Improving Accuracy of Genomic Prediction for Economically Important Traits in Canadian Holstein Dairy Cattle

by

Shadi Nayeri

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Animal Science

Department of Agricultural, Food and Nutritional Science
University of Alberta

ABSTRACT

Milk is a valuable source of energy, high quality protein, and several key minerals and vitamins for humans. Selection for milk production in dairy cattle was first based on phenotype and pedigree information and the knowledge of genetic parameters for the trait of interest. However, substantial advances in molecular genetics technology including bovine SNP discovery and sequencing projects have enabled researchers to apply new selection tools (such as genomic selection or GS) to identify genetically superior animals. GS is based on linkage disequilibrium (LD) between unknown functional variants and SNP genotypes that are spread out across the whole genome. It is hypothesized in this work that incorporation of candidate causal mutations into genotyping panels can increase the accuracy of genomic predictions. These variations are expected to more likely affect the trait and to be more effective across populations and generations due to persistent LD. The objectives of this study were to identify candidate causal genes and variants for production and fertility traits in Holstein dairy cattle, and then to include the candidate variants in genomic predictions by constructing and using a custom genotyping panel. In order to develop a more balanced selection tool, fertility traits were also included in this study. In the first study, genome-wide association analysis (GWAS) was performed to identify or refine the positions of genomic regions associated with milk production, milk components and fertility traits, and these positions were used to identify genes and pathways that may influence these traits. The identified QTL regions for production traits support previous findings, overlapping with genes with known relevant biological functions identified in earlier studies such as *DGAT1* and *CPSF1*. A significant region on chromosome 21 encompassing the gene *FAM181A* and not previous linked to fertility in dairy cattle was identified for the calving to first service interval and days open traits. A functional enrichment analysis of the association results

yielded GO terms consistent with the specific phenotypes tested; for example, GO term GO:0007595 (lactation) for milk production (MILK) and GO:0040019 (positive regulation of embryonic development) for calving to first service interval (CTFS). In the second study, GWAS was performed to determine the locations of genome regions affecting lifetime profit index (LPI), female fertility (age at first service, cow first service to conception, heifer and cow non-return rate) and longevity (direct and indirect herd life and daughter herd life) in the Canadian Holstein dairy cattle population. As with study 1, the results overlap in part with previous findings and some novel regions were discovered, specifically loci on BTA13 and BTA27 associated with lactation persistency. Previously proposed causative and candidate genes supported by this work include *GRINA* while new candidates are *SLC2A4RG* and *THRB*. In the third study, a custom genotyping panel was designed using the GWAS results from the first two studies, sequence and SNP information from a variety of sources, with the goal of including candidate causal mutations. The new Affymetrix panel, termed 80K, was evaluated as a tool for improving genomic predictions. The effects of combining the panel with the existing 50K panel (creating a 124K panel) and of using only those SNPs that overlap with transcribed sequences (transcriptome panel) were also investigated. The results showed that a small increase in the accuracy of genomic prediction (0.57% averaged across all traits) was achieved by incorporating the genotypes of candidate variants identified through GWAS. The accuracy of prediction using the transcriptome panel was better (0.72% averaged across all traits). In summary, GWAS results have detected several regions associated with milk production, LPI, longevity and fertility traits in Canadian Holstein cattle. Most of these regions were identified in other studies; however, novel regions of association were detected for days open, calving to first service interval and lactation persistency. These novel regions can be used to guide future mapping and functional

analysis to identify genes and sequence differences that explain variations in these traits. The genomic prediction results obtained through the use of custom genotyping panel show a small increase in the accuracy; however, the accuracy was better for a subset of variants selected within the transcribed regions. Coupling variant annotation information with more recent approaches, including imputation to the sequence data, may lead to better prediction accuracies.

Preface

Chapter two of this thesis has been published as Nayeri, Shadi and Paul Stothard, "Tissues, metabolic pathways and genes of key importance in lactating dairy cattle" in *Springer Science Reviews* (2016) 1:29. Shadi Nayeri drafted the manuscript and Paul Stothard contributed in revising and reviewing the drafted manuscript.

Chapter three of this thesis has been published as Shadi Nayeri, Mehdi Sargolzaei, Mohammed K. Abo-Ismail, Natalie May, Stephen P. Miller, Flavio Schenkel, Stephen S. Moore and Paul Stothard "Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle", in *BMC Genetics* (2016) 17:75. Paul Stothard, Stephen P. Miller, Flavio Schenkel, Stephen Moore and Shadi Nayeri developed the project, assisted with analysis and edited the manuscript. Shadi Nayeri performed data analysis and drafted the manuscript. Natalie May helped to draft the manuscript and assisted with the interpretation of results. Mehdi Sargolzaei and Mohammed K. Abo-Ismail assisted with data analysis and edited the manuscript. All authors contributed to the revision of the manuscript and approved the final version of the manuscript.

Chapter four of this thesis has been published as Shadi Nayeri, Mehdi Sargolzaei, Mohammed K. Abo-Ismail, Stephen Miller, Flavio Schenkel, Stephen Moore and Paul Stothard, "Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle" In the Journal of *Dairy Science*, (2016) 100:1-13. Paul Stothard, Stephen P. Miller, Flavio Schenkel, Stephen Moore and Shadi Nayeri developed the project, assisted with analysis and edited the manuscript. Shadi Nayeri performed data analysis and drafted the manuscript. Mehdi Sargolzaei and Mohammed K. Abo-Ismail assisted with data

analysis and edited the manuscript. All authors contributed to the revision of the manuscript and approved the final version of the manuscript.

Dedication

This work has been dedicated to my beloved Mother, Touran Faghan, who has been my inspiration, support and best friend during my whole life. My endless love for her for showing me anything is possible with faith, hard work and determination; and to my beloved late Father, Massihodin Nayeri, for his love and support and teaching me to celebrate and embrace life regardless of its obstacles. I could not have asked for better parents or role models.

Acknowledgements

First and foremost, I would like to express my deep gratitude and sincere thanks to my supervisor, Dr. Paul Stothard, for the patient guidance, support, encouragement and advice he has provided during this time as his student. I have been very lucky to have a supervisor who could wisely direct my research and me by his comprehensive and prompt responses to my questions throughout these years. This work would have not been possible without his support.

I would like to have my sincere and especial thanks to Dr. Stephen Moore, who gave me this opportunity to put my first step in this journey; for his continuous support, encouragement, invaluable and constructive suggestions and comments during all these years.

Special thanks to my committee members Dr. Stephen Miller and Dr. Michael Dyck for their valuable comments and suggestions and providing this opportunity for me to gain more experience working with other team members in their group.

My special thanks to Dr. Mehdi Sargolzaei and Dr. Flavio Schenkel from University of Guelph for advising and directing me in statistical analyses, their valuable suggestions and comments in drafting the result, and also helping me in learning new genomic methods in a friendly and professional environment in Guelph as well as my course of study.

I would like to thank the financial support provided by Natural Sciences and Engineering Research Council (NSERC) of Canada, Bovine Semex Alliance and Dr. Jacques Chesnais for his support. My special thanks to University of Alberta lab members, Dr. Urmila Basu, Jackson Mah and Meng Yan, for their help in making RNA-sequencing data. My special thank to Dr. Laki Goonewardene and Dr. Filippo Miglior for supporting me during my study.

I would like to thank my family for always supporting me regardless of distance. My regards to my parents Touran Faghan and Masihoddin Nayeri, and my beloved siblings, Shirin Nayeri, Amirhossein Nayeri and Azita Nayeri.

I would like to thank my bigger family, the livestock-Gentec group, Dr. Mohammad Abo Ismail, Dr. Ghader Manafiaz, Dr. Chinyere Ekine-Dzivenu, Dr. Alireza Akhavan, Homa Asaker, Dr. Younes Miar, Aidin Forutan Naddafi, Dr. Arun Kommadath, Dr. Brian Karisa, Bibaswan Goshal, Natali May, Dr. Everestus Akanno, Dr. Duc Lu, Dr. Chunyan Zhang, Anna Szenthe, Dr. Mary De Pauw, Robert ForYah for their help, friendship and continues support and assistance during the whole years of my study. My special thanks to my beloved friends, Hale Oguzlu-Baldelli, Alberto Baldelli, Firouzeh Toosi, Nima Fazeli, Maryam Nesari, Sahar Navidghasemi, Nabi Goudarzi, Vianey Landeros and Narges Zare who were always there for me in my bad and good times.

Finally, I praise my God for protecting me, leading me through this path and holding my back in my difficult times.

Table of Contents

List of Tables

List of Figures

List of Symbols

3βHSD	Δ^5-3β hydroxysteroid dehydrogenase isomerase
ABCA1	ATP binding cassette subfamily A member 1
ABCG2	ATP binding cassette subfamily G member 2
ACC	Acetyl-CoA carboxylase
ACSL1	Acyl-CoA synthetase long-chain family member 1
AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1
AGPAT6	Glycerol-3-phosphate acyltransferase 4
ATP	Adenosine triphosphate
BHBA	Beta-hydroxybutyrate
BTA	*Bos taurus* autosome
BTN1A1	Butyrophilin, subfamily I, member AI
cAMP	Cyclic adenosine monophosphate
CAST	Calpastatin
CDN	Canadian dairy network
ChREBP	Carbohydrate responsive element-binding protein
CL	Corpus luteum
CPT-I	Carnitine palmitoyltransferase I
CPT-II	Carnitine-acylcarnitine translocase
CTU1	Cytosolic thiouridylase subunit 1
CYP11A1	Cytochrome P450, family 11, subfamily A, polypeptide 1
DAG	Diacylglycerol
DD	Daughters deviations
DEBV	De-regressed bull proof

DGAT	Diacylglycerol acyltransferase
DGAT1	Diacylglycerol O-acyltransferase 1
DGV	Direct genomic value
E2	Estradiol
EBV	Estimated breeding value
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
eQTL	Expression QTL
ER	Endoplasmic reticulum
FA	Fatty acid
FABP	Fatty acid binding proteins
FAD^+	Flavin adenine dinucleotide
FADH	Flavin adenine dinucleotide hydrogen
FDR	False discovery rate
FSH	Follicle stimulating hormone
GBLUP	Genomic best linear unbiased prediction
GEBV	Genomic estimated breeding value
GHR	Growth hormone receptors
GnRH	Gonadotrophin releasing hormones
GO	Gene ontology
GPAM	Glyceol-3-phosphate acyltransferase
GPAT	Glycerol-3-phosphate acyltransferase
GQLS	Quasi-Likelihood Score
GS	Genomic selection
GTHR	Generalized thyroid hormone resistance

GWAS	Genome-wide association study
HD	High-density
HMG-CoA	3-hydroxy-3-methylglutaryl-CoA
HMG-CoA reductase	3-hydroxy-3-methylglutaryl CoA reductase
HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase
HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1
HSL	Hormone sensitive lipase
IGF1	Insulin-like growth factor 1
IGF1R	Insulin like growth factor 1 receptor
IGFBP	IGF binding protein
IGFBP4	IGF-binding protein 4
IL-1	Unterleukin-1
INSR	Insulin receptor
IPA	Ingenuity pathway analysis
IPKB	Ingenuity Pathway Knowledge Base
LCACoA	LC-acyl-CoA
LCFA	Long chain fatty acids
LD	Linkage disequilibrium
LE	Linkage equilibrium
LH	Luteinizing hormone
LPA	Lysophosphatidic acid
LPIN1	Lipin I
LPL	Lipoprotein lipase
MAF	Minor allele frequency
MDS	Multidimensional scaling

NAD^+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide hydrogen
NEB	Negative energy balance
NEFA	Non-esterified fatty acids
NGS	Next-generation sequencing
NPFFR2	Neuropeptide FF receptor 2
P4	Progesterone
PA	Phosphatidate
PA	Parental average
PAPP2-A2	Pregnancy-associated plasma protein-A2
PCCA	Propionyl-CoA carboxylase alpha subunit
PCK1	Phosphoenolpyruvate carboxykinase 1
PCK2	Phosphoenolpyruvate carboxykinase 2, mitochondrial
PEY	Persistency of milk energy yield
$PGF_{2\alpha}$	Prostaglandin F2 alpha
PMY	Persistency of milk yield
PNPLA2	Adipose triglyceride lipase
PPARGC1A	PPARG coactivator 1 alpha
PPARα	Peroxisome proliferator-activated receptor α
PPY	Persistency of protein yield
PSCS	Persistency of somatic cell score
Q-Q plots	Quantile-quantile plots
QC	Quality control
QTL	Quantitative trait loci
QTN	Quantitative trait nucleotide

SCD	Stearoyl-CoA desaturase
SNEB	Severe negative energy balance
SNP	Single nucleotide polymorphism
SREBF2	Sterol regulatory element binding transcription factor 2
SREBP-1c	Sterol regulatory element-binding protein-1c
SREBPs	Sterol regulatory element-binding proteins
T3	Triiodothyronine
T4	Thyroxine
TAG	Triacylglycerides
TBV	True breeding value
TG	Triglycerides
THRB	Thyroid hormone receptor, beta
TNFα	Tumor necrosis factor α
VLDL	Very low density lipoprotein

CHAPTER 1. GENERAL INTRODUCTION

1.1. Introduction

Genetic improvement has been one of the fundamental drivers behind the success of the Canadian dairy sector. Due to selection over the last 30 years, milk production has vastly increased particularly within North American Holstein-Friesian cattle (Walsh et al., 2011). However, one of the drawbacks of improvement in milk production has been a decline in fertility performance and reductions in genetic merit for other functional traits such as health and fitness traits (Walsh et al., 2011; Lucy, 2001; Egger-Danner et al., 2014). For example, conception rates to first service intervals have decreased by 30 to 52% between 1985 and 2003 and the incidence of production associated disease has increased (Dillon et al., 2006; Macdonald et al., 2008; Pryce et al., 2004; Norman et al., 2009; Oltenacu and Broom, 2010). Efforts to address fertility associated issues face several challenges: fertility traits have low heritability, are expressed late in life and are affected by multiple factors such as nutrition and management (Oltenacu and Broom, 2010). Demand for milk production is increasing, and based on recent estimations (in 2010) the American dairy industry (United State Department of Agriculture) is expected to maintain the same milk supply from eight million dairy animals by 2050 as opposed to the current population of nine million dairy cattle (Santos et al., 2010; Walsh et al., 2011). Understandably, breeding goals in dairy cattle are changing to consider both production (milk) and functional traits (such as health and fertility) (Groen et al., 1997; Miglior et al.).

Major improvements in increasing milk production have been achieved with quantitative genetics based on phenotypic and pedigree information and the knowledge of genetic parameters for the traits of interest (Dekkers and Hospital, 2002). Recent advances in molecular genetics

have enabled researchers to identify and use DNA markers for making selection decisions. In marker-assisted selection, markers that are in linkage equilibrium (LE) or linkage disequilibrium (LD) with quantitative trait loci (QTL) are used (Dekkers, 2004). However, the application of MAS by the dairy industry has been limited (Boichard et al., 2003). The reason is that many quantitative traits including milk production in dairy cattle are affected by many loci each with a small effect; the number of markers available for MAS was small and therefore only a small proportion of the total genetic variance could be captured with these markers. Consequently, relatively small gains were possible (Sallam et al., 2015). Furthermore, the cost of genotyping these early markers was high (Dekkers, 2004; Lande and Thompson, 1998; Sallam et al., 2015) The sequencing of the bovine genome and the subsequent HapMap project made a huge amount of markers available in the form of single nucleotide polymorphism (SNPs) (Meredith et al., 2012; Matukumalli et al., 2009; Gibbs et al., 2012). This progress, with the emergence of high throughput genotyping, allows for genomic selection (GS) to be performed. With GS, selection decisions are based on genomic breeding values (GEBV) calculated for selection candidates using estimates of single nucleotide polymorphism (SNP) marker effects that were trained in a reference population with both phenotypes and genotypes (Goddard and Hayes, 2007; Hayes et al., 2009). Long generation intervals and the fact that nearly all the economically important traits are expressed only in female cattle has made the dairy cow an especially good candidate for genome-based selection methods (Dekkers, 2004).

1.2. Research rationale and hypothesis

Genomic selection can be applied without knowledge of the underlying quantitative trait nucleotides (the QTN) that affect phenotype. However, the identification of the QTNs or markers which are in very strong linkage disequilibrium (LD) with causal mutations can give us a better

understanding of the biology underlying these traits. Also, the application of panels based on QTN information may provide greater accuracy in calculating genomic prediction due to a more direct link between phenotype and genotype, and may yield prediction equations that work more effectively across breeds and generations. There is also the potential to use QTN or near-to-QTN markers for the development of smaller, less expensive marker panels for genomic prediction evaluations. The goal of this work was to identify QTNs or markers near QTNs and to include these on a custom panel for use in genomic selection in dairy cattle. Using this panel, this work tested the hypothesis that the inclusion of candidate QTN genotypes in genomic prediction calculations can increase prediction accuracies.

1.3. Research objectives and chapter details

The objectives of this study were to 1) identify DNA markers targeted on causal mutations for milk production, longevity, lifetime profit index and fertility traits through genome-wide association studies and candidate gene identification and to 2) test whether application of these markers can increase accuracy of genomic selection in Canadian dairy cattle. Having known that potential QTNs can be identified using information from a variety of sources (Veerkamp and Beerda, 2007; Zhu and Zhao, 2007; Grant et al., 2011; Ron and Weller, 2007), our experiments were performed through the following steps. First, positional and functional candidate genes identified based on the results of association studies of a high-density (HD) marker panel and on known biological roles or expression patterns of genes. Second, DNA variations were identified within the selected candidate genes. For this step we used information from a public SNP database, genomic DNA sequencing and RNA-seq. Third, the potential impacts of variants were predicted (for example initiator-codon-variant or missense variant) using bioinformatics tools. A custom SNP panel was then designed containing variants identified through previous steps and

was used to genotype Holstein bulls in training and validation sets. The information derived from the training set was used to estimate marker allele substitution effects for all the SNPs in the panel and to calculate genomic breeding values (GEBVs) for the validation set. Finally, the accuracy of the GEBVs was assessed through comparison with the bull proofs.

This dissertation is organized as follows. **Chapter 2** presents background material covering tissues, metabolic pathways and genes that are involved in the physiological adaptations early in lactation in dairy cattle. **Chapter 3** reports results from a genome-wide association study for production (milk production, fat production, fat deviation, protein production and protein deviation) and fertility (calving to first service interval, days open, daughter fertility and heifer first service to calving interval) traits along with positional and functional candidate genes within the associated regions. Additionally, an enrichment analysis was performed to test for overrepresentation of significant SNPs in biological pathways. **Chapter 4** describes the result of the association analysis for lifetime profit index, longevity (herd-life, indirect herd-life and direct herd-life), lactation persistency and other female fertility-associated traits (age at first service, cow first service to conception and heifer and cow non-return rate). **Chapter 5** is a description of the development of a custom made SNP panel and calculation of the accuracy of GEBVs. The main results, limitations of the work and recommendations are summarized in **Chapter 6**.

1.4. References

Boichard DB, Grohs C, Bourgeois F, et al (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. Genet Sel Evol 35:77–101. doi: 10.1051/gse

Dekkers JC (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci. doi: 82/13_suppl/E313 [pii]

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3:22–32. doi: 10.1038/nrg701

Dillon P, Berry DP, Evans RD, et al (2006) Consequences of genetic selection for increased milk production in European seasonal pasture based systems of milk production. Livest Sci 99:141–158. doi: 10.1016/j.livprodsci.2005.06.011

Egger-Danner C, Cole JB, Pryce JE, et al (2014) Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal 9:191–207. doi: 10.1017/S1751731114002614

Gibbs R, Taylor JF, Van Tassell CP, et al (2012) NIH Public Access. Science (80-) 29:997–1003. doi: 10.1016/j.biotechadv.2011.08.021.Secreted

Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet 124:323–330.

Grant JR, Arantes AS, Liao X, Stothard P (2011) In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. Bioinformatics 27:2300–2301. doi: 10.1093/bioinformatics/btr372

Groen AF, Steine T, Colleau J-J, et al (1997) Economic values in dairy cattle breeding, with special reference to functional traits. Report of an EAAP-working group. Livest Prod Sci 49:1–21. doi: 10.1016/S0301-6226(97)00041-9

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–43. doi: 10.3168/jds.2008-1646

Lande R, Thompson R (1998) Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. Heredity (Edinb) 80 (Pt 4):489–498. doi: 10.1046/j.1365-2540.1998.00308.x

Lucy MC (2001) Reproductive loss in high-producing dairy cattle: where will it end? J Dairy Sci 84:1277–93. doi: 10.3168/jds.S0022-0302(01)70158-0

Macdonald K a, Verkerk G a, Thorrold BS, et al (2008) A comparison of three strains of holstein-friesian grazed on pasture and managed under different feed allowances. J Dairy Sci 91:1693–1707. doi: 10.3168/jds.2007-0441

Matukumalli LK, Lawley CT, Schnabel RD, et al (2009) Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS One. doi: 10.1371/journal.pone.0005350

Meredith BK, Kearney FJ, Finlay EK, et al (2012) Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet 13:21. doi: 10.1186/1471-2156-13-21

Miglior F, Chesnais J, Van Doormaal BJ Genetic improvement: a major component of increased

dairy farm profitability.

Norman HD, Wright JR, Hubbard SM, et al (2009) Reproductive status of Holstein and Jersey cows in the United States. J Dairy Sci 92:3517–3528. doi: 10.3168/jds.2008-1768

Oltenacu PA, Broom DM (2010) The impact of genetic selection for increased milk yield on the welfare of. 19:39–49.

Pryce JE, Royal MD, Garnsworthy PC, Mao IL (2004) Fertility in the high-producing dairy cow. Livest Prod Sci 86:125–135. doi: 10.1016/S0301-6226(03)00145-3

Ron M, Weller JI (2007) From QTL to QTN identification in livestock - Winning by points rather than knock-out: A review. Anim Genet 38:429–439. doi: 10.1111/j.1365-2052.2007.01640.x

Sallam AH, Endelman JB, Jannink J-L, Smith KP (2015) Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. Plant Genome 8:0. doi: 10.3835/plantgenome2014.05.0020

Santos JEP, Bisinotto RS, Ribeiro ES, et al (2010) Applying nutrition and physiology to improve reproduction in dairy cattle. Soc Reprod Fertil Suppl 67:387–403. doi: 10.5661/RDR-VII-387

Veerkamp RF, Beerda B (2007) Genetics and genomics to improve fertility in high producing dairy cows. Theriogenology 68:266–273. doi: 10.1016/j.theriogenology.2007.04.034

Walsh SW, Williams EJ, Evans ACO (2011) A review of the causes of poor fertility in high milk producing dairy cows. Anim Reprod Sci 123:127–138. doi:

10.1016/j.anireprosci.2010.12.001

Zhu M, Zhao S (2007) Candidate gene identification approach: Progress and challenges. Int J Biol Sci 3:420–427.

CHAPTER 2. LITERATURE REVIEW: TISSUES, METABOLIC PATHWAYS AND GENES of KEY IMPORTANCE IN LACTATING DAIRY CATTLE[1]

2.1. Abstract

Milk and dairy products are valuable sources of food for humans. Increased milk yield and changes in milk composition in dairy cows have been achieved through a variety of means including better nutrition, management and genetic selection. This selection can be performed without consideration of the specific genes and genetic variation involved. However, association analysis using dense SNP genotyping panels provides an approach for identifying genomic regions affecting milk production. Coupling physiological and metabolic information with association analysis results can provide greater insight into the specific genetic variants and molecular mechanisms affecting production traits as well as the potential effects of these variants on fertility in dairy cattle. To this end, this review highlights key tissues, metabolic pathways and genes of importance in lactating dairy cattle, particularly early in lactation. Physiological and metabolic adaptations in three key tissues (adipose, mammary gland and liver) are discussed, followed by the important endocrine adaptations during negative energy balance (NEB). Key genes mediating metabolic and endocrine adaptations are also highlighted. Finally, genes that account for variation in production traits are presented in relation to the tissues and processes described. Knowledge of the genes and pathways involved will be important for ongoing efforts aimed at finding other genes and variants that contribute to milk production and fertility traits.

[1] A version of this chapter has been published in the journal of *Springer Science Reviews*. Nayeri et al. 2016. Literature Review: Tissues, metabolic pathways and genes of key importance in lactating dairy cattle. pp 1:29 DOI: 10.1007/s40362-016-0040-3

Also, a better understanding of the molecular basis of these traits may lead to more accurate genomic predictions.

