeveen, G. P. Keefe, and S. Dufour. 2018. Herd-level mastitis-associated costs on Canadian dairy farms. Front. Vet. Sci. 5:100.

Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

Andersen, S., I. R. Dohoo, R. Olde Riekerink, and H. Stryhn, and Mastitis Research Workers' Conference. 2010. Diagnosing intramammary infections: Evaluating expert opinions on the definition of intramammary infection using conjoint analysis. J. Dairy Sci. 93:2966–2975.

Bates, A. J. and I. Dohoo. 2016. Risk factors for peri-parturient farm- er diagnosed mastitis in New Zealand dairy herds: Findings from a retrospective cohort study. Prev. Vet. Med. 127:70–76.

Berry, D. P. and W. J. Meaney. 2006. Interdependence and distribution of subclinical mastitis and intramammary infection among udder quarters in dairy cattle. Prev. Vet. Med. 75:81–91.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571.

Constable, P., K. Hinchcliff, S. Done, W. Grünberg, and O. Radostits. 2017. Veterinary medicine: A textbook of the diseases of cattle, horses, sheep, pigs and goats (Edition 11). Elsevier, St. Louis, MO, USA.

Cook, W. F., E. A. Fiez, and L. K. Fox. 1992. Mastitis in first lactation Southwest Idaho dairy cows. J. Dairy Sci. 75:158.

Harmon, R. J. 1994. Physiology of mastitis and factors affecting somatic cell counts. J. Dairy Sci. 77:2103–2112.

Heringstad, B., G. Klemetsdal, and J. Ruan. 2000. Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. Livest. Prod. Sci. 64:95–106.

Heringstad, B., G. Klemetsdal, and T. Steine. 2003. Selection responses for clinical mastitis and protein yield in two Norwegian dairy cattle selection experiments. J. Dairy Sci. 86:2990–2999.

Jamrozik, J., A. Koeck, F. Miglior, G. J. Kistemaker, F. S. Schenkel, D. F. Kelton, and B. J. Van Doormaal. 2013. Genetic and genomic evaluation of mastitis resistance in Canada. Interbull Bull. 47:43-51.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012. Health recording in Canadian Holsteins: Data and genetic parameters. J. Dairy Sci. 95:4099-4108.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54–65.

Martin, P., H. W. Barkema, L. F. Brito, S. G. Narayana, and F. Miglior. 2018. Symposium review: Novel strategies to genetically improve mastitis resistance in dairy cattle. J. Dairy Sci. 101:2724-2736.

Mills, R. E., K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, A. Chinwalla, D. F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L. M. Iakoucheva, Z. Iqbal, S. Kang, J. M. Kidd, M. K. Konkel, J. Korn, E. Khurana, D. Kural, H. Y. K. Lam, J. Leng, R. Li, Y. Li, C. Y. Lin, R. Luo, X. J. Mu, J. Nemesh, H. E. Peckham, T. Rausch, A. Scally, X. Shi, M. P. Stromberg, A. M. Stütz, A. E. Urban, J. A. Walker, J. Wu, Y. Zhang, Z. D. Zhang, M. A. Batzer, L. Ding, G. T. Marth, G.

McVean, J. Sebat, M. Snyder, J. Wang, K. Ye, E. E. Eichler, M. B. Gerstein, M. E. Hurles, C. Lee, S. A. McCarroll, and J. O. Korbel. 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470:59–65.

Oliver, S. P., M. J. Lewis, B. E. Gillespie, H. H. Dowlen, E. C. Jae- nicke, and R. K. Roberts. 2003. Prepartum antibiotic treatment of heifers: Milk production, milk quality and economic benefit. J. Dairy Sci. 86:1187–1193.

Parker Gaddis, K. L., J. B. Cole, J. S. Clay, and C. Maltecca. 2014. Genomic selection for producer-recorded health event data in US dairy cattle. J. Dairy Sci. 97:3190-3199.

Parker, K. I., C. Compton, F. M. Anniss, A. Weir, C. Heuer, and S. McDougall. 2007. Subclinical and clinical mastitis in heifers following the use of a teat sealant precalving. J. Dairy Sci. 90:207–218.

Piepers, S., L. De Meulemeester, A. De Kruif, G. Opsomer, H. W. Barkema, and S. De Vliegher. 2007. Prevalence and distribution of mastitis pathogens in subclinically infected dairy cows in Flanders, Belgium. J. Dairy Res. 74:478–483.

Piepers, S., K. Peeters, G. Opsomer, H. W. Barkema, K. Frankena, and S. De Vliegher. 2011. Pathogen group specific risk factors at herd, heifer and quarter levels for intramammary infections in early lactating dairy heifers. Prev. Vet. Med. 99:91–101.

Pyörälä, S. 2003. Indicators of inflammation in the diagnosis of mastitis. Vet. Res. 34:565–578.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A.R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. González, M. Gratacòs, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, J. Zhang, L. Armengol, D.F. Conrad,

X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K.W. Jones, S.W. Scherer, and M. E. Hurles. 2006. Global variation in copy number in the human genome. Nature 444:444–454.

Richert, R. M., K. M. Cicconi, M. J. Gamroth, Y. H. Schukken, K. E. Stiglbauer, and P. L. Ruegg. 2013. Risk factors for clinical mastitis, ketosis, and pneumonia in dairy cattle on organic and small conventional farms in the United States. J. Dairy Sci. 96:4269– 4285.

Ruud, L. E., K. E. Bøe, and O. Østerås. 2010. Associations of soft flooring materials in free stalls with milk yield, clinical mastitis, teat lesions, and removal of dairy cows. J. Dairy Sci. 93:1578–1586.

Saidi, R., D. Khelef, and R. Kaidi. 2013. Subclinical mastitis in cattle in Algeria: Frequency of occurrence and bacteriological isolates. J. S. Afr. Vet. Assoc. 84:E1–5.

Sampimon, O., H. W. Barkema, I. Berends, J. Sol, and T.J.G.M. Lam. 2009. Prevalence of intramammary infection in Dutch dairy herds. J. Dairy Res. 76:129–136.

Sharma, N., N. K. Singh, and M. S. Bhadwal. 2011. Relationship of somatic cell count and mastitis: An overview. Asian-Australalasian. J. Anim. Sci. 24:429-438.