2.2. Introduction

Humans have recognized milk and dairy products as a valuable source of sustenance since as early as 4000 BC (Lucy 2001; Bauman et al. 2006). Indeed, milk is a source of energy, high quality protein, several key minerals and vitamins (Bauman et al. 2006). The demand for milk and milk products continues to increase, as does the production capacity of individual cows. For example, from 2005 to 2012, milk production of the Canadian dairy herd increased by 6%, while the number of dairy cows declined by 11% (www.cdn.ca). Ongoing genetic selection as well as advances in the understanding of the biology of lactation and biosynthesis of milk such as improved understanding of the interrelations between dietary components, digestive processes in the rumen and the regulation of mammary synthesis of milk, fat have led to improvements in management and substantial increases in milk production and productivity (Bauman et al. 2006; Ha et al. 2015). This knowledge includes advanced understanding of: the biology of lactation in many mammals (such as goat, sheep, guinea pigs, mice, rats and several other species); the relationship between structure and function of mammary epithelial cells; the biochemical pathways for the synthesis of milk components; the role of hormones in the development of the mammary gland and the regulation of mammary gland function (Heid and Keenan 2005; Bauman et al. 2006; Keenan and Mather 2009). Specific genes and gene variants that account for variation in milk production traits have been identified and represent another source of knowledge of the molecular events that can contribute to increases in production. For example, a missense mutation in the *DGAT1* gene on chromosome 14 has been identified with major effects on milk composition and fat content in dairy cattle (Grisart et al. 2002). Two other loci with

major effects on milk yield and composition have been identified on chromosomes 6 and 20 and involving the *ABCG2* and *GHR* genes, respectively (Blott et al. 2003; Cohen-Zinder et al. 2005). The identification of such genes and variants has been guided in large part by existing knowledge of the roles of certain genes in lactation. This review seeks to present information on tissues, pathways and genes that can guide future studies aimed at understanding the underlying genetic differences that contribute to variation in milk production and fertility traits. Consideration of fertility is important because of the unfavourable correlations between milk yield and fertility traits (Pryce et al. 2004). Many reproductive disorders including late resumption of ovarian activity or poor conception rate are related to negative energy balance at the peak of lactation in dairy cows (Taylor et al. 2004; Wathes et al. 2007). The identification of the specific genetic variants responsible for variation in these traits should lead to more accurate approaches to genomic selection that work better across generations and breeds (Snelling et al. 2013), and may help us understand or predict potential negative effects on other traits.

Substantial advances in molecular genetics and genomics tools have made the identification of genes and mutations causing simply inherited Mendelian traits relatively straightforward (Dekkers 2004; Goddard et al. 2014). However, quantitative traits such as milk production in dairy cattle are polygenic, affected by multiple genes and mutations at many sites in the genome (Snelling et al. 2013; Goddard et al. 2014). Many of the mutations that affect these quantitative traits have small effects on the phenotype and explain only a small portion of the genetic variance. Therefore, successfully identifying casual mutations and regions affecting these traits is more difficult compared to simple, highly heritable traits (Goddard et al. 2014).

A major goal of quantitative trait loci (QTL) studies in livestock is to identify regions, genes and markers that can be used in breeding programs. A QTL is a section of DNA (a locus) that is

correlated with variation in a given phenotype (Khatkar et al. 2004). Identifying the QTL affecting a trait has previously involved low density markers and the application of linkage mapping (Meuwissen et al. 2001; Goddard et al. 2014). Subsequently, through the discovery of markers within new target regions, the use of fine mapping, and the application of high density SNP (single nucleotide polymorphism) panels, the accuracy of QTL detection increased (Kemper et al. 2015). Fine mapping relies on linkage disequilibrium (LD) between SNPs and unknown casual variants, which are also called quantitative trait nucleotides or QTN (Snelling et al. 2013; Höglund et al. 2014a). Recent association analyses using dense genetic markers have detected variants associated with milk production-related traits including mammary gland development, and prolactin signaling and involution pathways (Sutherland et al. 2007; Raven et al. 2014a; Höglund et al. 2014a; Höglund et al. 2015). After identifying a QTL region, however, it may still be difficult to determine which variants in the region truly affect the trait. Knowledge of the physiology of the trait and of relevant metabolic pathways can be valuable in this regard, as it can highlight genes in the QTL region of potential importance (Weng et al. 2011). One of the aims of this review is to facilitate discovery of the mechanisms underlying QTL associated with milk production and fertility related traits, through highlighting tissues, pathways, and genes known to play important roles in lactation.

2.3. Physiological and metabolic adaptations early in lactation in dairy cattle

The transition period in dairy cattle presents an enormous metabolic change and challenge to the high-yielding dairy cow (Bell 1995). During this time, the energy requirements of the cow increase to accommodate milk production and maintenance (Jorritsma et al. 2003; Walsh et al. 2011). This increase in energy requirements can be partially met by increased feed consumption but is limited due to low dry matter intake and decrease in appetite that tend to occur around this

time; the remainder is met by mobilization of body reservoirs (Grummer 2007). Adipose tissue is the predominant energy reserve in dairy cattle during periods of chronic energy deficit (Bell 1995; Roche et al. 2009). Through homeostasis mechanisms, adipose tissue optimizes non-esterified fatty acid mobilization to maintain physiological equilibrium and to provide the required energy early in lactation (Bauman and Currie 1980; Bell 1995; Roche et al. 2009). However, not only is the homeostasis mechanism in adipose tissue important to support changes in a cow's condition, a coordinated change in lipid metabolism of other body tissues is also necessary to support the physiological state of the animal (Bauman and Currie 1980; Roche et al. 2009). This mechanism of regulation is called homeorhesis and applies to nutritionally insensitive (genetically controlled) regulation of lipid metabolism in dairy cattle to support the physiological state of the animal (Bauman and Currie 1980; Roche et al. 2009). Liver is the main site for the uptake of serum free fatty acid, increased lipid β-oxidation and increased gluconeogenesis early in lactation in dairy cattle (Bauman and Currie 1980; Roche et al. 2009). Another metabolic adaptation associated with negative energy balance and homeostasis in dairy cows is related to increased use of nutrients and milk lipid droplet in the mammary gland (Roche et al. 2009). Therefore, the liver and the mammary gland in dairy cattle are also important tissues in homeostasis and homeorhetic control of lipid metabolism during early lactation (Fielding and Frayn 1998; Drackley 1999). The physiological and metabolic pathways as well as the regulatory components in these three tissues (adipose, liver, and mammary gland) are described in subsequent sections.

2.3.1. Adipose tissue

Early in lactation, lipid metabolism characteristics change in adipose cells (adipocytes). Endocrine profile changes and mobilization of fatty acid from adipocytes begins (Bauman and

Griinari 2001). Subsequently, the abundance of non-esterified fatty acids (NEFA) in the serum albumin increases to allow uptake by various tissues (Roche et al. 2009). Two main metabolic pathways optimizing NEFA mobilization to maintain physiological equilibrium are lipolysis and lipogenesis (Roche et al. 2009).

Lipogenesis

Major sites where lipogenesis generally occurs are the intestinal mucosal cells, the hepatocytes (liver cells) and the adipose tissue (Laliotis et al. 2010). In ruminants, the predominant sites are adipose tissue and the mammary gland of lactating dairy cows (Laliotis et al. 2010). These tissues are responsible for the uptake of pre-formed fatty acids from lipid circulation and for *de novo* fatty acid synthesis using acetyl-CoA derived from the catabolism of carbohydrates (Roche et al. 2009; Laliotis et al. 2010). Most of the carbohydrates in ruminants are fermented into acetate while butyrate and propionate are produced to a lesser extent. As such, acetate is the predominant lipogenic substrate in adipose tissue and the mammary gland for *de novo* fatty acid synthesis (Vernon et al. 2001; Laliotis et al. 2010).

Acetate is first transformed into pyruvate and then into acetyl-CoA through oxidation within mitochondria (Laliotis et al. 2010). Fatty acid synthesis (lipogenesis) begins with carboxylation of this acetyl-CoA to malonyl-CoA. This reaction is catalyzed by the rate-limiting enzyme, acetyl-CoA carboxylase (ACC) (Roche et al. 2009). Malonyl-CoA is then condensed with acetyl-CoA by Acyl-malonyl ACP condensing enzyme to produce a four-unit substrate and CO_2 as a result (Berg et al. 2002; Roche et al. 2009). The next three steps in fatty acid synthesis are reduction of a keto- group at C-3 to a methylene group and formation of butryl-ACP (Berg et al. 2002). With formation of burtryl-ACP, the first cycle of elongation completes. The elongation cycle continues with condensation of butryl ACP with malonyl ACP to form C_6-β-ketoacyl ACP,

and a similar cycle of reactions repeats until C_{16}-acyl ACP is formed; this intermediate is then hydrolyzed by thioestrase to yield palmitate (C16:00) and ACP (Berg et al. 2002). In the case of fatty acid uptake from circulating lipids, the second pathway of lipogenesis in the adipocyte starts, which is hydrolysis of plasma triacylglycerides (TAG) by lipoprotein lipase (LPL), producing NEFA and monoacylglycerides (Lehner and Kuksis 1996; Roche et al. 2009). Depending upon availability of glycerol-3-phosphate and monoacylglycerides, TAGs are synthesized through either phosphatidic or monoacylglycerol pathways (Lehner and Kuksis 1996; Roche et al. 2009).

Lipolysis

The hydrolysis of triacylglycerols (TAG) by lipase is activated by signals from molecules such as catecholamine (epinephrine and norepinephrine) and adrenocorticotropic hormones (Roche et al. 2009). These hormones trigger membrane receptors that activate adenylate cyclase (Fig. 1). Increased levels of cyclic adenosine monophosphate (cAMP) then simulate protein kinase A. This simulation leads to activation of lipase (hormone sensitive-lipase or HSL), which hydrolyzes fatty acids at the sn-1 and sn-3 positions (Berg et al. 2002). Then, monoacylglycerol lipase hydrolyzes the remaining fatty acid at the sn-2 position and generates 3 fatty acids (NEFA) and glycerol (Berg et al. 2002). Following this hydrolysis, NEFA mobilizes into circulation and quickly attaches to serum albumin for transport to various tissues (Roche et al. 2009).

2.3.2. Liver

The liver has a key role in lipid metabolism and maintaining lipid homeostasis in animals (Nguyen et al. 2008; Lin et al. 2013). Physiological, metabolic and endocrine adaptations that take place in the liver during early lactation support lipid metabolism in dairy cattle (Lin et al. 2013). Many metabolic disorders affecting transition cows, such as fatty liver syndrome and ketosis occur as a result of increased lipid and fatty acid oxidation in the liver during this period of metabolic challenge (Goff and Horst 1997; Dann and Drackley 2005). The oxidation of long chain fatty acids occurs in hepatic mitochondria and peroxisomes (Dann and Drackley 2005; Roche et al. 2009). Then triglycerides, the end product of liver β-oxidation, are carried by the lipoprotein VLDL (very low density lipoprotein). Lipoproteins are composed of triglycerides, cholesteryl esters, phospholipids and cholesterol (Kessler et al. 2014). Since cholesterol metabolism early in lactation has been a subject of intense investigation with regard to lipoprotein carriers, lipid metabolism related disorders, membrane fluidity and steroid hormone synthesis (Dann and Drackley 2005; Kessler et al. 2014), this section will review pathways and genes that are involved in lipid metabolism as well as cholesterol and steroid hormone synthesis in the liver.

Lipid metabolism in the liver

NEFA generated through lipid metabolism can be oxidized by liver mitochondria or peroxisomes for use as an energy source or used by the mammary glands as a source of milk fat (Drackley 1999; Roche et al. 2009). β-oxidation occurring in the mitochondria involves production of acetyl-CoA, and reduction of nicotinamide adenine dinucleotide (NAD^+) and flavin adenine dinucleotide (FAD^+) in order to produce adenosine triphosphate (ATP) in the citric acid cycle

and electron transport chain. The alternative pathway of hepatic fatty acid oxidation is through peroxisomes.

2.3.2.1.1. Oxidation of fatty acids in mitochondria

The β-oxidation of fatty acids in liver occurs mainly in the mitochondria. The carnitine palmitoyltransferase (CPT) system is recognized as a component of fuel homeostasis and transport system for these NEFA for β-oxidation (McGarry and Brown 1997). This system is composed of three enzymes, carnitine palmitoyltransferase I (CPT-I), carnitine-acylcarnitine translocase, and CPT-II (McGarry and Brown 1997; Dann and Drackley 2005). NEFA are delivered to the liver and converted to fatty acyl-CoA esters by acyl-CoA synthase. Then acyl-CoA is taken up by CPT-I on the outer mitochondrial membrane to be activated in the form of fatty acyl-carnitine (Berg et al. 2002). This step is believed to be the rate-limiting regulatory step in the metabolism of long chain fatty acids (McGarry and Brown 1997). Then, fatty acyl-carnitine permeates the inner membrane and by the enzyme CPT-II reforms fatty acyl-CoA (McGarry and Brown 1997). The activated fatty acyl then enters the pathway of β-oxidization in the mitochondrial matrix by a recurring sequence of four reactions: oxidation by flavin adenine dinucleotide (FAD), hydration, oxidation by NAD^+, and thiolysis by CoA (Table 2.1) (Mannaerts et al. 1979; Berg et al. 2002). The NADH (nicotinamide adenine dinucleotide hydrogen) and FADH produced during these reactions generate ATP in the citric acid cycle and electron transport chain, respectively (Berg et al. 2002). When fatty acid mobilization increases in adipocytes, excessive acetyl-CoA generated from β-oxidation is converted into acetoacetate and BHBA (beta-hydroxybutyrate), which are ketone bodies (Roche et al. 2009). Ketone bodies are an important energy-providing mechanism for vital organs such as the brain in early dairy cow's lactation (Roche et al. 2009). The remaining free fatty acids will be re-esterified to triglycerides

(TG) and exported as VLDL to the plasma (Fig. 1) (Kessler et al. 2014; Gross et al. 2015). Van den Top et al. (1995) and Kessler et al. (2014) showed that plasma VLDL-cholesterol, LPL-cholesterol (lipoprotein lipase) and TG concentrations decrease distinctively after parturition. Limited secretion of VLDL from liver and accumulation of TG in the liver can then lead to fatty liver syndrome (Van den Top et al. 1995).

2.3.2.1.2. Oxidation of fatty acid in peroxisomes

The oxidative pathway of NEFA in peroxisomes is similar to that in mitochondria. However, one of the products of these reactions is hydrogen peroxide instead of NADH. In addition, peroxisomes do not contain a respiratory chain linked to ATP which results in capture of less energy and more heat during peroxisomal β-oxidation (Drackley 1999). Therefore, peroxisomal β-oxidation may be considered an overflow pathway to oxidize fatty acids (FA) during extensive NEFA mobilization (Drackley 1999).

Cholesterol and steroid hormone metabolism

The transition period not only requires homeorhetic changes in glucose and lipid metabolism but also cholesterol metabolism (Bauman and Currie 1980; Kessler et al. 2014). Cholesterol is a fundamental lipid in modulating cell membrane fluidity and is the precursor of steroid hormones such as progesterone, testosterone, estradiol and cortisol (Berg et al. 2002). This section describes the biochemical pathways involved in cholesterol and steroid hormone synthesis, as well as regulatory components and interactions mediating cholesterol homeostasis in dairy cows during the transition period.

2.3.2.1.3. Cholesterol synthesis

Cholesterol and fatty acids are synthesized in the liver (Horton et al. 1998). Cholesterol has 27 carbon atoms in its structure, all of which are derived from acetyl-CoA. Synthesis of cholesterol starts with the formation of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) from acetyl-CoA and acetoacetyl-CoA. This stage is mediated by 3-hydroxy-3-methylglutaryle-CoA synthase (Viturro et al. 2009). Then, HMG-CoA is reduced to mevalonate for the synthesis of cholesterol. The synthesis of mevalonate is the main step in cholesterol formation and is catalyzed by 3-hydroxy-3-methylglutaryl CoA reductase (HMG-CoA reductase) (Berg et al. 2002; Viturro et al. 2009). Mevalonate is converted into 3-isopentenyl pyrophosphate and this molecule condenses in three steps to form farnesyl pyrophosphate (Berg et al. 2002). Then, two molecules of farnesyl pyrophosphate reduce to form squalene. This reaction is catalyzed by the endoplasmic reticulum enzyme, squalene synthase. Squalene is cyclized by oxidosqualene cyclase to form lanosterol (2. 1). The final stage of cholesterol synthesis is the conversion of lanosterol to cholesterol in a multistep process (Berg et al. 2002).

2.3.2.1.4. Steroid hormone synthesis from cholesterol

Cholesterol is the building block for all of the five major classes of steroid hormones: progestagens, glucocorticoids, mineralocorticoids, androgens and estrogens (Berg et al. 2002). Steroid hormone synthesis is stimulated and controlled by different peptide hormones in different organs. For example, follicle-stimulating hormone (FSH) controls the progesterone and estrogen synthesis in ovarian granulosa cells, whereas luteinizing hormone (LH) regulates progesterone synthesis in ovary luteinized granulosa-luteal cells, and androgen production in ovarian theca-interstitial cells (Hu et al. 2010). Steroid hormones contain 21 carbon atoms. The first stage in

the synthesis of steroid hormones is the removal of a six-carbon unit from cholesterol to form pregnenolone, a reaction catalyzed by the cytochrome P450 side-chain cleavage enzyme (P450scc, CYP11A1) on the inner mitochondrial membrane (Lambeth 1986; Hu et al. 2010). Progesterone is then synthesized from pregnenolone in two steps: 1) oxidization of the 3-hydroxyl group of pregnenolone and 2) isomerization of the Δ^5 double bond to a Δ^4 double bond (Berg et al. 2002). This step is catalyzed by the rate limiting Δ^5-3β hydroxysteroid dehydrogenase isomerase (3βHSD) enzyme in steroidogenic cells in the ovary (Payne and Hales 2004; Hu et al. 2010).

Androgens and estrogens are synthesized from progesterone in two steps: 1) hydroxylation of progesterone at C-17 and 2) cleavage of the side chain consisting of C-20 and C-21 carbons to yield androstenedione which is an androgen. This reaction is catalyzed by the 17 α -hydroxylase enzyme which uses Δ^5 as substrate for the lyase activity (Payne and Hales 2004). Testosterone, which is secreted from theca-interstitial cells of the ovary, is another androgen and is formed by the reduction of the 17-keto group of androstenedione (Berg et al. 2002; Payne and Hales 2004). This reaction is catalyzed by 17β-hydroxysteroid dehydrogenases (Payne and Hales 2004). The estrogens, estrone and estradiol (E2), are synthesized from androgens by the loss of the C-19 methyl group (Berg et al. 2002). Testosterone and androstenedione can be further metabolized to estradiol and estrone, respectively, in the ovary in a reaction catalyzed by the aromatase enzyme (CYP19A1) (Berg et al. 2002; Hu et al. 2010). The ovarian granulosa cells secret progesterone (P4) and estradiol, and ovarian theca cells predominantly synthesize androgens.

2.3.3. Mammary gland

The mammary gland synthesizes and secretes a large number of products in the milk including proteins (whey 20% and casein 80%), carbohydrates, coated lipid droplets, water and ions (Bauman et al. 2006). Milk fat is of major importance to the dairy industry, as it influences the manufacturing properties and other organoleptic qualities of milk and dairy products (Bauman et al. 2006; Bionaz and Loor 2008a). Several studies have defined and quantified major metabolic aspects of mammary lipid metabolism. These main lipid associated metabolic pathways are the ones involved in fatty acid uptake from the blood (through endothelial long-chain fatty acid transport), *de novo* fatty acid (FA) synthesis (in cytosol), FA synthesis in the mitochondria and milk lipid synthesis, droplet formation and secretion (in the endoplasmic reticulum (ER) membrane) (Bauman and Griinari 2003; Bauman et al. 2006; Bionaz and Loor 2008a). Fat production and milk FA composition are affected by the stage of lactation and level of production (Kay et al. 2005; Bernard et al. 2008; Bionaz and Loor 2008a). Transcriptional studies of the bovine mammary gland have highlighted a complex and coordinated set of molecular events that are involved in mammary adaptations to lactation (Lemay et al. 2007; Bionaz and Loor 2008a). This section will briefly review these molecular events from endothelial FA uptake to lipid droplet formation in the ER membrane.

2.3.3.1. Blood fatty acid uptake and de novo fatty acid synthesis

The mammary gland can use two sources of fatty acids for milk fat synthesis. One source is the *de novo* synthesized fatty acids produced by mammary epithelial cells; the other source is fatty acids that are obtained from the digestive tract or through mobilization of body reservoirs (Ma and Corl 2012). Short chain (4-8 carbons), medium chain (10-14 carbons) and a portion of long

chain fatty acids (16 carbons) are synthesized from acetate and β-hydroxybutyrate in the *de novo* FA synthesis process; the remaining long chain fatty acids (including the other half of 16 carbon FA and all FA longer than 16 carbons) are taken up from circulation by the mammary gland (Bauman and Griinari 2003). In ruminants, fatty acids are derived predominantly from intestinal absorption of dietary and microbial fatty acids (Bauman and Griinari 2003). Early in lactation, however, when the animal is in negative energy balance, the contribution from mobilized fatty acids (such as circulating lipoproteins and NEFA) increases (Bauman and Griinari 2001; Ma and Corl 2012). Mammary cells take up albumin-bound FA (or NEFA) and lipoproteins. The VLDL or chylomicrons are also anchored to mammary endothelium by lipoprotein lipase (LPL) which hydrolyzes triacylglycerol (TAG) in the lipoprotein to release the FA (Fielding and Frayn 1998). Most of these long chain fatty acids (LCFA) are then esterified with CoA to LC-acyl-CoA (LCACoA) in the inner face of the plasma membrane before participating in metabolic pathways (Bionaz and Loor 2008a). This step is regulated by the acyl-CoA synthetase long-chain family member 1 (*ACSL1*) gene which has been shown to be the most predominant among other acyl-CoA synthetase mRNA isoforms in the bovine mammary tissue during lactation (Bionaz and Loor 2008a; Bionaz and Loor 2008b). The *ACSL1* gene converts free long-chain fatty acids into fatty acyl-CoA esters (Fig. 1). Specific localization of *ACSL1* gene product in the plasma membrane, endoplasmic reticulum and the mitochondria-associated membrane supports channelling of LCFA and synthesis of TG from LCFAs (Coleman et al. 2002; Bionaz and Loor 2008a; Bionaz and Loor 2008b).

Triacylglycerol (TAG) synthesis and formation of milk lipid droplets

The activated long-chain fatty acids (LCACoA) bound to FABP3 (fatty acid binding protein 3) gene protein are used as substrate for the SCD (stearoyl-CoA desaturase) enzyme, which is

located on the ER membrane (Bionaz and Loor 2008a; Bionaz and Loor 2008b). SCD adds a double bond to the Δ^9 position of unsaturated fatty acids (myristoyl-, palmitoyl- and stearoyl-CoA) and triacylglycerol synthesis (TAG) begins through a series of sequential reactions carried out by the products of the *GPAM* (glyceol-3-phosphate acyltransferase), *LPIN1* (Lipin I) and *DGAT1* (Diacylglycerol acyltransferase I) genes (Fig. 1) (Bionaz and Loor 2008b). The first step in TAG synthesis is the acylation of glycerol-3-phosphate to form lysophosphatidic acid (LPA); this step is catalyzed by the glycerol-3-phosphate acyltransferase (GPAT) enzyme (Gimeno and Cao 2008). Then a fatty acid is transferred to LPA by LPA acyltransferase (also called AGPAT) enzyme to produce phosphatidate (PA) (Takeuchi and Reue 2009). The PA is then served as a precursor of diacylglycerol (DAG). Lipin enzyme (an endoplasmic reticulum enzyme) catalyzes this reaction (Table 2.1) (Reue and Dwyer 2008). Finally, DAG converts to TAG by way of the diacylglycerol acyltransferase (DGAT) enzyme (Shindou et al. 2008; Takeuchi and Reue 2009). The formed TAGs are enveloped by the ER plasma membrane and gradually move to the apical surface of the cell to the point that they dissociate from the cell (Keenan and Mather 2009). The bovine milk lipid droplet is dependent on the adipose differentiation related protein (adipophilin, ADFP) for differentiation from the ER membrane and the product of the butyrophilin, subfamily I, member AI (*BTN1A1*) gene for differentiation from cell membrane (Bionaz and Loor 2008a; Keenan and Mather 2009).

2.4. Endocrine adaptations in transition dairy cows

As a result of negative energy balance (NEB) early in lactation, major changes in hormonal regulation occur in high-yielding dairy cows (Djoković et al. 2014). This involves changes in concentrations of key hormones as well as tissue responsiveness. For example, an increase in lipolysis and decrease in lipogenesis occur in order to maintain physiological equilibrium of the

body and to satisfy the needs of the mammary gland through nutrient redistribution (Bauman and Currie 1980; Bauman and Griinari 2001; Roche et al. 2009; Djoković et al. 2014). Blood hormone concentrations have an important role in mammary gland development and lactogenesis during the periparturient period (Bauman and Currie 1980). Pituitary growth hormone (GH), the thyroid gland hormones, insulin, cathecholamines and leptin are some examples of the endocrine factors regulating lipid metabolism (Bauman and Currie 1980; Roche et al. 2009; Djoković et al. 2014).

The physiological effects of growth hormone are initiated when it binds to GH receptors (GHR) on target cells. Growth hormone enhances the lipolytic response of adipose tissue to β-adrenergic-signals and is reported to have a positive effect on hormone-sensitive lipase (HSL) activity in adipose tissue (Etherton and Bauman 1998; Roche et al. 2009). Binding of GH to its receptors (GHR-1A) in the liver initiates synthesis and secretion of insulin-like growth factor 1 (IGF-1) (Roche et al. 2009). Despite the increase in plasma GH concentrations early in lactation, the abundance of hepatic GH receptors decreases, and as a result plasma IGF-1 also decreases (Block et al. 2001; Lucy 2001; Roche et al. 2009). Decreasing liver GHR abundance initiates lipolysis (Roche et al. 2009). Since the concentration of IGF-1 does not fluctuate with feeding activity, it is a good indicator of nutritional status (Taylor et al. 2004). An optimum concentration of IGF-1 to maintain enough of a pool of circulating IGF-1 and its widespread actions is achieved by six binding proteins (IGFBPs 1 to 6) and the acid-labile subunit (ALS) (Rhoads et al. 2004; Taylor et al. 2004). In addition, many members of the somatotropic axis (hypothalamo-pituitary axis) are expressed locally within endometrium (Wathes et al. 2011). For example, IGF-1 and IGF-2 act through type 1 IGF receptor (IGF1R) and are also expressed in the postpartum uterus (Wathes et al. 2011). *IGF1* and *IGF2* are expressed in many organs of the

body and have an influence on proliferation, differentiation and metabolic activities. These genes may therefore play a role in uterine involution (Llewellyn et al. 2008; Wathes et al. 2011).

Insulin has a regulatory effect on lipogenesis and is an antagonist to the lipolytic actions of GH (Vernon et al. 2001). Hypoinsulinemia (low concentrations of insulin in the blood) and a decrease in responsiveness of skeletal muscle and adipose tissue to insulin occurs simultaneously in early lactation and leads to an insulin-independent uptake of the available glucose by the mammary gland and greater body lipid mobilization to the liver (Bell and Bauman 1997; Roche et al. 2009). This process begins with lower insulin concentration and elevated placental lactogen in the uterus during late pregnancy which stimulates adipose metabolism to provide nutrients for the growing fetus (Sivan and Boden 2003).

Leptin, secreted from adipocytes, decreases immediately postpartum as a consequence of energy deficit (Leury et al. 2003; Roche et al. 2009). This reduction in leptin production matches the plasma insulin profile early in lactation and is consistent with reduced adipose tissue glucose uptake (Leury et al. 2003; Roche et al. 2009). Leptin hormone secretion is regulated by a complex of different molecules and hormones such as insulin, glucocorticoids and cytokines (tumor necrosis factor (TNF) α), unterleukin-1 (IL-1), catecholamines, testosterone and PPARγ (Vernon et al. 2001).

Catecholamines, such as epinephrine and norepinephrine, act as lipolytic stimulators through activating cAMP and then PKA which activates subunits of both HSL and perilipin proteins that subsequently increase lipolysis (Fig. 1) (Berg et al. 2002; Roche et al. 2009). Perilipin phosphorylation, which occurs through a cAMP-dependent PKA cascade, is essential for

translocation of HSL hormone from cytosol to the surface of the lipid droplet (Roche et al. 2009). It has been reported that the transcription of the genes producing perilipin, ß-adrenergic receptors and HSL in adipose tissue increase early in lactation in dairy cows (Sumner and McNamara 2007). In addition, the responsiveness of bovine adipose tissue to catecholamines increases in early lactation (McNamara 1988).

Thyroid hormones have an important role in the dairy cattle transitional period and in determining cell metabolism intensity, metabolism of lipids and carbohydrates and lactation course in general (Djoković et al. 2007). These hormones are known for their importance in milk production through stimulation of metabolic rates with other hormones (Blum et al. 1983). It has been shown that there is a positive correlation between thyroid hormones in blood and energy metabolism (Reist et al. 2002). During negative energy balance and high lipid metabolism, however, the concentrations of T4 (thyroxine) and T3 (triiodothyronine which is four times more active than T4) are reduced (hypothyroidism) in the blood of dairy cows shortly before and after calving (Blum et al. 1983; Reist et al. 2002; Pezzi et al. 2003). Negative energy balance and an increase in lipid mobilization and hypothyroidism early in lactation in dairy cows are accompanied by metabolic disorders associated with carbohydrate and lipid metabolism such as ketosis and fatty liver (Djoković et al. 2014). Therefore, thyroid hormones are considered to be important indicators of homeorhetic adaptation to negative energy balance in dairy cows until energy balance is achieved (Djoković et al. 2007; Kasagic et al. 2011; Djoković et al. 2014).

2.5. Reproductive endocrinology and hormonal adaptations in cows in negative energy balance (NEB) stage

Reproductive function in dairy cattle is dependent on balanced and coordinated endocrine activity (Wathes et al. 2011). This includes homeostasis between different reproduction hormones such as gonadotrophin releasing hormones (GnRH), follicle stimulating hormone (FSH), luteinizing hormone (LH), prolactin (PRL) and gonadal steroids as well as prostaglandin F2 alpha ($PGF_{2\alpha}$) from the uterus (Leslie 1983). Shortly before parturition, $PGF_{2\alpha}$ increases and luteolysis of corpus luteum (CL) begins. Progesterone decreases rapidly as a result of CL regression. Plasma estrogen concentration drops immediately after calving to the values below those found during the normal estrous cycle (Jorritsma et al. 2003).

It has been shown that NEB is a reason for poor conception rates in transition dairy cows (Wathes et al. 2003; Wathes et al. 2007; Fenwick et al. 2008). The relationship between NEB and fertility stems from the effects of NEB on the resumption of cyclicity and on the quality of oocyte or the corpus luteum which is discussed in below.

Ovarian cyclicity and ovulation after calving are necessary for a successful insemination and cows should be able to ovulate approximately 2-3 months postpartum (Jorritsma et al. 2003). Gonadotropins play an important role in the onset of this activity. FSH concentrations remain at relatively constant levels throughout the post-pubertal life of a dairy cow, but LH concentrations become pulsatile in cyclic animals (Leslie 1983). Inhibition of LH pulsatility before puberty leads to decreased ovarian activity (Jorritsma et al. 2003). Besides, the onset of puberty has been shown to be associated with attainment of a critical level of body fat (Schillo et al. 1992;

Jorritsma et al. 2003). Therefore, it has been suggested that there is a relationship between the metabolic status of the animal and reproductive endocrine system of dairy cow not only for onset of puberty in heifers, but also for resumption of ovarian activity in non-adoptive dairy cows early in lactation (Schillo et al. 1992). In this regard, LH seems to have a more important role than FSH after calving (Stagg et al. 1998). Low glucose concentrations have also been associated with less amplitude of LH pulses. In addition, endogenous opioid peptides, which are secreted during stress, have negative effects on LH pulsatility and the onset of ovarian activity post-partum (Ahmadzadeh et al. 1998; Jorritsma et al. 2003).

The IGF-1 system is thought to influence the establishment and maintenance of pregnancy of dairy cows through affecting reproductive tract of cows (Taylor et al. 2004). The IGF-1 protein acts as a co-gonadotroph and amplifies the effects of FSH and LH on the growth and differentiation of ovarian follicles (Lucy 2001; Taylor et al. 2004). The IGF-1 system also plays an important role in the survival of the embryo and its plasma concentrations was shown to be associated with longer calving to conception intervals (Wathes et al. 2003; Taylor et al. 2004). It has been reported that regulation of *IGF1* and *IGF2* is positively associated with ovarian oestradiol production (Stevenson and Wathes 1996; Jorritsma et al. 2003; Fenwick et al. 2008). The concentration of IGF-1 increases notably at a time of increasing oestradiol dominance in the bovine oviduct (Stevenson and Wathes 1996). Moreover, IGF-1 and insulin have a stimulatory effect on ovarian granulosa cells, increasing proliferation, as well as progesterone and oestradiol production. Insulin and IGF-1 also stimulate androgen production in (ovarian) theca cells (Spicer and Echternkamp 1995; Jorritsma et al. 2003). Recent studies have further suggested the role of thyroid hormones in the onset of ovarian activity (Van den Top et al. 1995; Jorritsma et al. 2003).

The quality of oocytes at the time of insemination is important in non-adopting dairy cows and is dependent on the sufficient number of ovarian cycles and the time that an antral follicle needs to reach its ovulatory size after calving (Jorritsma et al. 2003). Several factors and hormones can affect the quality of oocytes. IGF-1 and its binding proteins might affect the quality of oocytes. Follicular development can be inhibited with increased IGFBPs that are known to function as IGF-1 inhibitors (Jorritsma et al. 2003). The second is the metabolic status of the dairy cow. As a result of an increase in body fat mobilization and to some extent body protein mobilization, plasma urea concentrations increase early in lactation (Jorritsma et al. 2003). The increase in ammonia concentrations may also occur as a result of accumulation of triacylglycerides and inhibition of ureagenesis during the transition period in dairy cows (Zhu et al. 2000). A high concentration of circulating urea and ammonia in the bloodstream of cows is associated with reduced fertility (Rhoads et al. 2004; Laven et al. 2007; Wathes et al. 2011). Exposure of oocytes in antral follicles to high levels of ammonia concentrations during fertilization may hamper cleavage and blastocyst formation (Sinclair et al. 2000). Increased urea concentration in the blood early in lactation is associated with declined cleavage ratios and blastocyst formation of the fertilized embryo (Jorritsma et al. 2003; Wathes et al. 2011). This increased level of urea concentration in the blood after calving may also influence the expression of endometrial IGF and insulin receptor (INSR) (Wathes et al. 2011). Wathes et al. (2011) reported that expression of *IGF1R* and *INSR* was not altered by the energy balance status of the dairy cow early in lactation but was positively correlated with the circulating urea concentration (Wathes et al. 2011).

2.6. Regulatory components and genes mediating metabolic and endocrine adaptations

Differences in the success of adaptation early in lactation between cows, under the same conditions and similar production level, suggest that adaptability may have a genetic basis (Van Dorland et al. 2009; Kessler et al. 2014; Ha et al. 2015). Many genes, pathways and key candidate metabolites in the plasma have been previously confirmed to be essentially involved in the regulation of metabolic and endocrine adaptations in dairy cow (Dann and Drackley 2005; Bionaz and Loor 2008b; Wathes et al. 2011; Ha et al. 2015). However, these genes and pathways might be expressed only at a certain point of time in the individual (Ha et al. 2015). Some genes, for example those affecting glucose levels, might be expressed in early lactation and others, affecting the abundance of non-esterified fatty acids (NEFA) for example, are expressed 4 weeks before or 13 weeks after calving (Ha et al. 2015). Identifying the genes and pathways regulating important biological functions during specific physiological states of dairy cattle may help in the identification of DNA variants that affect milk production and subsequent fertility (Snelling et al. 2013).

2.6.1. Genes and key pathways in multiple tissues

The onset of lactation in dairy cows is accompanied by an increase in milk synthesis and nutrient requirements, and eventually there is metabolism adaptation to lactation-associated challenges. These adaptations include metabolism adjustments in liver and peripheral tissues (including adipose tissue, mammary gland, skeletal muscle tissues and kidney), and mobilization of body reserves and increased lipid metabolism (Weikard et al. 2012). The increase in lipid metabolism results in an increase in concentrations of key metabolites NEFA and BHBA in plasma (Figure

1.1), and TAG in liver (Roche et al. 2009; Schlegel et al. 2012). Several genes and pathways in multiple tissues are involved in regulating these metabolites in lactating dairy cows (Mandard et al. 2004; Carlson et al. 2007; Bionaz and Loor 2008a; Ling and Alcorn 2008; Bionaz and Loor 2008b; Schlegel et al. 2012; Ha et al. 2015). A gene-based mapping and pathway analysis indicated that three pathways (steroid hormone biosynthesis, ether lipid metabolism and glycerophospholipid metabolism) jointly affect the concentrations of NEFA, BHBA and glucose in cows during the transition period (Ha et al. 2015). The key genes that are involved in regulating energy metabolism in multiple tissues include *PPRA, PCK1, PCK, ACACA, FASN, FBP2*, *FABP3*, *PPARGC1A*, *ACSL1*, *PPARGC1A*, *AGPAT6*, *PCCA*, *LPIN1, ACO*, *CPT-I*, *CPT-II*, and *ACSL* (Chmuzynska 2006; Rudolph et al. 2007; Loor et al. 2007; Bionaz and Loor 2008b; Baik et al. 2009; Weikard et al. 2012) These genes are involved in fatty acid uptake (mainly in the liver and mammary gland), mitochondrial and peroxisomal fatty acid oxidation, ketone body metabolism (ketogenesis), and cholesterol metabolism (in liver) early in lactation in dairy cattle (Schlegel et al. 2012) and are discussed in the following tissue-specific sections in more detail.

Gene expression studies are revealing the extent to which different genes are involved in different tissues (Baik et al. 2009; Schlegel et al. 2012; Weikard et al. 2012; Zhang et al. 2015). For example, genes that are involved in carbohydrate metabolism, such as those encoding gluconeogenesis and propionate metabolism enzymes (including *PCK1*), were expressed more in liver than mammary tissues (Baik et al. 2009; Zhang et al. 2015). However, the related *PCK2* gene shows a small difference in expression between mammary gland and liver (Baik et al. 2009; Weikard et al. 2012; Zhang et al. 2015). Other studies showed that the *PCK2* gene might also be active in glyceroneogenesis in lipogenic tissues (adipose tissues) during fasting or restricted feed intake (Hanson and Reshef 2003; Reshef et al. 2003) and in the epithelial cells of mammary

tissue during lactation (Hsieh et al. 2011). Weikard et al. (2012) reported that expression of the *PPARGC1A* gene was significantly increased in liver, mammary gland and skeletal muscle in lactating cows. The *PPARGC1A* gene coordinates expression of several proteins and in this way it controls the regulation of several metabolic pathways in response to metabolic challenges (Weikard et al. 2012). This gene has been reported to have a pivotal role in hepatic glucose synthesis (gluconeogenesis) (Puigserver and Spiegelman 2003), to be a key gene in mitochondrial oxidative phosphorylation metabolism (Mootha et al. 2003; Patti et al. 2003), and to independently regulate the expression of several lipogenic genes after the onset of lactation in dairy cattle (Bionaz and Loor 2008a). In a study comparing gene expression patterns between liver, mammary gland and skeletal muscle tissues in lactating cows, Weikard et al. (2012) indicated that the *PPARGC1A* and *PCCA* genes display a significantly altered mRNA abundance between the tissues and across all the cow groups under investigation: cows with different genetic potential for milk performance (high milk performance, medium and low milk performance) and cows with different genetic backgrounds (purebred and combined beef, dairy background). Fatty acid binding proteins (FABP) are the main transporters of long chain fatty acids (LCFA) to specific organelles for metabolism (Mcarthur et al. 1999). Different isoforms of FABPs have shown unique patterns of tissue-specific gene expression (Frolov et al. 1997) and are most abundant in tissues that are involved in active lipid metabolism (Baik et al. 2009). In this regard, FABP3 was shown to have a major role in bovine mammary gland lipid synthesis and is much more abundant in this tissue (Bionaz and Loor 2008a; Baik et al. 2009). FABP1 was shown to be more abundant in liver (Chmuzynska 2006), whereas expression of FABP4 was reported to be greater in mammary and adipose tissues (Hunt et al. 1986; Baik et al. 2009). There are nine isoforms of 1-acylglycerol-3-phosphate O-acyltransferase (AGPAT) in mammals, an

enzyme that catalyzes the transfer of fatty acyl-CoA to lysophosphatidic acid (Yamashita et al. 2007). The AGPAT1 isoform is the most abundant one in both liver and mammary tissues (Baik et al. 2009), but a knockout study in mice suggests an important roll for AGPAT6 as well in mammary tissues, in the biosynthesis of milk fat (Beigneux et al. 2006).

2.6.2. Genes and key pathways regulating liver lipid and cholesterol metabolism in transition dairy cow

NEFA, beta-hydroxybutryate (BHBA) and glucose are key factors in the metabolic status of transition dairy cows (Van Dorland et al. 2009; Graber et al. 2012; Ha et al. 2015). Ha et al. (2015) reported that several pathways jointly regulate concentrations of these metabolites, including three highly significant pathways: steroid hormone biosynthesis, ether lipid metabolism and glycerophospholipid metabolism. Several genes are associated with these pathways including *CD53*, *ABCC1*, *ADCYAP1R1*, *ZNF551*, *AHCYL1*, *WWC1* and *MED19* (Ha et al. 2015). Ha et al. (2015) also reported similar links to pathways and genes associated with cholesterol metabolism and NEFA concentrations in dairy cows. These results are in agreement with Kessler et al. (2014) who showed that mRNA abundance of genes involved in cholesterol synthesis (*SREBF2*, *HMGCS1* and *HMGCR* and *ABCG1*) markedly increased early in lactation (Horton et al. 1998; Kessler et al. 2014).

The high concentration of NEFAs early in lactation can act as signaling molecule, regulating the expression of hepatocyte genes that are involved in lipid metabolism (Duplus et al. 2000; Jump et al. 2005; Li et al. 2013). In this regard, AMP-activated protein kinase (AMPK) signaling pathways have been shown to be a key regulator of hepatic lipid metabolism, responding to hormones and metabolites including NEFAs (Li et al. 2013). AMPK acts as a mediator for

expression of transcriptional factors, peroxisome proliferator-activated receptor α (PPARα), sterol regulatory element-binding protein-1c (SREBP-1c) and carbohydrate responsive element-binding protein (ChREBP) (Viollet et al. 2006; Li et al. 2013). Activation of these transcriptional factors leads to expression of lipolytic and lipogenic genes (Zhou et al. 2001; Kawaguchi et al. 2002; Li et al. 2013). Activated PPRAα can increase expression of lipolytic genes (*ACO*, *CPT-I*, *CPT-II*, *L-FABP* and *ACSL*) and subsequently lipid oxidation (Li et al. 2013). This is while AMPKα inhibits transcription factors (SREBP-1c and ChREBP) which decrease expression of lipogenic genes (*ACC1*, *FAS* and *SCD-1*) and eventually lipid synthesis (Fig. 1) (Li et al. 2013). In addition, CPT-I activity increases by activated AMPKα; activated CPT-I increases downstream hepatic enzymatic activity and lipid metabolism (Li et al. 2013).

One of the important genes that control synthesis of sterols is *SREBF2* (Horton et al. 1998). Kessler et al. (2014) showed that there is a significant correlation between *SREBF2* mRNA expression and the hepatic gene expression of both 3-hydroxy-3-methylglutaryl-CoA reductase (*HMGCR*) and *HMGCS1*. In addition, several investigations showed that occurrence of fatty liver syndrome early in lactation is associated with cholesterol carrier lipoproteins such as VLDL (Kessler et al. 2014; Gross et al. 2015). Cholesterol is transported by high density lipoproteins (HDL) from peripheral tissues to the liver (Kessler et al. 2014). The *ABCA1* gene regulates formation of HDL. Furthermore, Viturro et al. (2009) reported a maximum increase in the expression levels of two transcription regulatory proteins, SREBP1 and SREBP2, on the week 2 postpartum that was coordinately and significantly correlated with an increase in expression levels of the enzymes 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) and 3-hydroxy-3-methylglutaryl-CoA synthesis (Viturro et al. 2009). These results indicate that there are complex regulatory mechanisms involved in homeostasis of cholesterol in transition dairy cows.

2.6.3. Genes and key pathways regulating milk fat synthesis in mammary gland in transition dairy cow

Early in lactation, fatty acid (FA) uptake from blood predominates relative to *de novo* FA synthesis in the mammary gland (Bionaz and Loor 2008a). This change in milk FA synthesis is mediated by up-regulation of genes and pathways that are associated with FA uptake from blood (such as *LPL*, *CD36*) and intracellular transport/channelling genes (such as *FABP3*) (Bionaz and Loor 2008a). Moreover, early in lactation, significant up-regulation of other genes and pathways has been observed, such as up-regulation of genes responsible in activation of fatty acids (e.g., *ACSL1*, *ACSS2*), *de novo* FA synthesis (e.g., *ACACA*, *FASN*), FA desaturation (e.g., *SCD*, *FADS1*), synthesis of TAG (e.g., *AGPAT6*, *GPAM*), lipid droplet formation (e.g., *BTNA1*) and ketone body utilization (e.g., *BDH1*, *OXCT1*) (Bionaz and Loor 2008a). Bionaz and Loor (2008) also suggested that expression of *SREBF1* is central to milk fatty acid synthesis and that the genes *PPARG*, *LASS2*, *INSIG1* and *OSBP* have a pivotal role in regulating lipid synthesis and mammary intracellular equilibrium between cholesterol and sphingolipids in lactating dairy cows.

2.6.4. Genes and key pathways regulating lipid metabolism in adipose tissue in transition dairy cow

With increasing milk production and onset of a period of negative energy balance, the expression of many genes and enzymes controlling lipid metabolism in adipose tissue changes (Khan et al. 2013; Li et al. 2013). These changes favour a decrease in lipogenesis and an increase in lipolysis. Khan et al. (2013) reported a decrease in expression of genes controlling adipogenesis including *PCK1*, *FASN*, *SCD*, *DGAT2*, *PPRAG*, *WNT10B* and *SREBF1*. These results are in agreement

with previous work by Sumner and McNamara (2007) and Bionaz and Loor (2012) which reported that adipose lipogenesis in cows during early lactation is primarily regulated through control of gene expression. Expression of key lipolytic enzyme genes (*LIPE*, *PNPLA2*, *MGLL* and *ADRB2*) followed a similar pattern early in lactation, indicating that the control of lipolysis in the adipose tissue is likely controlled by post-transcriptional events (McNamara and Murray 2001; Vernon et al. 2001; Khan et al. 2013). Post-transcriptional activation of HSL through simulation of the β-2-adrenergic receptor and the phosphorylation cascade has been shown to be the first step in beginning of lipolysis and providing fatty acids to the mammary gland and other tissues (Sumner-Thomson et al. 2011; Khan et al. 2013). The transcription of other lipolysis genes (*LIPE*, *PLIN1* and *ADRB2*) increases following an increase in the enzymatic capacity for continued supply of FA to other organs and rebuilding adipose stores (McNamara 1988). Adipose triglyceride lipase (PNPLA2) has been reported to be a highly-expressed lipolytic enzyme in the white adipose tissue of dairy cattle, which is associated with basal and β-2-adrenergic-simulated triacylglycerol hydrolysis (Miyoshi et al. 2007).

2.6.5. Influence of energy balance and metabolites early in lactation on gene expression in the endometrium of postpartum dairy cow

It has been reported that severe negative energy balance (SNEB) in high-producing postpartum dairy cows is associated with subsequent low fertility (Wathes et al. 2011). Excessive lipid metabolism, increased concentrations of NEFAs and BHB and reduced concentrations of glucose and IGF-1 are associated with reproductive disorders and poor conception rates (Bauman and Griinari 2003; Wathes et al. 2007). The failure of multiparous cows to conceive is correlated with low IGF-1 circulation in the first two weeks postpartum (Taylor et al. 2004; Wathes et al.

2011). It has been shown that expression of IGF-binding protein 4 (*IGFBP4*) and inflammatory response genes including matrix metalloproteinases (*MMP1*, *MMP3*, *MMP9* and *MMP13*), chemokines, cytokines and calgranulins significantly increase in the endometrium as a result of metritis in cows with SNEB (Wathes et al. 2009; Wathes et al. 2011). Wathes et al. (2011) also reported that the expression of hormone receptors in the endometrium (*IGF1R*, *IGF2R*, *INSR*, *GHR*, *NR3C1*, *ESR1* and *ESR2*) did not change according to the energy balance status and that there is a coordinated expression between hormone receptors *IGF1R*, *IGF2R* and *INSR* as well as *GHR* with *ESR1* and *NR3C1* with *ESR2* (Wathes et al. 2011). Furthermore, increased concentrations of blood urea as a result of dietary factors and tissue protein catabolism may influence the expression of endometrial *IGF* and *INSRs* (Wathes et al. 2011).

2.7. Candidate genes identified in the key tissues through association analysis for production and fertility traits

The availability of highly informative marker maps, genome-wide association analysis (Dekkers 2004), gene-based mapping (an association approach that tests each gene instead of each SNP separately as described in Ha et al (2015)) and pathway analysis (Ha et al. 2015) have resulted in the identification of several crucial regulated target genes and metabolic pathways in the mammary gland, liver and blood plasma that are responsible for the regulation of the metabolism early in lactation. For example, a QTL with a major effect on milk yield and composition has been identified on the centromeric end of the chromosome 14, and involves the *DGAT1* gene (Grisart et al. 2002). Figure 2 shows a strong association of SNPs with milk production on chromosome 14 close to the *DGAT1* gene in a genome-wide association study done on Canadian Holstein dairy cattle (Nayeri et al. 2016). Similarly, linkage disequilibrium (LD) analysis

highlighted a chromosomal region on bovine chromosome 20 harboring the *GHR* gene which affects milk yield and composition (Georges et al. 1995; Arranz et al. 1998; Blott et al. 2003). Another association analysis revealed highly significant SNPs (false discovery rate at P-value ≤ 10^{-8}) associated with fat and protein percentage on chromosome 19 residing within *ACLY*, which is a fatty acid biosynthesis gene (Raven et al. 2014b). Significant associations involving markers within or close to other fat metabolism associated genes such as *FASN*, *SREBPB1* and *STAST5A* have also been reported for milk production traits in dairy cattle (Bouwman et al. 2011). In a recent study of a German Holstein-Friesian population, two highly significant polymorphisms were found to be associated with milk fat content; one of these variants is located within the promoter region of the *EPS8* gene on chromosome 5 and the other variant is located near the *GPAT4* gene on chromosome 27 (Wang et al. 2012).

The product of the *EPS8* gene provides a substrate for receptor tyrosine kinases and physically interacts with the epidermal growth factor receptor (EGFR) (Fazioli et al. 1993). Interaction of the *EPS8* gene product with EGFR increases the signaling response to epidermal growth factor (EGF) (Raven et al. 2014b). The promoter SNP reported by Wang et al. (2012) in *EPS8* may mediate the binding of transcription factor TFAP2A to influence the transcription rate of *EPS8*. The expression of *TFAP2A* is correlated with the concentration of NEFA and liver triacylglycerol (Wang et al. 2012). It has been demonstrated that sterol regulatory element-binding proteins (SREBPs), which control the expression of genes required for the uptake and synthesis of cholesterol, fatty acid and triglycerides, are regulated by the epidermal growth factors (Chatterjee et al. 2009). Therefore, it is plausible that an increased milk fat biosynthesis in the lactating mammary gland is the result of an enhanced transcription rate of *EPS8*, conferred by binding of TFAP2A. The *GPAT4* gene is near a QTL region reported to contribute to the

genetic variation of milk fatty acid composition in the Dutch Holstein population (Bouwman et al. 2011; Wang et al. 2012). This gene plays an important role in lipid biosynthesis in mammals. The transcription rate of *GPAT4* is highly correlated with concentrations of diacylglycerols and triacylglycerols in milk (Beigneux et al. 2006; Bionaz and Loor 2008b). In a more recent study, this region on chromosome 27 was reported for associations with milk fat and milk volume, protein and lactose content in Holstein and Jersey crossbreds (Littlejohn et al. 2014). Mullen et al. (2011) have detected several novel and previously identified associations involving variants within introns of the *IGF1* gene associated with milk protein yield, milk fat yield, milk fat concentration, somatic cell score and carcass associated traits in Holstein dairy cattle (Mullen et al. 2011). The IGF-1 protein stimulates protein synthesis in the epithelial cells of the mammary gland and plays an important role in mammary gland growth and function (Burgos and Cant 2010). A more comprehensive list of major candidate genes associated with milk production traits is shown in Table 2.2.

The impact of poor fertility in the dairy industry has led to the inclusion of a female fertility index in some breeding programs and has undoubtedly contributed to interest in identifying the loci affecting fertility (Höglund et al. 2009; Höglund et al. 2012; Höglund et al. 2014a; Höglund et al. 2015). Association and QTL mapping studies have identified several candidate genes affecting fertility traits including interval from calving to first insemination, days from first to last insemination, 56-day non-return rate and insemination per conception. For example, an investigation of QTL regions affecting female fertility traits in Nordic Holstein cattle identified a strongly associated missense mutation within the multifunctional *CD36* gene on chromosome 4 (Höglund et al. 2014a). In two other association studies in Nordic Red dairy cattle using 50K SNP genotypes imputed into whole genome sequencing data, several other genes including

SLC6A17, *SDS5*, *ADCY1*, *SLC1A4* and *PPM1B* associated with cow and heifer non-return rate, calving to first service interval, number of inseminations per conception and days from first to last insemination were identified (Höglund et al. 2014b; Höglund et al. 2015). The *TGFB2*, *APOH* and *IGLL1* genes were reported as important candidate genes under significant peaks associated with non-return rate and days to first service in Italian Holstein cattle (Minozzi et al. 2013). The genes *TGFB2* and *APOH* are both involved in the process of the follicular development as they interact with the reproductive hormones LH and FSH (Minozzi et al. 2013). The *TGFB* isoforms can stimulate FSH receptor expression and amplify progesterone production and LH receptor induction (Knight and Glister 2006). The *IGLL1* gene has also been reported to be up regulated during the peripartum period in the endometrium of the lactating dairy cow and may play an important role in energy balance by influencing production and fertility traits at the same time (Cerri et al. 2012). A list of major candidate genes identified through association studies for fertility-associated traits is provided in Table 2.3.

2.8. Genomic regions and genes affecting multiple traits in dairy cattle

Milk production and fertility traits are polygenic, affected by many genes and variants, each with a small effect on the observed phenotype (Snelling et al. 2013; Nayeri et al. 2016). GWAS studies of different production and fertility traits in dairy cattle have identified shared quantitative regions and candidate genes—regions that appear to influence multiple traits (Chmuzynska 2006; Höglund et al. 2009; Reverter and Fortes 2012; Nayeri et al. 2016). In some cases the effects are confined to multiple production traits. For example, the underlying genomic region on chromosome 14 that includes *DGAT1* gene has been shown to have a major effect on milk fat content and several other production traits including milk yield, fat percentage and protein percentage (Ashwell et al. 2004; Pimentel et al. 2011; Maxa et al. 2012; Meredith et al.

2012; Xie et al. 2014; Nayeri et al. 2016). Similarly, several studies have reported associations of SNPs on chromosome 20 surrounding the *GHR* gene with milk yield, protein yield and protein percentage (Blott et al. 2003; Viitala et al. 2006; Meredith et al. 2012; Chamberlain et al. 2012; Raven et al. 2014b; Nayeri et al. 2016). More recently, variants close to the *MGST1* gene on chromosome 5 have been shown to be associated with increased fat yield, protein percentage and lactose percentage, and a decrease in protein yield, lactose yield and protein volume (Littlejohn et al. 2016). Through the collection and analysis of gene expression data the authors demonstrate that a strong *MGST1* eQTL (expression QTL) likely underlies these associations, however the specific role of *MGST1* in regulating milk composition is not known (Littlejohn et al. 2016). The identification of genes that influence multiple production traits is not surprising given the shared underlying molecular mechanisms (Solovieff et al. 2013). For example, it has been shown that the main functional pathways that are regulated by the *K232A* polymorphism in *DGAT1* gene (associated with reduced milk production and increased milk fat yield) were related to cell energy metabolism (lipid biosynthesis, oxidative phosphorylation, electron transport chain), protein degradation and cell signaling (Mach et al. 2012). This might reflect the underlying biological pleiotropic effect, where a single casual variant is related to the variations in multiple traits as explained by Solovieff et al. (2013).

Regions and candidate genes associated with multiple fertility traits have been described. For example, chromosome 21 was shown to harbor a region overlapping among two fertility traits, calving to first service interval and days open, and a candidate gene in that region has been proposed, *FAM181A* (Nayeri et al. 2016). In another GWAS study in Nordic Red cattle, a shared significant SNP (rs43271631) on chromosome 1 was associated with multiple fertility traits such as fertility index, 56-day non-return rate, number of insemination per conception and days from

first to last insemination (Höglund et al. 2015). This SNP is located within an intron of the *TRPC1* gene, which was shown to regulate osteoblast formation in mice (Ong et al. 2013). Cole et al (2009) reported a common SNP (ss86324977) on chromosome 18 in an intronic region of the sialic acid binding Ig-like lectine-5 (*SIGLEC5*) gene, associated with sire and daughter calving ease, that was also reported to affect direct calving traits in multiple studies (Kuehn et al. 2003; Thomasen et al. 2008; Cole et al. 2009; Sahana et al. 2011). In humans this gene is expressed in the placenta and has been suggested to have a role in the initiation of parturition (Brinkman-Van Der linden et al. 2007). In a GWAS study for calving traits in Danish and Swedish cows, the majority of the identified QTL showed an effect on more than one calving trait (such as birth index, stillbirth, calving ease, calf survival and calving index) (Sahana et al. 2011). The sharing of regions among fertility traits can reflect the similarity of some of the assessment procedures. For example, the number of inseminations per conception (number of insemination) is related to the days from first to last insemination (which measure time between first and last insemination) (Höglund et al. 2015). It also likely reflects, as with the overlap among production traits, shared mechanisms.

Perhaps the most interesting and challenging genes and variants are those that affect production and fertility. The success in increasing production in high-producing dairy cows is accompanied by a decline in reproductive performance, first-service pregnancy rate and reproductive efficiency (Lucy 2001; Pryce et al. 2004; Veerkamp and Beerda 2007). Cows with higher milk production at day 56 postpartum were shown to have significantly a longer commencement of luteal activity postpartum and a shorter first luteal phase (Royal et al. 2002). There is an antagonistic relationship between production and fertility traits due to pleiotropic gene effects, linkage or complex physiological associations (Kadarmideen et al. 2000; Veerkamp and Beerda

2007). For example, a region with effects on both production and fertility traits was reported on chromosome 5 (Kolbehdari et al. 2009; Pimentel et al. 2011; Guo et al. 2012; Saatchi et al. 2013; Nayeri et al. 2016). The significant variants identified in this region (at 87 to 100 Mb on chromosome 5) were reported to be associated with C22:1 milk fatty acid content, milk fat yield (Guo et al. 2012; Saatchi et al. 2013), protein yield (Cochran et al. 2013), calving to first service interval (Nayeri et al. 2016) and sire conception rate in Angus, Brown Swiss and Holstein cattle (Peñagaricano et al. 2012; Höglund et al. 2014b). Several candidate genes were reported within this region including *ST8SIA1*, *ABCC9*, *GABARAPL1* and *SLO1C1* (Pimentel et al. 2011; Nayeri et al. 2016). The *ABCC9* gene is thought to form ATP-sensitive potassium channels in cardiac, vascular and non-vascular smooth muscles (Gene ID: 10060). This gene is reported as a potential inhibitor of human myometrial contractility (Curley et al. 2002) through opening ATP-sensitive potassium channels, flowing K^+ ion and reducing cellular excitability (Khan et al. 2001), and was speculated to be a candidate gene in dairy cattle for calving to first service interval (Nayeri et al. 2016). This gene has also been reported to be associated with protein yield in dairy cattle in Holstein cattle (Cochran et al. 2013; Nayeri et al. 2016). Olsen et al. (2011), reported a region on chromosome 12 significant for non-return rate in Norwegian Red cattle previously reported to be associated with several milk production traits (milk, fat and protein yield). They showed that the most significant SNP in this region had a positive effect on milk traits and a negative effect on non-return rate (mainly for cows returning to oestrus after insemination) (Olsen et al. 2011). A GWAS of fertility and production traits in Italian Holstein cattle revealed one SNP on chromosome 5 (at 199 Mb) associated with protein yield, calving interval, fertility index, angularity and days to first service (Minozzi et al. 2013). This SNP (BTA-27242-no-rs) was reported with a positive effect for protein yield but had negative effects on calving interval,

fertility index, days to first service and angularity, and is not located within any gene, however five genes (*DUSP6*, *POC1B*, *ATP2B1*, *C12orf12*, *EPYC*) were reported within 1 Mb of the SNP (Minozzi et al. 2013). In another recent GWAS study using imputed whole-genome sequenced data, Iso-Touru et al. (2016) identified SNPs in five genes (*ENSBTAG00000034643*, *GBF1*, *TMEM180*, *ACTR1B*, and *bta-mir-146b*) associated with both fertility and milk yield. *GBF1* and *bta-mir-146b* may influence fat yield through a gene network linked to lipid and carbohydrate metabolism and to reproduction through a network connected to inflammatory response and cell-to-cell signaling (Iso-Touru et al. 2016). The authors suggest that application of whole-genome sequence data in GWAS analysis along with gene-network and pathway information may help to better identify candidate genes and variations affecting multiple production and fertility traits and indicates possible pathways that correlate these traits.

2.9. Conclusion

Genetic selection in the dairy industry has contributed to impressive gains in productivity that will help address increasing demand for milk and milk products. Knowledge of the biology of lactation including the key tissues, metabolic pathways, hormones and genes that are involved can help researchers identify the underlying variants that contribute to phenotypic differences. Indeed, association studies have highlighted several polymorphisms potentially accounting for variation in production and fertility traits. Continued studies of gene function and expression in the context of lactation and reproduction in cattle and other species will likely improve our ability to identify causative genes and variants for these traits, and may eventually lead to more accurate approaches to genomic selection that work better across generations and breeds.

2.10. References

Ahmadzadeh A, Barnes M, Pearson R. (1998) Effect of Naloxone on Serum Luteinizing Hormone Concentration in Anovulatory Holstein Cows During the Early Postpartum Period. Domest Anim Endocrinol 15:177–181. doi: 10.1016/S0739-7240(98)00005-8

Arranz JJ, Coppieters W, Berzi P, et al (1998) A QTL affecting milk yield and composition maps to bovine chromosome 20: A confirmation. Anim Genet 29:107–115. doi: 10.1046/j.1365-2052.1998.00307.x

Ashwell MS, Heyen DW, Sonstegard TS, et al (2004) Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. J Dairy Sci 87:468–75. doi: 10.3168/jds.S0022-0302(04)73186-0

Baik M, Etchebarne BE, Bong J, VandeHaar MJ (2009) Gene expression profiling of liver and mammary tissues of lactating dairy cows. Asian-Australasian J Anim Sci 22:871–884.

Bauman DE, Currie WB (1980) Partitioning of nutrients during pregnancy and lactation: a review of mechanisms involving homeostasis and homeorhesis. J Dairy Sci 63:1514–1529. doi: 10.3168/jds.S0022-0302(80)83111-0

Bauman DE, Griinari JM (2001) Regulation and nutritional manipulation of milk fat: Low-fat milk syndrome. Livest Prod Sci 70:15–29. doi: 10.1016/S0301-6226(01)00195-6

Bauman DE, Griinari JM (2003) Nutritional regulation of milk fat synthesis. Annu Rev Nutr 23:203–227. doi: 10.1146/annurev.nutr.23.011702.073408

Bauman DE, Mather IH, Wall RJ, Lock AL (2006) Major advances associated with the

biosynthesis of milk. J Dairy Sci 89:1235–1243. doi: 10.3168/jds.S0022-0302(06)72192-0

Beigneux AP, Vergnes L, Qiao X, et al (2006) Agpat6—a Novel Lipid Biosynthetic Gene Required for Triacylglycerol Production in Mammary Epithelium. 47:734–744. doi: 10.1021/ja8019214.Optimization

Bell AW (1995) Regulation of Organic Nutrient Late Pregnancy Metabolism During Transition from to Early Lactation. J Anim Sci 73:2804–2819. doi: /1995.7392804x

Bell AW, Bauman DE (1997) Adaptations of glucose metabolism during pregnancy and lactation. J Mammary Gland Biol Neoplasia 2:265–278.

Berg J., Tymoczko J., Stryer L (2002) Biochemistry, 5th edition.

Bernard L, Leroux C, Chilliard Y (2008) Bioactive Components of Milk. Springer Science+BusinessMedia, Inc. New York, USA

Bionaz M, Loor JJ (2008a) Gene networks driving bovine milk fat synthesis during the lactation cycle. 21:1–21. doi: 10.1186/1471-2164-9-366

Bionaz M, Loor JJ (2008b) ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J Nutr 138:1019–1024. doi: 138/6/1019 [pii]

Block SS, Butler WR, Ehrhardt RA, et al (2001) Decreased concentration of plasma leptin in periparturient dairy cows is caused by negative energy balance. J Endocrinol 171:339–348. doi: 10.1677/joe.0.1710339

Blott S, Kim J, Moisio S, et al (2003) Molecular dissection of a quantitative trait locus: a

phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163:253–66.

Blum JW, Kunz P, Leuenberger H, et al (1983) Thyroid hormones, blood plasma metabolites and haematological parameters in relationship to milk yield in dairy cows. Anim Prod 36:93–104. doi: 10.1017/S0003356100039982

Bouwman AC, Bovenhuis H, Visker MH, van Arendonk JA (2011) Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet 12:43. doi: 10.1186/1471-2156-12-43

Brinkman-Van Der linden ECM, Hurtado-Ziola N, Hayakawa T, et al (2007) Human-specific expression of Siglec-6 in the placenta. Glycobiology 17:922–931. doi: 10.1093/glycob/cwm065

Burgos SA, Cant JP (2010) IGF-1 stimulates protein synthesis by enhanced signaling through mTORC1 in bovine mammary epithelial cells. Domest Anim Endocrinol 38:211–221. doi: 10.1016/j.domaniend.2009.10.005

Carlson DB, McFadden JW, D'Angelo A, et al (2007) Dietary l-Carnitine Affects Periparturient Nutrient Metabolism and Lactation in Multiparous Cows. J Dairy Sci 90:3422–3441. doi: 10.3168/jds.2006-811

Cerri RLA, Thompson IM, Kim IH, et al (2012) Effects of lactation and pregnancy on gene expression of endometrium of Holstein cows at day 17 of the estrous cycle or pregnancy. J Dairy Sci 95:5657–75. doi: 10.3168/jds.2011-5114

Chamberlain a J, Hayes BJ, Savin K, et al (2012) Validation of single nucleotide polymorphisms associated with milk production traits in dairy cattle. J Dairy Sci 95:864–75. doi: 10.3168/jds.2010-3786

Chatterjee S, Szustakowski JD, Nanguneri NR, et al (2009) Identification of Novel Genes and Pathways Regulating SREBP Transcriptional Activity. PLoS One 4:e5197. doi: 10.1371/journal.pone.0005197

Chmuzynska (2006) The multigene family of fatty acid-binding proteins (FABPs): Function, structure and polymorphism. J Appl Genet 47:39–48.

Cochran SD, Cole JB, Null DJ, Hansen PJ (2013) Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genet 14:49. doi: 10.1186/1471-2156-14-49

Cohen-Zinder M, Seroussi E, Larkin DM, et al (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res 15:936–944. doi: 10.1101/gr.3806705

Cole JB, VanRaden PM, O'Connell JR, et al (2009) Distribution and location of genetic effects for dairy traits. J Dairy Sci 92:2931–2946. doi: 10.3168/jds.2008-1762

Coleman RA, Lewin TM, Van Horn CG, Gonzalez-Baró MR (2002) Do long-chain acyl-CoA synthetases regulate fatty acid entry into synthetic versus degradative pathways? J Nutr 132:2123–2126.

Curley M, Cairns MT, Friel AM, et al (2002) Expression of mRNA transcripts for ATP-sensitive potassium channels in human myometrium. Mol Hum Reprod 8:941–945.

Dann HM, Drackley JK (2005) Carnitine palmitoyltransferase I in liver of periparturient dairy cows: effects of prepartum intake, postpartum induction of ketosis, and periparturient disorders. J Dairy Sci 88:3851–3859. doi: 10.3168/jds.S0022-0302(05)73070-8

Dekkers JC (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci. doi: 82/13_suppl/E313 [pii]

Djoković R, Cincović M, Kurćubić V, et al (2014) Endocrine and Metabolic Status of Dairy Cows during Transition Period. Thai J Vet Med Thai J Vet Med 44:59–66.

Djoković R, Šamanc H, Jovanović M, Nikolić Z (2007) Blood Concentrations of Thyroid Hormones and Lipids and Content of Lipids in the Liver of Dairy Cows in Transitional Period. Acta Vet Brno 76:525–532. doi: 10.2754/avb200776040525

Drackley JK (1999) Biology of dairy cows during the transition period: the final frontier? J Dairy Sci 82:2259–2273. doi: 10.3168/jds.S0022-0302(99)75474-3

Duplus E, Glorian M, Forest C (2000) Fatty acid regulation of gene transcription. J Biol Chem 275:30749–30752. doi: 10.1074/jbc.R000015200

Etherton TD, Bauman DE (1998) Biology of somatotropin in growth and lactation of domestic animals. Physiol Rev 78:745–761.

Fazioli F, Minichiello L, Matoska V, et al (1993) Eps8, a substrate for the epidermal growth factor receptor kinase, enhances EGF-dependent mitogenic signals. EMBO J 12:3799–808.

Fenwick MA, Llewellyn S, Fitzpatrick R, et al (2008) Negative energy balance in dairy cows is associated with specific changes in IGF-binding protein expression in the oviduct. Reproduction 135:63–75. doi: 10.1530/REP-07-0243

Fielding BA, Frayn KN (1998) Lipoprotein lipase and the disposition of dietary fatty acids. Br J Nutr 80:495–502. doi: 10211047

Frolov A, Cho TH, Murphy EJ, Schroeder F (1997) Isoforms of rat liver fatty acid binding protein differ in structure and affinity for fatty acids and fatty acyl CoAs. Biochemistry 36:6545–6555. doi: 10.1021/bi970205t

Georges M, Georges M, Nielsen D, et al (1995) Mapping Quantitative Trait Loci Controlling Milk Production in Dairy Cattle by Exploiting Progeny Testing. Genet Soc Am 139:907–920.

Gimeno RE, Cao J (2008) Thematic review series: glycerolipids. Mammalian glycerol-3-phosphate acyltransferases: new genes for an old activity. J Lipid Res 49:2079–2088. doi: 10.1194/jlr.R800013-JLR200

Goddard ME, MacLeod L., Kemper K., et al (2014) A Research Plan for the Identification of QTL. In: Proceeding, 10th World Congress of Genetics Applied to Livestock Production.

Goff JP, Horst RL (1997) Physiological changes at parturition and their relationship to metabolic disorders. J Dairy Sci 80:1260–1268. doi: 10.3168/jds.S0022-0302(97)76055-7

Graber M, Kohler S, Müller A, et al (2012) Identification of plasma and hepatic parameters related to metabolic robustness in dairy cows. J Anim Physiol Anim Nutr (Berl) 96:75–84.

doi: 10.1111/j.1439-0396.2010.01124.x

Grisart B, Coppieters W, Farnir F, et al (2002) Positional candidate cloning of a QTL in dairy cattle: identi cation of a missense mutation in the bovine DGAT gene with major effect on milk yield and composition. Genome Res 12:222–231. doi: 10.1101/gr.224202

Gross JJ, Kessler EC, Albrecht C, Bruckmaier RM (2015) Response of the Cholesterol Metabolism to a Negative Energy Balance in Dairy Cows Depends on the Lactational Stage. PLoS One 10:e0121956. doi: 10.1371/journal.pone.0121956

Grummer RR (2007) Strategies to improve fertility of high yielding dairy farms: Management of the dry period. Theriogenology 68:281–288. doi: 10.1016/j.theriogenology.2007.04.031

Guo J, Jorjani H, Carlborg Ö (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. doi: 10.1186/1471-2156-13-82

Ha N-T, Gross JJ, van Dorland A, et al (2015) Gene-Based Mapping and Pathway Analysis of Metabolic Traits in Dairy Cows. PLoS One 10:e0122325. doi: 10.1371/journal.pone.0122325

Hanson RW, Reshef L (2003) Glyceroneogenesis revisited. Biochimie 85:1199–1205. doi: 10.1016/j.biochi.2003.10.022

Heid HW, Keenan TW (2005) Intracellular origin and secretion of milk fat globules. Eur J Cell Biol 84:245–258. doi: 10.1016/j.ejcb.2004.12.002

Höglund JK, Buitenhuis AJ, Guldbrandtsen B, et al (2009) Overlapping chromosomal regions

for fertility traits and production traits in the Danish Holstein population. J Dairy Sci 92:5712–9. doi: 10.3168/jds.2008-1964

Höglund JK, Buitenhuis B, Guldbrandtsen B, et al (2015) Genome-wide association study for female fertility in Nordic Red cattle. BMC Genet 16:110. doi: 10.1186/s12863-015-0269-x

Höglund JK, Guldbrandtsen B, Lund MS, Sahana G (2012) Analyzes of genome-wide association follow-up study for calving traits in dairy cattle. BMC Genet 13:71. doi: 10.1186/1471-2156-13-71

Höglund JK, Sahana G, Brøndum RF, et al (2014a) Fine mapping QTL for female fertility on BTA04 and BTA13 in dairy cattle using HD SNP and sequence data. BMC Genomics 15:790. doi: 10.1186/1471-2164-15-790

Höglund JK, Sahana G, Guldbrandtsen B, Lund MS (2014b) Validation of associations for female fertility traits in Nordic Holstein, Nordic Red and Jersey dairy cattle. BMC Genet 15:8. doi: 10.1186/1471-2156-15-8

Horton JD, Shimomura I, Brown MS, et al (1998) Activation of cholesterol synthesis in preference to fatty acid synthesis in liver and adipose tissue of transgenic mice overproducing sterol regulatory element-binding protein-2. J Clin Invest 101:2331–2339. doi: 10.1172/JCI2961

Hsieh C-W, Huang C, Bederman I, et al (2011) Function of phosphoenolpyruvate carboxykinase in mammary gland epithelial cells. J Lipid Res 52:1352–62. doi: 10.1194/jlr.M012666

Hu J, Zhang Z, Shen W-J, Azhar S (2010) Cellular cholesterol delivery, intracellular processing

and utilization for biosynthesis of steroid hormones. Nutr Metab (Lond) 7:47. doi: 10.1186/1743-7075-7-47

Hunt CR, Ro JH, Dobson DE, et al (1986) Adipocyte P2 gene: developmental expression and homology of 5'-flanking sequences among fat cell-specific genes. Proc Natl Acad Sci U S A 83:3786–90. doi: 10.1073/pnas.83.11.3786

Iso-Touru T, Sahana G, Guldbrandtsen B, et al (2016) Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. BMC Genet 17:55. doi: 10.1186/s12863-016-0363-8

Jorritsma R, Wensing T, A.M. K, et al (2003) Metabolic changes in early lactation and impaired reproductive performance in diary cows. Vet Res 34:11–26. doi: 10.1051/vetres

Jump DB, Botolin D, Wang Y, et al (2005) Fatty Acid Regulation of Hepatic. J Nutr 135:2503–2506.

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

Kasagic D, Radojicic B, Gvozdic D, et al (2011) Endocrine and metabolic profile in Holstein and red Holstein heifers during peripartal period. Acta Vet Brno 61:555–565. doi: 10.2298/AVB1106555K

Kawaguchi T, Osatomi K, Yamashita H, et al (2002) Mechanism for fatty acid "sparing" effect on glucose-induced transcription: Regulation of carbohydrate-responsive element-binding protein by AMP-activated protein kinase. J Biol Chem 277:3829–3835. doi:

10.1074/jbc.M107895200

Kay JK, Weber WJ, Moore CE, et al (2005) Effects of week of lactation and genetic selection for milk yield on milk fatty acid composition in Holstein cows. J Dairy Sci 88:3886–3893. doi: 10.3168/jds.S0022-0302(05)73074-5

Keenan TW, Mather IH (2009) Intracellular origin of milk fat globules and the nature of the milk fat globule membrane. Adv Dairy Chem 2:137–171. doi: 10.1007/0-387-28813-9_4

Kemper KE, Reich CM, Bowman PJ, et al (2015) Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. Genet Sel Evol 47:1–17. doi: 10.1186/s12711-014-0074-4

Kessler EC, Gross JJ, Bruckmaier RM, Albrecht C (2014) Cholesterol metabolism, transport, and hepatic regulation in dairy cows during transition and early lactation. J Dairy Sci 97:5481–5490. doi: 10.3168/jds.2014-7926

Khan MJ, Hosseini a, Burrell S, et al (2013) Change in subcutaneous adipose tissue metabolism and gene network expression during the transition period in dairy cows, including differences due to sire genetic merit. J Dairy Sci 96:2171–82. doi: 10.3168/jds.2012-5794

Khan RN, Matharoo-Ball B, Arulkumaran S, Ashford ML (2001) Potassium channels in the human myometrium. Exp Physiol 86:255–264.

Khatkar MS, Thomson PC, Tammen I, Raadsma HW (2004) Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genet Sel Evol 36:163–190. doi: 10.1186/1297-9686-36-2-163

Knight PG, Glister C (2006) TGF-beta superfamily members and ovarian follicle development. Reproduction 132:191–206. doi: 10.1530/rep.1.01074

Kolbehdari D, Wang Z, Grant JR, et al (2009) A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. J Anim Breed Genet 126:216–27. doi: 10.1111/j.1439-0388.2008.00793.x

Kuehn C, Bennewitz J, Reinsch N, et al (2003) Quantitative Trait Loci Mapping of Functional Traits in the German Holstein Cattle Population. J Dairy Sci 86:360–368. doi: 10.3168/jds.S0022-0302(03)73614-5

Laliotis GP, Bizelis I, Rogdakis E (2010) Comparative Approach of the de novo Fatty Acid Synthesis (Lipogenesis) between Ruminant and Non Ruminant Mammalian Species: From Biochemical Level to the Main Regulatory Lipogenic Genes. Curr Genomics 11:168–183. doi: 10.2174/138920210791110960

Lambeth JD (1986) Cytochrome P-450 $_{scc}$ A Review of the Specificity and Properties of the Cholesterol Binding Site. Endocr Res 12:371–392. doi: 10.3109/07435808609035446

Laven RA, Scaramuzzi RJ, Wathes DC, et al (2007) Recent research on the effects of excess dietary nitrogen on the fertility of dairy cows. Vet Rec 160:359–362. doi: 10.1136/vr.160.11.359

Lehner R, Kuksis A (1996) BIOSYNTHESIS OF TRIACYLGLYCEROLS. Prog Lipid Res 35:169–201. doi: O163-7827/96/$32.00

Lemay DG, Neville MC, Rudolph MC, et al (2007) Gene regulatory networks in lactation:

identification of global principles using bioinformatics. BMC Syst Biol 1:56. doi: 10.1186/1752-0509-1-56

Leslie KE (1983) The events of normal and abnormal postpartum reproductive endocrinology and uterine involution in dairy cows: a review. Can Vet J 24:67–71.

Leury BJ, Baumgard LH, Block SS, et al (2003) Effect of insulin and growth hormone on plasma leptin in periparturient dairy cows. Am J Physiol Regul Integr Comp Physiol 285:R1107–R1115. doi: 10.1152/ajpregu.00320.2003

Li X, Li X, Chen H, et al (2013) Non-Esterified Fatty Acids Activate the AMP-Activated Protein Kinase Signaling Pathway to Regulate Lipid Metabolism in Bovine Hepatocytes. Cell Biochem Biophys 67:1157–1169. doi: 10.1007/s12013-013-9629-1

Lin X-Z, Luo J, Zhang L-P, et al (2013) MiR-27a suppresses triglyceride accumulation and affects gene mRNA expression associated with fat metabolism in dairy goat mammary gland epithelial cells. Gene 521:15–23. doi: 10.1016/j.gene.2013.03.050

Ling B, Alcorn J (2008) Acute administration of cefepime lowers L-carnitine concentrations in early lactation stage rat milk. J Nutr 138:1317–22.

Littlejohn MD, Tiplady K, Fink TA, et al (2016) Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. Sci Rep 1–14. doi: 10.5061/dryad.457br

Littlejohn MD, Tiplady K, Lopdell T, et al (2014) Expression variants of the lipogenic AGPAT6 gene affect diverse milk composition phenotypes in Bos taurus. PLoS One 9:e85757. doi:

10.1371/journal.pone.0085757

Llewellyn S, Fitzpatrick R, Kenny D a., et al (2008) Endometrial expression of the insulin-like growth factor system during uterine involution in the postpartum dairy cow. Domest Anim Endocrinol 34:391–402. doi: 10.1016/j.domaniend.2007.11.003

Loor JJ, Everts RE, Bionaz M, et al (2007) Nutrition-induced ketosis alters metabolic and signaling gene networks in liver of periparturient dairy cows. Physiol Genomics 32:105–16. doi: 10.1152/physiolgenomics.00188.2007

Lucy MC (2001) Reproductive loss in high-producing dairy cattle: where will it end? J Dairy Sci 84:1277–93. doi: 10.3168/jds.S0022-0302(01)70158-0

Ma L, Corl B a (2012) Transcriptional regulation of lipid synthesis in bovine mammary epithelial cells by sterol regulatory element binding protein-1. J Dairy Sci 95:3743–55. doi: 10.3168/jds.2011-5083

Mach N, Blum Y, Bannink a, et al (2012) Pleiotropic effects of polymorphism of the gene diacylglycerol-O-transferase 1 (DGAT1) in the mammary gland tissue of dairy cows. J Dairy Sci 95:4989–5000. doi: 10.3168/jds.2012-5348

Mandard S, Müller M, Kersten S (2004) Peroxisome proliferator-activated receptor a target genes. Cell Mol Life Sci 61:393–416. doi: 10.1007/s00018-003-3216-3

Mannaerts GP, Debeer LJ, Thomas J, De Schepper PJ (1979) Mitochondrial and peroxisomal fatty acid oxidation in liver homogenates and isolated hepatocytes from control and clofibrate-treated rats. J Biol Chem 254:4585–95.

Maxa J, Neuditschko M, Russ I, et al (2012) Genome-wide association mapping of milk production traits in Braunvieh cattle. J Dairy Sci 95:5357–64. doi: 10.3168/jds.2011-4673

Mcarthur MJ, Atshaves BP, Frolov A, et al (1999) Cellular uptake and intracellular trafficking of long chain fatty acids. 40:1371–1383.

McGarry JD, Brown NF (1997) The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. Eur J Biochem 244:1–14. doi: 10.1111/j.1432-1033.1997.00001.x

McNamara JP (1988) Regulation of bovine adipose tissue metabolism during lactation. 4. Dose-responsiveness to epinephrine as altered by stage of lactation. J Dairy Sci 71:643–649. doi: 10.3168/jds.S0022-0302(88)79602-2

McNamara JP, Murray CE (2001) Sympathetic nervous system activity in adipose tissues during pregnancy and lactation of the rat. J Dairy Sci 84:1382–1389. doi: 10.3168/jds.S0022-0302(01)70169-5

Meredith BK, Kearney FJ, Finlay EK, et al (2012) Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet 13:21. doi: 10.1186/1471-2156-13-21

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829.

Minozzi G, Nicolazzi EL, Stella A, et al (2013) Genome wide analysis of fertility and production traits in Italian Holstein cattle. PLoS One 8:e80219. doi: 10.1371/journal.pone.0080219

Miyoshi H, Perfield JW, Souza SC, et al (2007) Control of adipose triglyceride lipase action by serine 517 of perilipin A globally regulates protein kinase a-stimulated lipolysis in adipocytes. J Biol Chem 282:996–1002. doi: 10.1074/jbc.M605770200

Mootha VK, Lindgren CM, Eriksson K-F, et al (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34:267–273. doi: 10.1038/ng1180

Mullen MP, Berry DP, Howard DJ, et al (2011) Single nucleotide polymorphisms in the insulin-like growth factor 1 (IGF-1) gene are associated with performance in Holstein-Friesian dairy cattle. Front Genet 2:1–9. doi: 10.3389/fgene.2011.00003

Nayeri S, Sargolzaei M, Abo-Ismail MK, et al (2016) Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet 17:75. doi: 10.1186/s12863-016-0386-1

Nguyen P, Leray V, Diez M, et al (2008) Liver lipid metabolism. J Anim Physiol Anim Nutr (Berl) 92:272–283. doi: 10.1111/j.1439-0396.2007.00752.x

Olsen HG, Hayes BJ, Kent MP, et al (2011) Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. Anim Genet 42:466–74. doi: 10.1111/j.1365-2052.2011.02179.x

Ong EC, Nesin V, Long CL, et al (2013) A TRPC1 protein-dependent pathway regulates osteoclast formation and function. J Biol Chem 288:22219–22232. doi: 10.1074/jbc.M113.459826

Patti ME, Butte AJ, Crunkhorn S, et al (2003) Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. Proc Natl Acad Sci U S A 100:8466–71. doi: 10.1073/pnas.1032913100

Payne AH, Hales DB (2004) Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr Rev 25:947–970. doi: 10.1210/er.2003-0030

Peñagaricano F, Weigel KA, Khatib H (2012) Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. Anim Genet 43:65–71. doi: 10.1111/j.1365-2052.2012.02350.x

Pezzi C, Accorsi PA, Vigo D, et al (2003) 5′-Deiodinase Activity and Circulating Thyronines in Lactating Cows. J Dairy Sci 86:152–158. doi: 10.3168/jds.S0022-0302(03)73595-4

Pimentel ECG, Bauersachs S, Tietze M, et al (2011) Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. Anim Genet 42:251–262. doi: 10.1111/j.1365-2052.2010.02148.x

Pryce JE, Royal MD, Garnsworthy PC, Mao IL (2004) Fertility in the high-producing dairy cow. Livest Prod Sci 86:125–135. doi: 10.1016/S0301-6226(03)00145-3

Puigserver P, Spiegelman BM (2003) Peroxisome Proliferator-Activated Receptor-gamma Coactivator 1 alpha (PGC-1alpha): Transcriptional Coactivator and Metabolic Regulator. Endocr Rev 24:78–90. doi: 10.1210/er.2002-0012

Raven L-A, Cocks BG, Goddard ME, et al (2014a) Genetic variants in mammary development,

prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genet Sel Evol 46:29. doi: 10.1186/1297-9686-46-29

Raven L-A, Cocks BG, Hayes BJ (2014b) Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:62. doi: 10.1186/1471-2164-15-62

Reist M, Erdin D, von Euw D, et al (2002) Estimation of Energy Balance at the Individual and Herd Level Using Blood and Milk Traits in High-Yielding Dairy Cows,. J Dairy Sci 85:3314–3327. doi: 10.3168/jds.S0022-0302(02)74420-2

Reshef L, Olswang Y, Cassuto H, et al (2003) Glyceroneogenesis and the triglyceride/fatty acid cycle. J Biol Chem 278:30413–30416. doi: 10.1074/jbc.R300017200

Reue K, Dwyer JR (2008) Lipin proteins and metabolic homeostasis. J Lipid Res 50:S109–S114. doi: 10.1194/jlr.R800052-JLR200

Reverter A, Fortes MRS (2012) Building single nucleotide polymorphism-derived gene regulatory networks: Towards functional genomewide association studies. J Anim Sci 91:530–536. doi: 10.2527/jas.2012-5780

Rhoads ML, Gilbert RO, Lucy MC, Butler WR (2004) Effects of urea infusion on the uterine luminal environment of dairy cows. J Dairy Sci 87:2896–2901. doi: 10.3168/jds.S0022-0302(04)73420-7

Roche JR, Friggens NC, Kay JK, et al (2009) Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. J Dairy Sci 92:5769–5801. doi:

10.3168/jds.2009-2431

Royal MD, Flint APF, Woolliams JA (2002) Genetic and phenotypic relationships among endocrine and traditional fertility traits and production traits in Holstein-Friesian dairy cows. J Dairy Sci 85:958–67. doi: 10.3168/jds.S0022-0302(02)74155-6

Rudolph MC, McManaman JL, Phang T, et al (2007) Metabolic regulation in the lactating mammary gland: a lipid synthesizing machine. Physiol Genomics 28:323–336. doi: 10.1152/physiolgenomics.00020.2006

Saatchi M, Garrick DJ, Tait RG, et al (2013) Genome-wide association and prediction of direct genomic breeding values for composition of fatty acids in Angus beef cattle. BMC Genomics 14:730. doi: 10.1186/1471-2164-14-730

Sahana G, Guldbrandtsen B, Lund MS (2011) Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. J Dairy Sci 94:479–486. doi: 10.3168/jds.2010-3381

Schillo KK, Halls JB, Hileman SM (1992) effects of nutrition and season in the onset of puberty in the beef heifer. J Anim Sci 70:3994–4005.

Schlegel G, Keller J, Hirche F, et al (2012) Expression of genes involved in hepatic carnitine synthesis and uptake in dairy cows in the transition period and at different stages of lactation. BMC Vet Res 8:28. doi: 10.1186/1746-6148-8-28

Shindou H, Hishikawa D, Harayama T, et al (2008) Recent progress on acyl CoA: lysophospholipid acyltransferase research. J Lipid Res 50:S46–S51. doi: 10.1194/jlr.R800035-JLR200

Sinclair KD, Sinclair LA, Robinson JJ (2000) Nitrogen metabolism and fertility in cattle: I. Adaptive changes in intake and metabolism to diets differing in their rate of energy and nitrogen release in the rumen. J Anim Sci 78:2659–2669.

Sivan E, Boden G (2003) Free fatty acids, insulin resistance, and pregnancy. Curr Diab Rep 3:319–322. doi: 10.1007/s11892-003-0024-y

Snelling WM, Cushman R a., Keele JW, et al (2013) Networks and pathways to guide genomic selection. J Anim Sci 91:537–552. doi: 10.2527/jas2012-5784

Solovieff N, Cotsapas C, Lee PH, et al (2013) Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 14:483–495. doi: 10.1126/scisignal.2001449.Engineering

Spicer LJ, Echternkamp SE (1995) The ovarian insulin and insulin-like growth factor system with an emphasis on domestic animals. Domest Anim Endocrinol 12:223–245. doi: 10.1016/0739-7240(95)00021-6

Stagg K, Spicer LJ, Sreenan JM, et al (1998) Effect of calf isolation on follicular wave dynamics, gonadotropin and metabolic hormone changes, and interval to first ovulation in beef cows fed either of two energy levels postpartum. Biol Reprod 59:777–783. doi: 10.1095/biolreprod59.4.777

Stevenson KR, Wathes DC (1996) Insulin-like growth factors and their binding proteins in the ovine oviduct during the oestrous cycle. J Reprod Fertil 108:31–40. doi: 10.1530/jrf.0.1080031

Sumner-Thomson JM, Vierck JL, McNamara JP (2011) Differential expression of genes in

adipose tissue of first-lactation dairy cattle. J Dairy Sci 94:361–369. doi: 10.3168/jds.2010-3447

Sumner JM, McNamara JP (2007) Expression of lipolytic genes in the adipose tissue of pregnant and lactating Holstein dairy cattle. J Dairy Sci 90:5237–46. doi: 10.3168/jds.2007-0307

Sutherland KD, Lindeman GJ, Visvader JE (2007) The molecular culprits underlying precocious mammary gland involution. J Mammary Gland Biol Neoplasia 12:15–23. doi: 10.1007/s10911-007-9034-8

Takeuchi K, Reue K (2009) Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis. Am J Physiol Endocrinol Metab 296:E1195–E1209. doi: 10.1152/ajpendo.90958.2008

Taylor VJ, Cheng Z, Pushpakumara PGA, et al (2004) Relationships between the plasma concentrations of insulin-like growth factor-I in dairy cows and their fertility and milk yield. Vet Rec 155:583–588. doi: 10.1136/vr.155.19.583

Thomasen JR, Guldbrandtsen B, Sørensen P, et al (2008) Quantitative trait loci affecting calving traits in Danish Holstein cattle. J Dairy Sci 91:2098–2105. doi: 10.3168/jds.2007-0602

Van den Top AM, Wensing T, Geelen MJ, et al (1995) Time trends of plasma lipids and enzymes synthesizing hepatic triacylglycerol during postpartum development of fatty liver in dairy cows. J Dairy Sci 78:2208–2220. doi: 10.3168/jds.S0022-0302(95)76848-5

Van Dorland HA, Richter S, Morel I, et al (2009) Variation in hepatic regulation of metabolism during the dry period and in early lactation in dairy cows. J Dairy Sci 92:1924–1940. doi:

10.3168/jds.2008-1454

Veerkamp RF, Beerda B (2007) Genetics and genomics to improve fertility in high producing dairy cows. Theriogenology 68:266–273. doi: 10.1016/j.theriogenology.2007.04.034

Vernon RG, Denis RGP, Sørensen a. (2001) Signals of adiposity. Domest Anim Endocrinol 21:197–214. doi: 10.1016/S0739-7240(01)00121-7

Viitala S, Szyda J, Blott S, et al (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. doi: 10.1534/genetics.105.046730

Viollet B, Foretz M, Guigas B, et al (2006) Activation of AMP-activated protein kinase in the liver: a new strategy for the management of metabolic hepatic disorders. J Physiol 574:41–53. doi: 10.1113/jphysiol.2006.108506

Viturro E, Koenning M, Kroemer A, et al (2009) Cholesterol synthesis in the lactating cow: Induced expression of candidate genes. J Steroid Biochem Mol Biol 115:62–7. doi: 10.1016/j.jsbmb.2009.02.011

Walsh SW, Williams EJ, Evans ACO (2011) A review of the causes of poor fertility in high milk producing dairy cows. Anim Reprod Sci 123:127–138. doi: 10.1016/j.anireprosci.2010.12.001

Wang X, Wurmser C, Pausch H, et al (2012) Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. PLoS One 7:e40711. doi: 10.1371/journal.pone.0040711

Wathes DC, Cheng Z, Chowdhury W, et al (2009) Negative energy balance alters global gene expression and immune responses in the uterus of postpartum dairy cows. 1–13. doi: 10.1152/physiolgenomics.00064.2009.

Wathes DC, Cheng Z, Fenwick MA, et al (2011) Influence of energy balance on the somatotrophic axis and matrix metalloproteinase expression in the endometrium of the postpartum dairy cow. Reproduction 141:269–281. doi: 10.1530/REP-10-0177

Wathes DC, Fenwick M, Cheng Z, et al (2007) Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. Theriogenology 68:232–241. doi: 10.1016/j.theriogenology.2007.04.006

Wathes DC, Taylor VJ, Cheng Z, Mann GE (2003) Follicle growth, corpus luteum function and their effects on embryo development in postpartum dairy cows. Reprod Suppl 61:219–237.

Weikard R, Goldammer T, Brunner RM, Kuehn C (2012) Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows. Physiol Genomics 44:728–739. doi: 10.1152/physiolgenomics.00007.2012

Weng L, Macciardi F, Subramanian A, et al (2011) SNP-based pathway enrichment analysis for genome-wide association studies. BMC Bioinformatics 12:99. doi: 10.1186/1471-2105-12-99

Xie Y, Yang S, Cui X, et al (2014) Identification and expression pattern of two novel alternative splicing variants of EEF1D gene of dairy cattle. Gene 534:189–96. doi: 10.1016/j.gene.2013.10.061

Yamashita A, Nakanishi H, Suzuki H, et al (2007) Topology of acyltransferase motifs and substrate specificity and accessibility in 1-acyl-sn-glycero-3-phosphate acyltransferase 1. Biochim Biophys Acta - Mol Cell Biol Lipids 1771:1202–1215. doi: 10.1016/j.bbalip.2007.07.002

Zhang Q, Koser SL, Bequette BJ, Donkin SS (2015) Effect of propionate on mRNA expression of key genes for gluconeogenesis in liver of dairy cattle. J Dairy Sci 98:8698–709. doi: 10.3168/jds.2015-9590

Zhou G, Myers R, Li Y, et al (2001) Role of AMP-activated protein kinase in mechanism of metformin action. J Clin Invest 108:1167–1174. doi: 10.1172/JCI200113505

Zhu LH, Armentano LE, Bremmer DR, et al (2000) Plasma Concentration of Urea, Ammonia, Glutamine Around Calving, and the Relation of Hepatic Triglyceride, to Plasma Ammonia Removal and Blood Acid-Base Balance. J Dairy Sci 83:734–740. doi: 10.3168/jds.S0022-0302(00)74935-6

Table 2. 1: Major enzymes and their physiological function early in lactation in dairy cattle.

Enzyme name	Tissue	Function	Reference
Acetyl-CoA carboxylase (ACC)	Adipocytes	Lipogenesis (carboxylation of acetyl-CoA to malonyl-CoA	Bauman (1980)
Acyl-malonyl ACP condensing	Adipocytes	Lipogenesis (condensation of acetyle-CoA to form malonyl-CoA)	Bauman (1980); Fielding and Frayn (1998)
Lipoprotein lipase (LPL)	Adipocytes	Lipogenesis (Hydrolization of plasma TAG to form NEFA and monoacylglycerides)	Bauman (1980); Nguyen et al. (2008)
Thioesterase	Adipocytes	Lipogenesis (Hydrolization of C_{16}-acyl ACP to palmitate)	Lehner and Kuksis (1996)
Hormone sensitive-lipase (HSL)	Adipocytes	Lipolysis (hydrolysis of fatty acids at sn-1 and sn-3 position)	Lehner and Kuksis (1996)
Monoacylglycerol lipase	Adipocytes	Lipolysis (hydrolysis of the remaining fatty acid at the sn-2 position to generates NEFA)	Berg et al. (2002)
Acyl-CoA synthase	Hepatocyte (cytoplasm)	Conversion of NEFA to fatty acyl-CoA	Berg et al. (2002); McGarry and Brown (1997)
Carnitine palmitoyltransferase I (CPT-I)	Hepatocytes (outer membrane mitochondria)	Fatty acid β-oxidation (uptake and formation of fatty acids to fatty acyl-CoA)	Berg et al. (2002)
Carnitine-acylcarnitine translocase	Hepatocytes (inner membrane of mitochondria)	Fatty acid β-oxidation (translocation of fatty acyl-CoA into the mitochondria)	Berg et al. (2002)
Carnitine-palmitoyltransferase II (CPT-II)	Hepatocytes (inner membrane of mitochondria)	Fatty acid β-oxidation (reforming acyl-CoA in mitochondria matrix)	Berg et al. (2002); McGarry and Brown (1997)
Acyl-CoA dehydrogenase	Hepatocytes (mitochondria matrix)	Fatty acid β-oxidation (Dehydrogenation of acyl-CoA by FAD)	Mannaerts et al. (1979)
Enoyl-CoA hydratase	Hepatocytes (mitochondria matrix)	Fatty acid β-oxidation (Hydration of enoyl-CoA to hydroxyacyl-CoA)	Mannaerts et al. (1979)

3-hydroxyacyl-CoA dehydrogenase	Hepatocytes (mitochondria matrix)	Fatty acid β-oxidation (oxidation of β-hydroxyacyl-CoA to β-ketoacyl-CoA by NAD^+)	Mannaerts et al. (1979)
ß-ketothiolase	Hepatocytes (mitochondria matrix)	Fatty acid β-oxidation (thiolysis of β-ketoacyl-CoA)	Mannaerts et al. (1979)
3-hydroxy-3-methylglutaryle-CoA synthase	Liver	Cholesterol synthesis (formation of 3-hydroxy-3-methylglutaryl-CoA from acetyl-CoA and acetoacetyl-CoA)	Horton et al. (1998); Viturro et al. (2009)
3-hydroxy-3-methylglytaryl CoA reductase	Liver	Cholesterol synthesis (reduction of 3-hydroxy-3-metyhylglutaryl-CoA to mevalonate)	Berg et al. (2002); Viturro et al. (2009)
Squalene synthesis	Liver (endoplasmic reticulum)	Cholesterol synthesis (reduction of two farnesyl pyrophosphate to form squalene)	Berg et al. (2002)
Oxidosqualene cyclase	Liver	Cholesterol synthesis (cyclizes of squalene to lanosterol)	Berg et al. (2002)
Glycerol-3-phosphate acyltransferase (GPAT)	Mammary gland (endoplasmic reticulum and/or mitochondria)	TAG synthesis (acylation of glycerol3-phosphate to form lysophosphatidic acid (LPA))	Gimeno and Cao (2008); Takeuchi and Reue (2009)
1-acylglycerol-3-phoshplte acyltransferase (AGPAT; also known as LPA acyltransferase)	Mammary gland (endoplasmic reticulum and/or mitochondria)	TAG synthesis (transfer of an additional fatty acid to LPA to form phosphatidate (PA))	Takeuchi and Reue (2009)
Lipin	Mammary gland (endoplasmic reticulum)	TAG synthesis (conversion of the phosphatidate to diacylglycerol)	Takeuchi and Reue (2009); Reue and Dwyer (2008)
Diacylglycerol acyltransferase (DGAT)	Mammary gland (endoplasmic reticulum)	TAG synthesis (acylation of DAG to TAG)	Takeuchi and Reue (2009); Shindou et al. (2008); Bionaz and Loor

			(2008a)
P450 side-chain cleavage enzyme (P450scc, CYP11A1)	Inner mitochondrial membrane of steroidogenic cells (ovary)	Steroid hormone synthesis (Pregnenolone synthesis)	Hu et al. (2010); Lambeth (1986)
Δ^5-3ß hydroxysteroid dehydrogenase isomerase (3ßHSD)	Steroidogenic cells (granulosa cells) in ovary	Steroid hormone synthesis (progesterone synthesis)	Hu et al. (2010); Payne and Hales (2004)
17α-Hydroxypregnenolone	Steroidogenic cells (theca cells) ovary	Synthesis of estrogen and androstenedione from progesterone	Payne and Hales (2004)
17β-Hydroxysteroid dehydrogenases	Ovary	Synthesis of testosterone from androstenedione	Payne and Hales (2004)
CYP19A1 (aromatase)	Theca-interstitial cells of ovary	Testosterone	Hu et al. (2010); Lambeth (1986)

Table 2. 2: List of major candidate genes identified through association studies for production traits in dairy cattle.

Gene	Gene name	Chromosome	PUBMED-ID	Trait	Year
POU1F1	POU class 1 homeobox 1	1	18557974	Milk yield, productive life	2008
DIP2A	DIP2 disco-interacting protein 2 homolog A	1	21048968	Protein yield	2010
TNFSF10	Tumor necrosis factor	1	21198698	Fat yield, protein yield, fat percentage, interval from first to successful insemination (cow)	2011
MIS18A	MIS18 kinetochore protein homolog A (S. pombe)	1	24456127	Somatic cell score	2014
SLC37A1	solute carrier family 37 member 1	1	26613780	Milk production	2016
STAT1	Signal transducer and activator of transcription 1	2	17033032	Milk yield, fat yield, protein yield	2006
CYP27A1	Cytochrome P450, family 27, subfamily A, polypeptide 1	2	21198698	Milk yield, somatic cell score	2011
IFIH1	Interferon induced with helicase C domain 1	2	21198698	Milk yield, fat yield, fat percentage	2011
IGFBP2	Insulin-like growth factor binding protein 2, 36kDa	2	21198698	Lactation, establishment of pregnancy	2011
SLC40A1	Solute carrier family 40	2	25148050	Milk yield	2014
SP110	SP110 nuclear body protein	2	24456127	Fat percentage	2014
SDC3	syndecan 3	2	24456127	Mammary system	2014
SMARCAL1	SWI/SNF related, matrix assoc., actin dep. Reg. of chromatin, subfamily a-like 1	2	24456127	Mammary system	2014
GBA	Glucosidase beta, acid	3	24456127	Protein percentage	2014
CTTNBP2NL	CTTNBP2 N-terminal like	3	24456127	Somatic cell score	2014
MUC1	mucin 1, cell surface	3	26613780	Milk production	2016

	associated				
LEP	Leptin	4	15905454, 18565947, 18650297, 15927775	Milk protein, milk fat, lactation performance, health, daily milk reproduction, postpartum luteal activity	2005, 2008
OLR1	Oxidized low density lipoprotein	5	16606746	Milk fat yield, milk fat percentage	2006
GABARAPL1	GABA type A receptor associated protein like 1	5	21198698, 27287773	Milk yield, fat percentage, fat production	2011, 2016
MGP	Matrix Gla protein	5	21198698	Milk yield, fat percentage	2011
EPS8	Epidermal growth factor receptor pathway substrate 8	5	24456127	Milk yield	2014
MGST1	Microsomal glutathione S-transferase 1	5	24456127	Fat yield, fat percentage	2014
RPAP3	RNA polymerase II associated protein 3	5	24456127	Milk yield, protein percentage	2014
SOCS2	Suppressor of cytokine signaling 2	5	24779965	Mammary development pathways, prolactin signaling pathways, lactation	2014
ATF4	Activating transcription factor 4	5	24779965	Lactation yields, involution pathways	2014
CCDC91	Coiled-coil domain containing 91	5	25148050	Fat percentage	2014
ITPR2	inositol 1,4, 5-triphosphate receptor, type 2	5	25148050	Fat percentage	2014
ACSS3	Acyl-CoA synthetase short-chain family member 3	5	25511820	Milk fat composition, milk fat percentage	2014
MKL1	megakaryoblastic leukemia	5	27006194	Milk yield	2016

	(translocation) 1				
VDR	vitamin D (1,25-dihydroxyvitamin D3) receptor	5	26613780	Milk production	2016
CSF2RB	colony stimulating factor 2 receptor beta common subunit	5	26613780	Milk production	2016
NCF4	neutrophil cytosolic factor 4	5	26613780	Milk production	2016
CSN1S2	Casein alpha-S2	6	15040897	Protein yield, protein percentage, fat yield, fat percentage, milk yield)	2004
CSN2	Casein beta	6	15040897	Protein yield, protein percentage, fat yield, fat percentage, milk yield)	2004
PPARGC1A	Peroxisome proliferator-activated receptor gamma, coactivator 1 alpha	6	15781588	Milk fat	2005
SPP1	Secreted phosphoprotein	6	16230712	Milk production, milk protein percentage, milk fat percentage	2005
IL8	Interleukin 8	6	17433017	Milk yield, fat yield, protein yield, somatic cell score	2007
IGFBP7	Insulin-like growth factor binding protein 7	6	21198698	Milk yield, 56 day non-return rate, interval from first service to successful insemination (heifer)	2011
FAM13A1	Family with sequence similarly 13, member A	6	21257065	Milk yield, fat yield, fat percentage	2011
IGFBP-5	Insulin-like growth factor binding protein-5	6	21338820	Calving ability, milk yield, protein yield, mammary gland	2011

				involution	
PKD2	Polycystic kidney disease 2	6	25148050	Protein percentage	2014
CSN1S1	Casein alpha s1	6	12939094, 15905454, 16840633	Milk yield, fat yield, protein yield, milk fat percentage, milk protein percentage	2003, 2005, 2006
CSN3	Casein kappa	6	12939094, 18666558	Milk yield, fat yield, protein yield, milk fat percentage, milk protein percentage	2003, 2008
ABCG2	ATP-binding cassette, sub-family G	6	15998908, 17584938, 17106124	Milk yield, milk fat and protein concentration	2005, 2007
PPARGC1A	Proliferative peroxisome-activated receptor, coactivator 1	6	22669841	Milk performance	2012
CAS1A	Casein alpha s1	6	24456127	Protein percentage	2014
LARP1	La ribonucleoprotein domain family, member 1	7	24456127	Somatic cell score	2014
IRF1	Interferon regulatory factor 1	7	24779965	Lactation yields, involution pathways	2014
GRIA1	Glutamate receptor, ionotropic, AMPA1	7	25511820	Milk fat composition, milk fat percentage	2014
CAST	Calpastatin	7	16734705, 23759029	Daughter pregnancy rate, productive life, protein yield, milk yield, fat yield, somatic cell score, net merit, conception rate (heifer and cow)	2006, 2013
FBP1	Fructose 1,6 bisphosphatase 1	8	22669841	Milk performance	2012
FBP2	Fructose 1,6 bisphosphatase 2	8	22669841	Milk performance	2012
TP53	Tumor protein p53	9	17584498	Lactation and	2007

				involution, pregnancy, puberty	
TEP1	telomerase associated protein 1	10	27506634	Test day protein yield, test day fat yield	2016
PCK2	Phosphoenolpyruvate carboxykinase 2, mitochondrial isoform	10	22669841	Milk performance	2012
GFI1B	Growth factor independent 1B transcription receptor	11	21048968	Fat percentage	2010
LGB	Lactoglobulin, beta	11	22192223, 19032698, 12836958	Milk protein composition, milk beta-lactoglobulin protein concentration	2012, 2009, 2003
NLRP6	NLR family, pyrin domain containing 6	11	24456127	Mammary system	2014
PRKCE	Protein kinase C, epsilon	11	24456127	Mammary system	2014
NRXN1	neurexin 1	11	24456127	Somatic cell score	2014
PAEP	progestagen-associated endometrial protein	11	26613780	Milk production	2016
RNF219	Ring finger protein 219	12	24456127	Fat production	2014
ACSS2	Acyl-CoA synthetase short-chain family member 2	13	21569316	Fat yield, milk fatty acids	2011
PLK1S1	kizuna centrosomal protein	13	27506634	Somatic cell score	2016
PCK1	Phosphoenolpyruvate carboxykinase 1, cytosolic isoform	13	22669841	Milk performance	2012
CYP11B1	Cytochrome P450, subfamily XI B, polypeptide 1	14	17179535	Milk production, somatic cell score, maternal calving ease, 90-day non-return rate (maternal and paternal)	2007
VPS28	Vacuolar protein sorting 28 homolog	14	21048968	Milk yield, protein percentage, fat yield, fat percentage	2010

MAF1	MAF1 homolog	14	21048968	Milk yield, fat percentage	2010
OPLAH	5-oxoprolinase	14	21048968	Fat percentage	2010
MAPK15	Mitogen-activated protein kinase 15	14	21048968	Fat percentage	2010
ZNF623	Zinc finger protein 623	14	21048968	Fat percentage, milk yield	2010
EEF1D	Eukaryotic translation elongation factor 1 delta	14	21048968	Fat percentage	2010
ZC3H3	Zinc finger CCCH-type containing 3	14	21048968	Fat percentage	2010
GML	Glycosylphosphatidylinositol anchored molecule like	14	21048968	Fat percentage, milk yield	2010
GPIHBP1	Glycosylphosphatidylinositol anchored high density lipoprotein binding protein 1	14	21048968	Milk yield, protein yield, fat percentage	2010
RHPN1	Rhophilin, Rho GTPase binding protein 1	14	21048968	Fat percentage	2010
PTK2	Protein tyrosine kinase 2	14	21048968	Fat percentage	2010
KCNK9	Potassium channel, subfamily K, member 9	14	21048968	Fat percentage	2010
COL22A1	Collagen, type XXII, alpha 1	14	21048968	Milk yield, protein yield, fat percentage	2010
KHDRBS3	KH domain containing, RNA binding, signal transduction associated 3	14	21048968	Fat percentage	2010
NIBP	IKKβ binding protein	14	21831322	Fat percentage	2011
CEBPD	CCAAT/enhancer binding protein	14	24779965	Involution pathways	2014
MYC	v-myc avian myelocytomatosis viral oncogene homolog	14	24779965	Involution pathways	2014
CYHR1	Cysteine/histidine-rich 1	14	25511820, 27287773	Milk fat composition, milk fat percentage, milk production	2014, 2016

ARHGAP39	Rho GTPase activating protein 39	14	25511820, 27287773	Milk fat composition, milk fat percentage, fat production	2014, 2016
CPSF1	Cleavage and polyadenylation specific factor 1	14	25511820	Milk fat composition, milk fat percentage	2014
GRINA	Glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1	14	25511820	Milk fat composition, milk fat percentage	2014
FAM83H	Family with sequence similarity 83, member H	14	25511820	Milk fat composition, milk fat percentage	2014
DGAT1	Diacylglycerol O-acyltransferase 1	14	11827942, 16621755, 17179535, 18650297, 18666558, 18669245	Milk yield and composition, milk protein yield, milk fat yield, milk fat and protein percentage, somatic cell score, maternal non-return rate, productive life, conformation traits	2002, 2006, 2007, 2008
CYP11B1	Cytochrome P450, subfamily XI B, polypeptide 1	14	21048968, 17179535	Milk yield, fat yield, protein yield, milk fat percentage, milk protein percentage, somatic cell score, maternal calving ease, 90-day non-return rate (paternal and maternal)	2010, 2007
ADCK5	AarF domain containing kinase 5	14	27506634	Test day fat yield (milk fat%)	2016
TONSL	Tonsoku-like, DNA repair	14	27506634,	Test day fat yield	2016

	protein		27287773	(milk fat%), milk production	
PPP1R16A	Protein phosphatase 1 regulatory subunit 16A	14	27506634, 27287773	Test day fat yield (milk fat%), milk production	2016
TRAPPC9	Trafficking protein particle complex 9	14	27506634	Test day fat yield (milk fat%)	2016
LRRC14	Leucine rich repeat containing 14	14	27506634	305 day fat yield, lactose percentage	2016
FOXH1	forkhead box H1	14	27287773	milk production	2016
PPP1R16A	protein phosphatase 1 regulatory subunit 16A	14	27287773	fat production, and fat percentage	2016
SMPD5	Sphingomyelin phosphodiesterase 5	14	27287773	Fat production, and fat percentage	2016
MROH1	Maestro heat like repeat family member 1	14	27287773	Fat production, and fat percentage	2016
EIF2C2	Argonaute 2, RISC catalytic component	14	25510969	Milk yield, fat yield, protein yield	2014
TRAPPC9	Trafficking protein particle complex 9	14	25510969	Milk yield, fat yield, protein yield	2014
HEATR7A	Maestro heat like repeat family member 1	14	25510969	Fat percentage	2014

Table 2. 3: List of major candidate genes identified through association studies for fertility traits in dairy cattle.

Gene	Gene name	Chromosome	PUBMED-ID	Trait	Year
PCCB	Propionyl CoA carboxylase, beta polypeptide	1	23759029	Daughter pregnancy rate	2013
TRPC1	Transient receptor potential cation channel subfamily C member 1	1	26369327	Female fertility index, 56-day non-return rate, number of inseminations per conception	2015
IGFBP2	Insulin-like growth factor binding protein 2, 36kDa	2	21198698	Lactation, establishment of pregnancy	2011
TSHB	Thyroid stimulating hormone, beta	3	23759029	Daughter pregnancy rate	2013
WDR77	*Bos taurus* WD repeat domain 77	3	25216717	Heifer non-return rate	2015
VAV3	Vav guanine nucleotide exchange factor 3	3	25216717	Days from first to last insemination	2015
CD36	Platelet glycoprotein 4	4	25216717	Number of inseminations, 56-day non-return rate, days from first to last insemination, the interval from calving to first insemination	2014
LEP	Leptin	4	15905454, 18565947, 18650297, 15927775	Milk protein, milk fat, lactation performance, health, daily milk reproduction, postpartum luteal activity	2005, 2008
ADCY1	Adenylate cyclase 1	4	24428918, 18565942	Number of inseminations, 56-day	2014, 2008

				non-return rate, days from first to last insemination, the interval from calving to first insemination, mammary system, conformation traits, daughter fertility, calving ease	
SEMA3C	Sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	4	25216717	Number of insemination per conception; days from first to last insemination; 56-day non-return rate; the length in days of the interval from calving to first insemination	2014
GNAT3	G protein subunit alpha transducin 3	4	25216717	Number of insemination per conception; days from first to last insemination; 56-day non-return rate; the length in days of the interval from calving to first insemination	2014
CSNK1E	Casein kinase 1, epsilon	5	23759029	Daughter pregnancy rate, heifer conception rate, productive life	2013
IGF1	Insulin-like growth factor 1	5	24265800	Days to first service	2013
AMHR2	Anti-mullerian hormone receptor type II	5	24265800	Calving interval	2013
CPT1B	Carnitine palmitoyltransferase 1B	5	24265800	56-day non-return rate	2013

ATP2B1	ATPase, Ca++ transporting plasma membrane 1	5	24265800, 19448026, 12926772	Calving interval, 56-day non-return rate, days to first service, 305-day first parity lactation, fat yield, protein yield	2013, 2009, 2003
SOX5	SRY (sex determining region Y)-box 5	5	24456127	Fertility	2014
IGFBP7	Insulin-like growth factor binding protein 7	6	21198698	Milk yield, 56 day non-return rate, interval from first service to successful insemination (heifer)	2011
IGFBP-5	Insulin-like growth factor binding protein-5	6	21338820	Calving ability, milk yield, protein yield, mammary gland involution	2011
CLOCK	Clock circulation regulator	6	23759029	Daughter pregnancy rate	2013
GPR125	Adhesion G protein-coupled receptor A3	6	26369327	Female fertility index	2015
NPFFR2	Neuropeptide FF receptor 2	6	24456127	Fertility	2014
EPGN	Epithelial mitogen	6	26613780	Calving interval	2016
CSF2	Colony stimulating factor 2	7	23759029	daughter pregnancy rate	2013
CAST	Calpastatin	7	16734705, 23759029	daughter pregnancy rate, productive life, protein yield, milk yield, fat yield, somatic cell score, net merit, conception rate (heifer and cow)	2006, 2013
TP53	Tumor protein p53	9	17584498	Lactation and involution,	2007

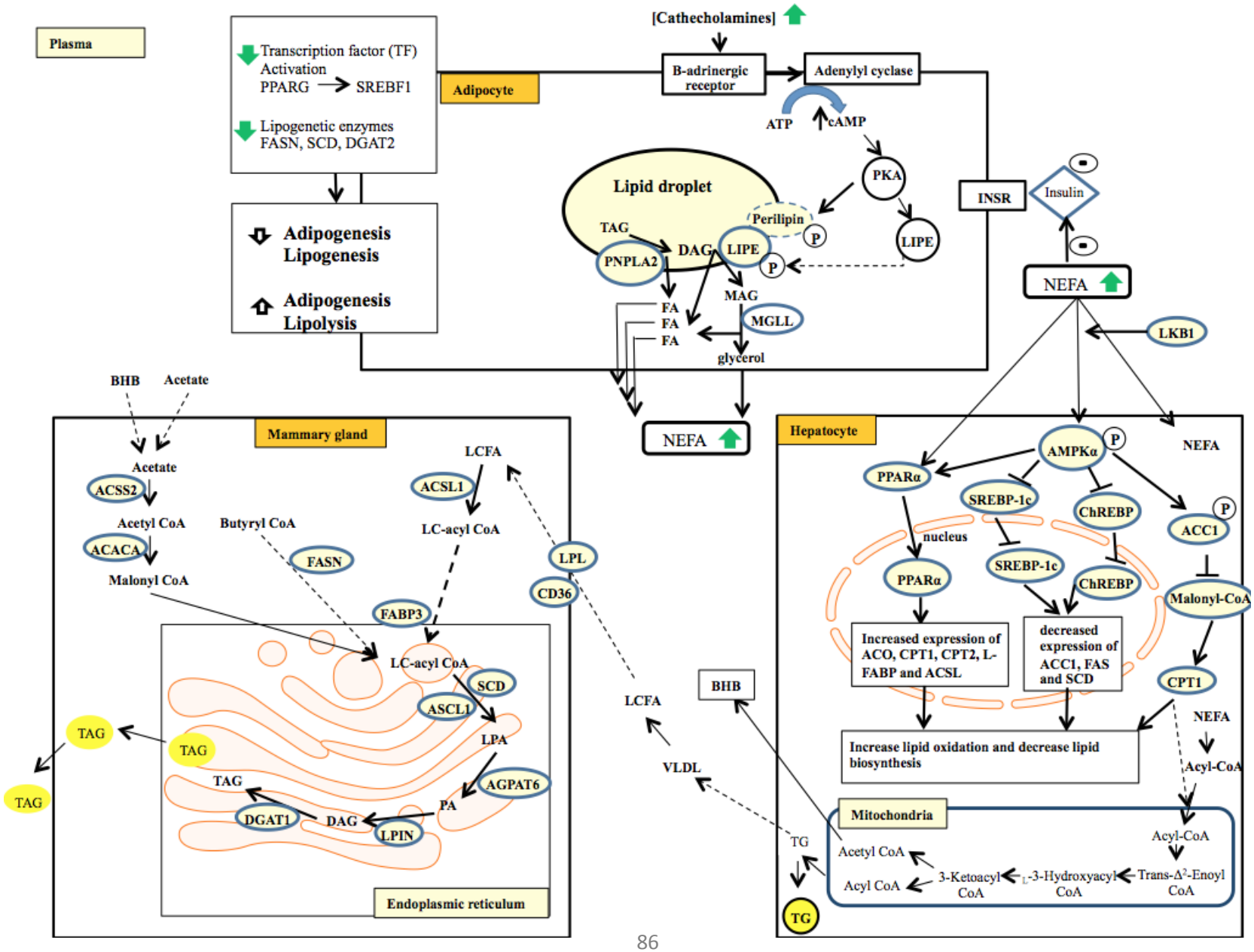
				pregnancy, puberty	
ACAT2	Acetyl-CoA acetyltransferase 2	9	23759029	Daughter pregnancy rate, conception rate (cow and heifer), productive life	2013
WDR27	WD repeat domain 27	9	24456127	Survival	2014
SLC1A4	Solute carrier family 1	11	24428918	56-day non-return rate (cow), days interval from calving to first service	2014
PPM1B	Protein phosphatase Mg2+/Mn2+ dependent,1B	11	24428918	56-day non-return rate (cow), days interval from calving to first service	2014
FSHR	Follicle stimulating hormone receptor	11	23759029, 20207511	Conception rate (heifer), productive life, superovulation response	2013, 2010
NLRP6	NLR family, pyrin domain containing 6	11	24456127	Survival	2014
HNF4A	Hepatocyte nuclear factor 4, alpha	13	23759029	Daughter pregnancy rate	2013
CACNB2	Calcium channel, voltage-dependent, beta 2 subunit	13	25216717	Number of inseminations, 56-day non-return rate, days from first to last insemination, the interval from calving to first insemination	2014
ZEB1	Zinc finger E-box binding homeobox 1	13	25216717	Number of inseminations, 56-day non-return rate, days from first to last	2014

				insemination, the interval from calving to first insemination	
ARHGAP12	Rho GTPase activating protein 12	13	25216717	Number of inseminations, 56-day non-return rate, days from first to last insemination, the interval from calving to first insemination	2014
ANKRD60	Ankyrin repeat domain 60	13	26369327	Female fertility index	2015
ANKRD60	Ankyrin repeat domain 60	13	25216717	Female fertility index; days from first to last insemination	2015
CYP11B1	Cytochrome P450, subfamily XI B, polypeptide 1	14	17179535	Milk production, somatic cell score, maternal calving ease, 90-day non-return rate (maternal and paternal)	2007
PLAG1	PLAG1 zinc finger	14	22100599	Calving ease	2012
MOS	V-mos Moloney murine sarcoma viral oncogene homolog	14	22100599	Reproduction rate	2012
TOX	Thymocyte selection-associated high mobility group box	14	22100599	Age at puberty	2012
CSPP1	Centrosome and spindle pole associated protein 1	14	23759029	Daughter pregnancy rate	2013
CPSF1	Cleavage and polyadenylation specific factor 1, 160kDa	14	23759029	Daughter pregnancy rate	2013
DGAT1	Diacylglycerol O-acyltransferase 1	14	11827942, 16621755, 17179535, 18650297, 18666558, 18669245	Milk yield and composition, milk protein yield, milk fat	2002, 2006, 2007, 2008

				yield, milk fat and protein percentage, somatic cell score, maternal non-return rate, productive life, conformation traits	
CYP11B1	Cytochrome P450, subfamily XI B, polypeptide 1	14	21048968, 17179535	Milk yield, fat yield, protein yield, milk fat percentage, milk protein percentage, somatic cell score, maternal calving ease, 90-day non-return rate (paternal and maternal)	2010, 2007
CD82	CD82 molecule	15	21831322	Daughter stillbirth	2011
PGR	Progesterone receptor-like	15	23759029, 23076525	In vitro fertilization or development, daughter pregnancy rate	2013
HSD17B12	Hydroxysteroid 17-beta dehydrogenase 12	15	23759029	Daughter pregnancy rate	2013
NEU3	Neuraminidase 3	15	23759029	Conception rate (heifer and cow), productive life	2013
GRAMD1B	GRAM domain containing 1B	15	26369327	Female fertility index	2015
PAPPA2	Pappalysin 2	16	22100599	Pregnancy rate, daughter calving ease	2012
MTOR	Mechanistic target of rapamycin	16	22100599	Reproduction rate (regulation of GnRH release before the initiation of ovarian cycling)	2012

DYRK3	Dual-specificity tyrosine-	16	23759029	Daughter pregnancy rate	2013
TGFB2	Transforming growth factor, beta 2	16	24265800	Number of days open	2013
IGLL1	Immunoglobulin lambda-like polypeptide 1	17	24265800	Fertility index	2013
PGLYRP1	Peptidoglycan recognition protein 1	18	21831322	Fat yield, protein yield, service-sire, daughter calving-ease, net merit, milk yield, productive life, fat percentage, protein percentage	2011
IGFL1	Insulin growth factor-like family member 1	18	21831322	Fat yield, protein yield, service-sire, daughter calving-ease, net merit, milk yield, productive life, fat percentage, protein percentage	2011

Figure 2. 1: Proposed steps, genes and pathways controlling energy metabolism in different tissues early in lactation. Solid/thick lines denote processes that are activated. Solid non-arrow lines denote processes that are inhibited. Adapted from Khan et al. (2013); Li et al. (2013); Bionaz and Loor (2008) and Roche et al. (2006).



CHAPTER 3. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle[2]

3.1. Abstract

Genome-wide association studies (GWAS) are a powerful tool for detecting genomic regions explaining variation in phenotype. The objectives of the present study were to identify or refine the positions of genomic regions affecting milk production, milk components and fertility traits in Canadian Holstein cattle, and to use these positions to identify genes and pathways that may influence these traits. Several QTL regions were detected for milk production (MILK), fat production (FAT), protein production (PROT) and fat and protein deviation (FATD, PROTD respectively). The identified QTL regions for production traits (including milk production) support previous findings and some overlap with genes with known relevant biological functions identified in earlier studies such as *DGAT1* and *CPSF1*. A significant region on chromosome 21 overlapping with the gene *FAM181A* and not previous linked to fertility in dairy cattle was identified for the calving to first service interval and days open. A functional enrichment analysis of the GWAS results yielded GO terms consistent with the specific phenotypes tested, for example GO terms GO:0007595 (lactation) and GO:0043627 (response to estrogen) for milk production (MILK), GO:0051057 (positive regulation of small GTPase mediated signal transduction) for fat production (FAT), GO:0040019 (positive regulation of embryonic development) for first service to calving interval (CTFS) and GO:0043268 (positive regulation of potassium ion transport) for days open (DO). In other cases, the connection between the enriched

[2] A version of this chapter has been published in the *Journal BMC Genetics*. Nayeri et al., 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. 10;17(1):75. DOI: 10.1186/s12863-016-0386-1.

GO terms and the traits were less clear, for example GO:0003279 (cardiac septum development) for FAT and GO:0030903 (notochord development) for DO trait. The chromosomal regions and enriched pathways identified in this study confirm several previous findings and highlight new regions and pathways that may contribute to variation in production or fertility traits in dairy cattle.

3.2. Introduction

Milk production and fertility are two economically important traits affecting profitability in dairy cattle. These traits are polygenic, affected by many genes and variants, each with small effects on the observed phenotype (Snelling et al., 2013). Improvements in management and nutrition, along with intense genetic selection have increased milk production in recent decades. However, selection has also changed the reproductive physiology of the cow and led to a decrease in reproductive efficiency (Lucy, 2001). For example, time in estrus has been reduced to less than 8 hours in lactating dairy cows (Nebel et al., 2000), pregnancy rate has decreased, days open and services per conception has increased (Lucy, 2001). In the last decade, advances in genome sequencing technologies and availability of a tremendous number of genetic variants in the form of single nucleotide polymorphisms (SNP) (Bovine HapMap Consortium, 2009) have led to the application of genomic selection (GS) (Meuwissen et al., 2001). Genomic selection is based on linkage disequilibrium (LD) of unknown functional mutations and SNP genotypes that are spread out across the whole genome (Meuwissen et al., 2001). Incorporation of functional mutations into genotyping panels could increase GS accuracy and applicability across populations (Meuwissen and Goddard, 2010; Snelling et al., 2013).

Many quantitative regions and candidate genes associated with milk production and fertility have been identified by means of genome-wide association analysis studies (GWAS) (Pimentel et al., 2009; Höglund et al., 2009b; Reverter and Fortes, 2012). Several QTL regions and genes associated with milk yield, fat yield, protein yield, fat deviation and protein deviation have been reported in previous studies (Boichard et al., 2003; Ashwell et al., 2004; Schnabel et al., 2005; Kolbehdari et al., 2009). Earlier studies have also identified strong functional candidate genes that affect milk production traits such as *DGAT1* and *GHR* (Blott et al., 2003; Grisart et al., 2004). Similarly, important genomic associations for fertility traits were found in previous GWAS studies, including significant QTLs for calving to first service interval (Daetwyler et al., 2008; Höglund et al., 2009a; Sahana et al., 2010), days open (Schulman et al., 2008), cow non-return rate (Höglund et al., 2009a; Olsen et al., 2011), heifer non-return rate (Holmberg and Andersson-Eklund, 2006), daughter pregnancy rate (Schnabel et al., 2005; Cole et al., 2011), age at puberty (Hawken et al., 2012) and interval from first service to last service for cows and heifers (Höglund et al., 2009a).

Most of the previously described genetic variants, however, are not causal, but rather are in linkage disequilibrium with the functional mutation. The level of LD is a limiting factor for the precision of QTL location detection in dairy cattle populations (Pryce et al., 2010). This is because even SNPs at long distances from the QTL may show associations with the phenotypic trait of interest due to extended LD (Höglund et al., 2014). Identifying pathways and genes that are associated with significant SNPs can give us a deeper biological insight into expression mechanisms of the trait under study (Snelling et al., 2013; Szkiba et al., 2014). Refining the position of QTL regions harboring candidate genes, and identifying causal mutations underlying

variation in complex traits, can lead to an increase in accuracy of selection for these traits (Hirschhorn and Daly, 2005).

The goals of the current study were to identify or refine the position of QTL regions for milk production (MILK), fat production (FAT), protein production (PROT), fat deviation (FATD), protein deviation (PROTD), heifer first service to calving interval (FSTCh), calving to first service interval (CTFS), daughter fertility (DF), and days open (DO) in Canadian Holstein dairy cattle. Additionally we performed an enrichment analysis to test for overrepresentation of significant SNPs in biological pathways.

3.3. Materials and Methods

3.3.1. Animals and data

A population of 3,729 North American Holstein bulls was used in this study, which examined nine production and fertility traits: milk production (MILK), fat production (FAT), protein production (PROT), fat and protein deviation (FATD and PROTD respectively), daughter fertility (DF), first service to calving interval (FSTCh), calving to first service interval (CTFS), and days open (DO). The Canadian Dairy Network (CDN) provided available pedigree, genotypes and official evaluations for proven bulls born between 1956 and 2009. Individuals were genotyped using the BovineSNP50K (50k) panel (3,729 bulls) (Illumina, San Diego, CA) or the high density (HD, 777k) SNP panel (2,387 bulls), respectively. In this work only autosomal SNPs were included. For 50k panel, SNP list used for official genomic evaluation by CDN was considered. These SNPs have passed standard quality control measures used by CDN. Quality control (QC) was performed on the HD genotyping data using snp1101 software (Sargolzaei, 2014) and 116,619 SNPs were excluded on the basis of Mendelian error rate higher

> 0.05 (3,566 SNPs), low call rate < 0.9 (6,446 SNPs), with low MAF (1e-06 – 0.5; 61,577 SNPs), excess of heterozygosity > 0.15 (90 SNPs) and excluded by user (46,433 SNPs excluded from sex chromosome and misplaced SNPs). The number of SNPs kept in the analyses before imputation was 40,666 and 657,986 SNPs for 50k and HD panels, respectively.

Genotypes of 3,729 50k animals were imputed to HD with a reference population of 2,387 HD individuals using FImpute V2.2 software (Sargolzaei et al., 2014). Quality control (QC) was performed on the imputed data as well. SNPs with minor allele frequency (MAF) of less than 1% (55,817 SNPs) and high Mendelian error rate more than 5% (74 SNPs) were excluded. After quality control, 602,095 SNPs remained for use in the subsequent association analysis.

Animal Care and Use Committee approval was not obtained for this study because analyses were performed on existing data obtained under standard farm management from commercial dairy farmers and breeders. As previously stated, data used in this research was provided by CDN, which is the organization that runs the national dairy cattle genetic evaluations in Canada. All dairy farmers in Canada must follow "The Code of Practice for the Care and Handling of Dairy Cattle" developed by the National Farm Animal Care Council of Canada (http://www.nfacc.ca/).

3.3.2. Calculating de-regressed proofs:

In this study de-regressed Holstein bull proofs were used as pseudo phenotypes. A bull's published estimated breeding value (EBV) is a weighted mean of his daughters deviations (DD) and his parental average (PA) (VanRaden et al., 2009). The de-regressed bull proofs were computed by CDN as shown below (VanRaden et al., 2009):

$$DE_{prg} = \frac{Rel_{EBV}}{1 - Rel_{EBV}} - \frac{Rel_{PA}}{1 - Rel_{PA}}$$

$$Rel_{DD} = \frac{DE_{prg}}{DE_{prg} + 1}$$

$$DEBV = PA + \frac{(EBV - PA)}{Rel_{DD}}$$

Where, DE_{prg} is the daughter equivalent from progeny information, Rel_{EBV} and Rel_{PA} are the reliabilities of EBV and PA, respectively, Rel_{DD} is the reliability of DD, and DEBV is the de-regressed bull proof.

3.3.3. Genome wide association analysis (GWAS)

Association analysis was performed using a single SNP regression mixed linear model implemented in the snp1101 software (VanRaden, 2008; Sargolzaei, 2014). The mixed linear model was:

$$Y_i = \mu + \beta g_i + a_i + e_i$$

where Y_i is pseudo phenotype of the ith bull (de-regressed bull proofs, DEBV); μ is the overall mean; β is the linear regression coefficient (allele substitution effect) of the SNP; g_i is the SNP genotype of the ith bull, which was coded as 0, 1 and 2 for SNP genotypes BB, AB and AA, respectively; a_i is the random additive polygenic effect of the ith bull and e_i is the random error term.

Assumptions for the model are a_i: $a \sim N\ (0, G\sigma_a^2)$ where G is the genomic relationship matrix (VanRaden, 2008) and σ_a^2 is the polygenic additive genetic variance; e_i: $e \sim N\ (0, R\sigma_e^2)$ where σ_e^2 is the residual variance. R is a diagonal matrix containing weights for the residual variance based on the reliabilities of the de-regressed bull proofs (VanRaden, 2008).

To account for multiple tests, 5% and 1% genome-wise false discovery rate (FDR) were used to identify significant and highly significant associations, respectively. The inflation factor λ (Devlin and Roeder, 1999) and quantile-quantile (Q-Q) plots were calculated to compare observed distributions of –log (P-value) to the expected distribution under the no association model for each trait.

3.3.4. Candidate SNP enrichment analysis

The SNP2GO R package was used for functional analysis of the genome-wide association results (Szkiba et al., 2014). Genomic annotations, associated Gene Ontology terms (GO terms) and the list of significant (P-value < 0.01) and non-significant SNPs (termed "candidate" and "non-candidate" SNPs in the SNP2GO documentation) for each trait were provided as input to SNP2GO, which reports biological pathways or processes that are enriched for significant SNPs. For this analysis, the Ensembl version 78 genomic annotation file for *Bos taurus* UMD 3.1 assembly was used in conjunction with the Ensembl gene ID file from Ensembl version 78, which contains gene ID to GO term associations. The SNP2GO "extension" value was set to 50 nucleotides, which expands the gene region (by 50 nucleotides upstream and 50 nucleotides downstream) when identifying overlaps between genes and markers. The "runs" parameter was set to 100,000 and a false discovery rate (FDR) of 1% was used to correct for multiple testing.

3.4. Results and discussion:

3.4.1. Association analysis

Association analysis identified strong associations for most of the production traits (Figure 3.1 to Figure 3.5) and some of the fertility traits (Figure 3.6 to Figure 3.9) in this study. Representative Manhattan plots are shown in Figure 3.1 for milk production (MILK) and calving to first service interval (CTFS) traits. No significant associations were detected for FSTCh and DF. For those traits yielding significant associations, the number of significant SNPs identified at a genome-wise FDR of 5% varied from 1,416 for PROTD to 8 for DO (Table 3.1). Q-Q plots comparing the observed distribution of –log (P-value) to the expectation under null hypothesis are shown in Figure 3.10 to Figure 3.14 (milk production traits) and Figure 3.15 to Figure 3.18 (fertility traits). The plots show a distribution close to the expected distribution line for CTFS (λ_{median} = 1.0567), DO (λ_{median} = 1.0115) and some production traits (MILK λ_{median} = 1.0056; FAT λ_{median} = 0.9558; PROT λ_{median} = 1.05), whereas there were strong deviations from expectation for FATD (λ_{median} = 0.7844) and PROTD (λ_{median} = 0.8969). When a high-density marker panel is used in single maker association analysis potentially a large number of markers in linkage disequilibrium may display association (and similar low P-values) with the same QTL region. This yields many significant tests that are not independent and, therefore, deviate from the expected distribution of test statistics. However, this does not imply in an overall inflation of P-values. The deviations from expectations observed for milk, fat, protein and fat deviations appear to be due largely to the strong effect of the *DGAT1* gene and the many SNPs in linkage disequilibrium that show some degree of association with it. When BTA14 was excluded from the analyses to assess this possibility, the λ_{median} values for milk production (MILK), fat production (FAT), fat deviation

(FATD) and protein deviation (PROTD) were 1.0605, 1.0896, 1.1005 and 0.9709, respectively, much closer to the expectation.

3.4.2. Production traits:

3.4.2.1. Milk production (MILK)

In total 292 SNPs were found to be significant for milk production at a genome-wise FDR of 5% (Additional file[3] S3.1). Highly significant SNPs (genome-wise 1% FDR) were mostly localized on BTA 5, 6, 14, 15, 20 (Figure 3.1).

Many of the strong associations detected in this study support previously reported regions for the same or correlated traits. For example, a region containing numerous highly significant SNPs on BTA14 includes *DGAT1*, a gene with a major effect on milk fat content (Grisart et al., 2004) and several other production traits (Ashwell et al., 2004; Pimentel et al., 2009; Maxa et al., 2012; Meredith et al., 2012; Xie et al., 2014). A peak on BTA20 overlaps with growth hormone receptor (*GHR*) (Blott et al., 2003; Viitala et al., 2006; Flori et al., 2009) and a nearby peak at 31-32 Mb contains several genes previously identified as potentially influencing milk production, milk fat percentage, lean meat yield, carcass weight, residual feed intake, age at puberty and male fertility traits in beef cattle (Additional file S3.2) (Blaschek et al., 2011; Rolf et al., 2012; Fortes et al., 2013; Doran et al., 2014; Jiang et al., 2014). We have also detected regions that were previously identified on BTA5 at position 87 and 107 Mb for milk yield (Chmuzynska, 2006). The presence of a QTL at confidence interval of 45-52 Mb on BTA6 was confirmed by previously reported significant QTLs found for clinical mastitis in Norwegian Red dairy cattle (Sodeland et al., 2011). A significant peak on BTA15 at the genomic interval of 54 to

[3] Additional files are accessible through the link http://dx.doi.org/10.7939/DVN/10939.

58 Mb in this study is close to a significant SNP that was reported to be associated with persistency of milk production in dairy cattle by Kolbehdari et al. (2009), which is located in an intronic region of the *CD44* gene.

The most significant SNP (BovineHD1400000216: rs134432442) for this trait was located within the *CPSF1* gene on BTA14. Cochran et al. (2013) has reported this SNP to be significantly associated with fat yield and fat percentage in Holstein cattle. Other highly significant (genome-wise FDR ≤ 1%) SNPs in this study are located within genes *TONSL* (BovineHD1400000206: rs137472016), *CYHR1* (BovineHD1400000204: rs137727465), *FOXH1* and *PPP1R16A* (ARS-BFGL-NGS-57820: rs109146371) (Additional file S3.1).

Of the 7,586 SNPs that were introduced as candidate SNPs (P-value < 0.01) to SNP2GO for enrichment analysis, a total of 1,576 were enriched in biological pathways that may provide better insight into key pathways and genes associated with milk production. Enrichment analysis found 545 significant GO terms (genome-wise FDR ≤ 1%) with a minimum of 10 genes associated with each GO term (Additional file S3.3). Included among the top 10 most significant GO terms (FDR ≤ 1%) are those with clear relevance to the trait, such as GO:0007595 (lactation) and GO:0043627 (response to estrogen) whereas for others the relevance is less clear, for example GO:0003785 (actin monomer binding) and GO:0010569 (regulation of double-strand break repair via homologous recombination). The complete list of enriched terms is given in Additional file S3.3.

3.4.2.2. Fat production (FAT) and fat deviation (FATD)

Significant SNPs identified for fat production (FAT) were located on BTA5 and 14 (Additional file S3.1). Similarly, GWAS identified highly significant peaks for fat deviation (FATD) on

BTA5, 6, 14 and 20 (Additional file S3.1). Significant SNPs were mostly located on BTA14 for these two traits (Figure 3.2 and Figure 3.3).

This association analysis supports the presence of QTLs on BTA5 (76 to 107 Mb), BTA14 (0 to 20 Mb and 31 to 63 Mb), BTA20 (22 Mb and 31 to 33 Mb) related to milk production traits reported in multiple studies (Boichard et al., 2003; Kolbehdari et al., 2009; Jiang et al., 2010; Cochran et al., 2013; Jiang et al., 2014; Raven et al., 2014). In addition to a significant peak at 0 to 20 Mb on BTA14, the location of *DGAT1*, we identified another peak with highly significant SNPs at a genomic interval of 60 to 70 Mb associated with FATD and MILK (Additional file S3.1). The presence of significant SNPs associated with milk production has been reported for this region at 63 Mb by Kolbehdari et al. (2009). Heyen et al. (1999) and Ashwell et al. (2004) have also reported three regions on BTA14 at 55 and 69 Mb (between markers D14S55-ILSTS39 and D14S31-CSSM66) and 69 to 79 Mb (between markers BM4305-IRNA100), with a significant effect on fat yield and fat percentage, milk yield and protein yield respectively. The common highly significant SNPs among FAT and FATD were mostly associated with BTA14 and assigned to several known and some newly identified genes including *ARHGAP39* (BovineHD1400000188: rs134892687), *PPP1R16A* (BovineHD1400000199: rs134839376), *GRINA* (BovineHD1400000275: rs133271979), *SMPD5* (BovineHD1400000262: rs135549651) and *MROH1* (BovineHD1400000243: rs133119726).

Chromosome 5 contained the second highest number of significant SNPs for both FAT and FATD (135 and 185, respectively; genome-wise FDR $\leq$ 1%). A significant peak on this chromosome was detected extending between 87 to 100 Mb in this study. This region was reported to be associated with a SNP (ss117963826) in the *GABARAPL1* gene with an antagonistic effect on milk yield and fat percentage (Chmuzynska, 2006). Kolbehdari et al.

(2009) also reported a SNP (rs41592948) in this region linked to the *GABARAPL1* gene with an effect on dairy strength. Furthermore, we identified five significant SNPs (genome-wise FDR ≤ 5%) within previously known QTL regions (BovineHD0500025075: rs137830740; BovineHD0500025146: rs133732696; BovineHD0500025147: rs42406616; BovineHD0500025415: rs109234621; BovineHD0500025488: rs109374096) on BTA5 for FAT trait (Additional file S3.2). These SNPs are all intron variants and are located within genes *ST8SIA1*, *ABCC9*, *SLO1C1* and *PDE3A*. The QTL region that these SNPs are located in were found associated with C22:1 fatty acid content, milk fat yield and sire conception rate in Angus, Brown Swiss and Holstein cattle (Additional file S3.2) (Guo et al., 2012; Saatchi et al., 2013).

Several SNPs within the identified significant peak on BTA20 for FATD in this study (31-32 Mb) are located within growth hormone receptor (*GHR*) gene including BovineHD2000009236: rs109719726; ARS-BFGL-NGS-118998: rs110482506 and UA-IFASA-7069: rs41639261 (Additional file S3.1). Association of significant SNPs affecting milk fat content within *GHR* gene was already reported in German Holstein-Friesian cattle (Wang et al., 2012). This region was also reported to be associated with early embryonic survival, sire conception rate (Ma et al., 2012) and carcass weight in Holstein-Friesian cattle (Doran et al., 2014) (Additional file S3.2). Another highly significant peak region for FATD was located on BTA6 and within a known QTL region (at 37-47 Mb) (Saatchi et al., 2014). This region has also been reported significant for milk protein content, carcass weight and fat production in US cattle breeds (Huang et al., 2012; Saatchi et al., 2014).

In the candidate SNP enrichment analysis, 1,958 SNPs were enriched in biological pathways for FAT (from 7,024 candidate SNPs P-value < 0.01). These SNPs were overrepresented in 330 significant GO terms (Additional file S3.3). Included among the top 10 most significant GO

terms the most relevant ones with the trait are (genome-wise FDR ≤ 1%) GO:0051057 (positive regulation of small GTPase mediated signal transduction) and GO:0030513 (positive regulation of BMP signaling pathway). For other terms the relevance is less clear such as GO:0003281 (ventricular septum development), GO:0003279 (cardiac septum development) and GO:0003215 (cardiac right ventricle morphogenesis). The complete list of enriched GO terms for this trait is shown in Additional file S3.3.

A total of 5,290 candidate SNPs (P-value < 0.01) associated with FATD were overrepresented in 445 enriched GO terms (Additional file S3.3). Among the top seven most significant GO terms the ones with clear relevance to the FATD trait are GO:0042403 (thyroid hormone metabolic process) and GO:0035357 (peroxisome proliferator activated receptor signaling pathway), whereas the relevance is less clear for other GO terms including GO:0033158 (regulation of protein import into nucleus, translocation).

3.4.2.3. Protein production (PROT) and protein deviation (PROTD)

Association analysis detected significant SNPs on BTA 5, 9 and 14 (Additional file S3.1). The largest number of significant SNPs for protein production (PROT) was located on chromosome 14, with 41 significant SNPs (genome-wise 1% FDR) (Figure 3.4). Significant regions detected for protein deviation (PROTD) were identified on chromosomes 3, 5, 6, 10, 14, 15, 20, 26 and 29 (Additional file S3.1) with the majority of significant SNPs located on chromosomes 3, 6, 14 and 20 (Figure 3.5).

Our study detected strong associations on chromosome 20 from 28 to 38 Mb for PROTD. These results are consistent with previous studies reporting QTL regions surrounding *GHR* on chromosome 20 for milk yield and milk composition trait (Blott et al., 2003; Meredith et al.,

2012; Raven et al., 2014). SNPs affecting protein yield and protein percentage on BTA20 associated with growth hormone receptor gene has been reported for several breeds including Holstein (Blott et al., 2003; Chamberlain et al., 2012), and Ayrshire dairy cattle (Viitala et al., 2006). We also identified highly significant SNPs including (ARS-BFGL-NGS-118998: rs110482506) in our study at 32 Mb on BTA20 which supports a QTL reported by Wang et al. (2012) with a SNP downstream of the *GHR* gene.

Significant peaks were detected on chromosome 6, spanning 24 to 40 and 80 to 90 Mb for PROTD. The presence of QTLs on BTA6 in dairy cattle affecting milk production traits, near the casein gene cluster (around 87 Mb) has been reported in multiple breeds and populations including Dutch, US Holstein cattle for milk fat and milk protein yield (Bovenhuis et al., 1992; Georges et al., 1995; Zhang et al., 1998) German Holsteins (Kühn et al., 1999; Freyer et al., 2003) and Brazilian Holstein cattle (Silva et al., 2011) for milk yield and fat yield. Several highly significant SNPs (FDR ≤ 5%) on BTA6 in this study associated with PROT and PROTD (including ARS-BFGL-NGS-42501: rs110388088; BovineHD0600006866: rs109858710; BovineHD0600009641: rs135525961; BovineHD0600009643: rs135142364; BovineHD0600009650: rs137464778) are located within a known QTL region associated with milk whey protein in dairy cattle (Huang et al., 2012). The complete list of these SNPs is given in Additional file S3.2. Also two highly significant SNPs (genome-wise FDR ≤ 1%) (Hapmap24324-BTC-062449 and BTA-121739-no-rs) on BTA6 for PROTD in this study were reported to be very close to *ABCG2* gene (in a distance of 100 kb) (Cohen-Zinder et al., 2005; Fang et al., 2014). Variants close and within this gene were associated with protein percentage in Chinese dairy cattle (Cohen-Zinder et al., 2005; Fang et al., 2014).

The peak detected for PROTD on chromosome 3 at 10 to 34 Mb is supported by previous investigations (Additional file S3.2) (Lipkin et al., 1998; Chmuzynska, 2006; Kolbehdari et al., 2009; Jiang et al., 2014). For example, four highly significant SNPs (genome-wise FDR ≤ 1%) within this region (at 11 Mb) in this study (BovineHD0300003802: rs109857972; BovineHD0300003805: rs136467848; BovineHD0300003811: rs132784836 and BovineHD0300003779: rs110122034) are located within a reported QTL region associated with protein percentage in German Holstein cattle (Chmuzynska, 2006). Two other SNPs in our study (BTB-00604223: rs41769311) at 52 Mb and (ARS-BFGL-NGS-4613: rs110428369) at 54 Mb on BTA15 were also found to be associated with protein percentage in Irish Holstein cattle (Meredith et al., 2012).

We identified five significant SNPs in this study for PROT on BTA5 at 88 Mb including BovineHD0500025150: rs42406611; BovineHD0500025181: rs109795387 and BovineHD0500025189: rs136903701 within the gene *ABCC9*. This region was detected to be significantly associated with protein yield (at 75 to 110 Mb) in Cochran et al. (2013) study.

Enrichment analysis of the candidate SNPs for PROT found 2,280 SNPs overrepresented in 658 GO terms (FDR ≤ 1%). Included among the top 10 most highly significant GO terms (FDR ≤ 1%) are those GO terms with a more clear relevance to the trait such as GO:0040037 (negative regulation of fibroblast growth factor receptor signaling pathway) and GO:0021903 (rostrocaudal neural tube patterning), whereas for others the relevance is less clear such as GO:0097228 (sperm principal piece). The complete list of enriched GO terms is provided in Additional file S3.3.

Candidate SNP enrichment analysis for PROTD identified 447 significant GO terms potentially involved in pathways affecting protein production deviation (Additional file S3.3). These

enriched GO terms were overrepresented by a total number of 2,557 highly significant SNPs. The most relevance GO terms among the top 10 most significant GO terms (FDR ≤ 1%) are GO:0007595 (lactation), GO:1902742 (apoptotic process involved in development), GO:0010257 (NADH dehydrogenase complex assembly) and GO:0033108 (mitochondrial respiratory chain complex assembly). The complete list of the significant GO terms is given in Additional file S3.3.

3.4.3. Fertility traits:

In balanced genomic selection programs, one aim is to select bulls which sire daughters with higher reproductive performance (daughters showing an early heat in the mating period with a high probability of conception) (Haile-Mariam et al., 2003). To examine fertility, we first focused on daughter performance traits including daughter fertility and then examined the regions associated with success of insemination traits such as calving interval and days open. Association analysis did not detect any significant SNP for daughter fertility or heifer first service to calving interval (Figure 3.7 and Figure 3.8). The identified chromosomes and significant regions for CTFS and DO are discussed in detail below. In order to detect the regions with potential significant pleiotropic effects on both production and fertility traits, we also investigated the overlapping regions between these two traits (Table 3.2 and Table 3.3). No overlapping region was detected between production and fertility traits in this study.

3.4.3.1. Calving to first service interval (CTFS)

A total of 20 SNPs (genome-wise FDR ≤ 5%) were found to be associated with calving to first service interval (Figure 1 B). These SNPs were mostly located on BTA13 and 21 (Additional file

S3.4). The most significant SNP (genome-wise 1% FDR) was BovineHD2100017054: rs136777407 (on BTA21), which is an intronic variant within the gene *FAM181A*.

Little is known about the function of this gene, although it was shown that methylation of this gene increases during mid-secretory phase of progesterone (P4) hormone in human endometrium (Houshdaran et al., 2014); This region (at 47-59 Mb) on BTA21 has not been previously reported to be associated with fertility traits in cattle or other species. The region also includes the *ASB2* gene. Therefore, these two genes can be considered as potential candidate genes affecting CTFS and calving interval associated traits in dairy cattle. The two other highly significant SNPs (genome-wise FDR $\leq$ 5%) in this study on BTA21 at the same region (47-57 Mb) were BovineHD2100016620: rs136994701, within the *SLC24A4* gene, and BovineHD2100013476: rs136407309 a 3-prime UTR variant within the *NKX2-1* gene. The *SLC24A4* gene encodes a member of potassium-dependent sodium or calcium exchanger protein family and has been identified to be associated with hair color, skin pigmentation (Han et al., 2008) and eye color in humans (Liu et al., 2010). The gene *NKX2-1* encodes a transcription factor that regulates the expression of thyroid-specific genes (Gene ID: 7080) as well as genes that are involved in morphogenesis (Gene ID: 7080). However, no association of these two genes with fertility trait has been previously reported.

We also detected two significant SNPs associated with CTFS at 30-32 Mb on BTA13; this region has previously been reported by Sahana et al (2010) as associated with interval from calving to first insemination and fertility index traits (Sahana et al., 2010). These two SNPs on BTA13 (BovineHD1300008936: rs136296491 and BovineHD1300009502: rs109815929) are located within candidate genes *FAM188A* and *MRC1,* respectively. The location of these SNPs is within known QTL regions associated with Inhibin level, shear force and lean meat yield in

several cattle breeds (Additional file S3.5). The single significant SNP on BTA5 (BovineHD0500025143: rs42718239) is located within gene *ABCC9* at 88 Mb. This region previously has been validated for calving to first insemination (ICF) trait in Nordic Holstein, Nordic Red and Jersey cattle breed but the gene *ABCC9* was not assigned to any of the associated SNPs (Höglund et al., 2014). This SNP is also located within a known QTL region associated with sire conception rate in Holstein dairy cattle (Peñagaricano et al., 2012). The protein coded by *ABCC9* gene is thought to form ATP-sensitive potassium channels in cardiac, skeletal, vascular and non-vascular smooth muscle (Gene ID: 10060). ATP-sensitive potassium channels are expressed in many tissues and regulate different cellular functions by coupling cell metabolism with membrane potential (Curley et al., 2002). The opening of these channels results in a flow of K^+ ions and thus, reducing cellular excitability and contractility (Khan et al., 2001). The K^+ channel opening components are reported as potent inhibitors of human myometrial contractility (Curley et al., 2002). It has been indicated that down-regulation of ATP-sensitive potassium channel (K_{ATP} channel) subunits may facilitate myometrial function during late pregnancy in humans (Curley et al., 2002). In another study in humans, it has been revealed that estrogen may induce the activation of K_{ATP} channels to promote cell proliferation in the myometrial (Park et al., 2008). Therefore, the *ABCC9* gene can be considered as a potential candidate gene in dairy cattle and may have a role in cell proliferation of myometrial cells and resuming the reproduction cycle after calving.

Of 8,834 SNPs that were introduced to SNP2GO (P-value < 0.01), 2,330 SNPs were overrepresented in biological pathways and molecular functions for CTFS. These SNPs were associated with 459 significant GO terms with FDR ≤ 1%. The GO terms with the clearest relevance for CTFS trait among the top nine most significant GO terms (FDR ≤ 1%) are

GO:0040019 (positive regulation of embryonic development) and GO:0045663 (positive regulation of myoblast differentiation), while for others the relevance is less clear, for example GO:0045836 (positive regulation of mitotic nuclear division). The complete list of significant GO terms is given in Additional file S3.6.

3.4.3.2. Days open (DO)

Association analysis identified a total of eight highly significant SNPs (genome-wise 1% FDR) on BTA21 associated with DO (Additional file S3.4; Figure 3.9). These significant SNPs span the 53-59 Mb region. The peak on chromosome 21 overlaps with the one detected for CTFS, and has not been previously reported for this trait. As noted above, this peak includes gene *FAM181A*. Schulman et al. (2008) reported a QTL associated with days open EBV on chromosomes 1, 2, 5, and 25 for Finnish Ayrshire dairy cattle. These regions were not detected in our study.

The number of candidate SNPs overrepresented in the enrichment analysis was 1,897 (P-value $\leq$ 0.01). These SNPs were involved in 381 significant GO terms for days open (Additional file S3.6). Among the top 10 most significant GO terms, the ones with clear relevance to the trait are GO:0043268 (positive regulation of potassium ion transport), GO:0071353 (cellular response to interleukin-4). Additional GO terms, for which the relationship to the trait is less clear, include GO:0000786 (nucleosome) and GO:0005070 (SH3/SH2 adaptor activity). The complete list of significant GO terms is given in Additional file S3.6.

3.4.4. Overlapping regions among milk production traits

The only overlapping peak among all of the production traits (MILK, FAT, FATD, PROT and PROTD) was the region identified on BTA14 (Table 3.2). This region spans around 1.4 to 2.9 Mb and includes 74 highly significant SNPs (genome-wise FDR ≤ 1%).

3.4.5. Overlapping regions among fertility traits

Investigating overlapping regions among fertility traits has resulted in identifying eight significant SNPs on BTA21 at 53-59 Mb between two fertility traits CTFS and DO in this study (Table 3.3). The overlapping region among these fertility traits was not reported to be associated with days open or calving to first interval traits in previous GWAS.

3.5. Conclusion

Genome wide association analysis in this study detected several regions associated with milk production and female fertility in Canadian Holstein cattle. Most of the regions in this study were identified in other independent studies. However, novel regions of association were detected. Our result shows a novel significant region on chromosome 21 (at 47-59 Mb), which overlaps among CTFS and DO, which was not reported for fertility traits in previous association studies. This region includes several genes including *FAM181A*, *SLC24A4* and *NKX2-1*. The inclusion of several traits in one study allowed us to more easily compare overlaps that might, for example, highlight regions with pleiotropic effects. Although overlaps were observed within production and fertility traits, we did not see any overlap between production and fertility in this study. GO term enrichment analysis of the GWAS results identified terms consistent with the known physiology of the traits as well as novel or unexpected terms. The chromosomal regions

identified in this study confirm several previous findings affecting production traits. Our result could also highlight new regions and pathways that may contribute to variation in fertility trait in dairy cattle. These novel regions can be used for further functional analysis to identify genes, gene networks and variants that explain variation in these traits.

3.6. Legend

Additional files are accessible through http://dx.doi.org/10.7939/DVN/10939.

3.7. References

Ashwell MS, Heyen DW, Sonstegard TS, et al (2004) Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. J Dairy Sci 87:468–75. doi: 10.3168/jds.S0022-0302(04)73186-0

Blaschek M, Kaya A, Zwald N, et al (2011) A whole-genome association analysis of noncompensatory fertility in Holstein bulls. J Dairy Sci 94:4695–4699. doi: 10.3168/jds.2010-3728

Blott S, Kim J, Moisio S, et al (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163:253–66.

Boichard DB, Grohs C, Bourgeois F, et al (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. Genet Sel Evol 35:77–101. doi: 10.1051/gse

Bovenhuis H, Van Arendonk J a, Korver S (1992) Associations between milk protein polymorphisms and milk production traits. J Dairy Sci 75:2549–2559. doi: 10.3168/jds.S0022-0302(92)78017-5

Chamberlain a J, Hayes BJ, Savin K, et al (2012) Validation of single nucleotide polymorphisms associated with milk production traits in dairy cattle. J Dairy Sci 95:864–75. doi: 10.3168/jds.2010-3786

Chmuzynska (2006) The multigene family of fatty acid-binding proteins (FABPs): Function,

structure and polymorphism. J Appl Genet 47:39–48.

Cochran SD, Cole JB, Null DJ, Hansen PJ (2013) Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genet 14:49. doi: 10.1186/1471-2156-14-49

Cohen-Zinder M, Seroussi E, Larkin DM, et al (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res 15:936–944. doi: 10.1101/gr.3806705

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

Curley M, Cairns MT, Friel AM, et al (2002) Expression of mRNA transcripts for ATP-sensitive potassium channels in human myometrium. Mol Hum Reprod 8:941–945.

Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JAB (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. J Dairy Sci 91:3225–36. doi: 10.3168/jds.2007-0333

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004. doi: 10.1111/j.0006-341X.1999.00997.x

Doran AG, Berry DP, Creevey CJ (2014) Whole genome association study identifies regions of

the bovine genome and biological pathways involved in carcass trait performance in Holstein-Friesian cattle. BMC Genomics 15:837. doi: 10.1186/1471-2164-15-837

Fang M, Fu W, Jiang D, et al (2014) A Multiple-SNP Approach for Genome-Wide Association Study of Milk Production Traits in Chinese Holstein Cattle. PLoS One 9:e99544. doi: 10.1371/journal.pone.0099544

Flori L, Fritz S, Jaffrézic F, et al (2009) The genome response to artificial selection: a case study in dairy cattle. PLoS One 4:e6595. doi: 10.1371/journal.pone.0006595

Fortes MRS, Deatley KL, Lehnert SA, et al (2013) Genomic regions associated with fertility traits in male and female cattle: advances from microsatellites to high-density chips and beyond. Anim Reprod Sci 141:1–19. doi: 10.1016/j.anireprosci.2013.07.002

Freyer G, Sørensen P, Kühn C, et al (2003) Search for pleiotropic QTL on chromosome BTA6 affecting yield traits of milk production. J Dairy Sci 86:999–1008. doi: 10.3168/jds.S0022-0302(03)73683-2

Georges M, Georges M, Nielsen D, et al (1995) Mapping Quantitative Trait Loci Controlling Milk Production in Dairy Cattle by Exploiting Progeny Testing. Genet Soc Am 139:907–920.

Grisart B, Farnir F, Karim L, et al (2004) Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc Natl Acad Sci U S A 101:2398–403.

Guo J, Jorjani H, Carlborg Ö (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. doi: 10.1186/1471-2156-13-82

Haile-Mariam M, Morton JM, Goddard ME (2003) Estimates of genetic parameters for fertility traits of Australian Holstein-Friesian cattle. Anim Sci 76:35–42.

Han J, Kraft P, Nan H, et al (2008) A Genome-Wide Association Study Identifies Novel Alleles Associated with Hair Color and Skin Pigmentation. PLoS Genet 4:e1000074. doi: 10.1371/journal.pgen.1000074

Hawken RJ, Zhang YD, Fortes MRS, et al (2012) Genome - wide association studies of female reproduction in tropically adapted beef cattle. J Anim Sci 90:1398–1410. doi: 10.2527/jas2011-4410

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108. doi: 10.1038/nrg1521

Höglund JK, Buitenhuis AJ, Guldbrandtsen B, et al (2009a) Overlapping chromosomal regions for fertility traits and production traits in the Danish Holstein population. J Dairy Sci 92:5712–9. doi: 10.3168/jds.2008-1964

Höglund JK, Guldbrandtsen B, Su G, et al (2009b) Genome scan detects quantitative trait loci affecting female fertility traits in Danish and Swedish Holstein cattle. J Dairy Sci 92:2136–43. doi: 10.3168/jds.2008-1104

Höglund JK, Sahana G, Guldbrandtsen B, Lund MS (2014) Validation of associations for female fertility traits in Nordic Holstein, Nordic Red and Jersey dairy cattle. BMC Genet 15:8. doi:

10.1186/1471-2156-15-8

Holmberg M, Andersson-Eklund L (2006) Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. J Dairy Sci 89:3664–71. doi: 10.3168/jds.S0022-0302(06)72406-7

Houshdaran S, Zelenko Z, Irwin JC, Giudice LC (2014) Human Endometrial DNA Methylome Is Cycle-Dependent and Is Associated With Gene Expression Regulation. Mol Endocrinol 28:1118–1135. doi: 10.1210/me.2013-1340

Huang W, Peñagaricano F, Ahmad KR, et al (2012) Association between milk protein gene variants and protein composition traits in dairy cattle. J Dairy Sci 95:440–9. doi: 10.3168/jds.2011-4757

Jiang L, Liu J, Sun D, et al (2010) Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5:e13661. doi: 10.1371/journal.pone.0013661

Jiang L, Liu X, Yang J, et al (2014) Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. BMC Genomics 15:1105. doi: 10.1186/1471-2164-15-1105

Khan RN, Matharoo-Ball B, Arulkumaran S, Ashford ML (2001) Potassium channels in the human myometrium. Exp Physiol 86:255–264.

Kolbehdari D, Wang Z, Grant JR, et al (2009) A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. J Anim Breed Genet 126:216–27. doi: 10.1111/j.1439-0388.2008.00793.x

Kühn C, Freyer G, Weikard R, et al (1999) Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim Genet 30:333–340. doi: 10.1046/j.1365-2052.1999.00487.x

Lipkin E, Mosig MO, Darvasi A, et al (1998) Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: Analysis of milk protein percentage. Genetics 149:1557–1567.

Liu F, Wollstein A, Hysi PG, et al (2010) Digital Quantification of Human Eye Color Highlights Genetic Association of Three New Loci. PLoS Genet 6:e1000934. doi: 10.1371/journal.pgen.1000934

Lucy MC (2001) Reproductive loss in high-producing dairy cattle: where will it end? J Dairy Sci 84:1277–93. doi: 10.3168/jds.S0022-0302(01)70158-0

Ma L, Wiggans GR, Wang S, et al (2012) Effect of sample stratification on dairy GWAS results. BMC Genomics 13:536. doi: 10.1186/1471-2164-13-536

Maxa J, Neuditschko M, Russ I, et al (2012) Genome-wide association mapping of milk production traits in Braunvieh cattle. J Dairy Sci 95:5357–64. doi: 10.3168/jds.2011-4673

Meredith BK, Kearney FJ, Finlay EK, et al (2012) Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet 13:21. doi: 10.1186/1471-2156-13-21

Meuwissen T, Goddard M (2010) Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185:623–31. doi: 10.1534/genetics.110.116590

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829.

Nebel R., Dransfield M., Jobst S., Bame J. (2000) Automated electronic systems for the detection of oestrus and timing of AI in cattle. Anim Reprod Sci 60–61:713–723. doi: 10.1016/S0378-4320(00)00090-7

Olsen HG, Hayes BJ, Kent MP, et al (2011) Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. Anim Genet 42:466–74. doi: 10.1111/j.1365-2052.2011.02179.x

Park S-H, Ramachandran S, Kwon S-H, et al (2008) Upregulation of ATP-sensitive potassium channels for estrogen-mediated cell proliferation in human uterine leiomyoma cells. Gynecol Endocrinol 24:250–256. doi: 10.1080/09513590801893315

Peñagaricano F, Weigel KA, Khatib H (2012) Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. Anim Genet 43:65–71. doi: 10.1111/j.1365-2052.2012.02350.x

Pimentel ECG, Tietze M, Simianer H, et al (2009) Study of Relationships Between Production and Fertility Traits in Dairy Cattle Using Genomic Data.

Pryce JE, Bolormaa S, Chamberlain a J, et al (2010) A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J Dairy Sci 93:3331–45. doi: 10.3168/jds.2009-2893

Raven L-A, Cocks BG, Hayes BJ (2014) Multibreed genome wide association can improve

precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:62. doi: 10.1186/1471-2164-15-62

Reverter A, Fortes MRS (2012) Building single nucleotide polymorphism-derived gene regulatory networks: Towards functional genomewide association studies. J Anim Sci 91:530–536. doi: 10.2527/jas.2012-5780

Rolf MM, Taylor JF, Schnabel RD, et al (2012) Genome-wide association analysis for feed efficiency in Angus cattle. Anim Genet 43:367–374. doi: 10.1111/j.1365-2052.2011.02273.x

Saatchi M, Garrick DJ, Tait RG, et al (2013) Genome-wide association and prediction of direct genomic breeding values for composition of fatty acids in Angus beef cattle. BMC Genomics 14:730. doi: 10.1186/1471-2164-14-730

Saatchi M, Schnabel RD, Taylor JF, Garrick DJ (2014) Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics 15:442. doi: 10.1186/1471-2164-15-442

Sahana G, Guldbrandtsen B, Bendixen C, Lund MS (2010) Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. Anim Genet 41:579–588. doi: 10.1111/j.1365-2052.2010.02064.x

Sargolzaei M (2014) SNP1101 User ' s Guide. Version 1.0.

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. doi: 10.1186/1471-

2164-15-478

Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS (2005) Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Anim Genet 36:408–16. doi: 10.1111/j.1365-2052.2005.01337.x

Schulman NF, Sahana G, Lund MS, et al (2008) Quantitative trait loci for fertility traits in Finnish Ayrshire cattle. Genet Sel Evol 40:195–214. doi: 10.1186/1297-9686-40-2-195

Silva AA, Azevedo ALS, Verneque RS, et al (2011) Quantitative trait loci affecting milk production traits on bovine chromosome 6 in zebuine Gyr breed. J Dairy Sci 94:971–980. doi: 10.3168/jds.2009-2970

Snelling WM, Cushman R a., Keele JW, et al (2013) Networks and pathways to guide genomic selection. J Anim Sci 91:537–552. doi: 10.2527/jas2012-5784

Sodeland M, Kent MP, Olsen HG, et al (2011) Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. Anim Genet 42:457–465. doi: 10.1111/j.1365-2052.2010.02165.x

Szkiba D, Kapun M, von Haeseler A, Gallach M (2014) SNP2GO: functional analysis of genome-wide association studies. Genetics 197:285–9. doi: 10.1534/genetics.113.160341

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423. doi: 10.3168/jds.2007-0980

VanRaden PM, Van Tassell CP, Wiggans GR, et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24. doi:

10.3168/jds.2008-1514

Viitala S, Szyda J, Blott S, et al (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. doi: 10.1534/genetics.105.046730

Wang M, Moisá S, Khan MJ, et al (2012) MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation. J Dairy Sci 95:6529–35. doi: 10.3168/jds.2012-5748

Xie Y, Yang S, Cui X, et al (2014) Identification and expression pattern of two novel alternative splicing variants of EEF1D gene of dairy cattle. Gene 534:189–96. doi: 10.1016/j.gene.2013.10.061

Zhang Q, Boichard D, Hoeschele I, et al (1998) Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics 149:1959–1973

Table 3. 1: The number of significant SNPs (5% and 1% genome-wise FDR) using single SNP regression mixed linear model on imputed BovineHD (777k) genotypes in Holstein dairy cattle.

Trait	Heritability	No. of sig SNPs (1% FDR)	No. of sig. SNPs (5% FDR)
Milk production	0.410	221	292
Fat production	0.340	595	813
Protein production	0.370	41	87
Fat deviation (%)	0.370	998	1,230
Protein deviation (%)	0.370	912	1,416
Daughter fertility	0.070	0	0
First service to calving interval (heifer)	0.033	0	0
Calving to first service interval	0.072	8	20
Days open	0.102	8	0

* FDR: false discovery rate

** Sig. SNPs: Significant SNPs

Heritability is calculated by CDN and used in national genetic evaluation.

Table 3. 2: Common significant SNPs (genome-wise 1% FDR) identified among all production related traits (MILK, FAT, PROT, PROTD) using single SNP regression mixed linear model on imputed BovineHD (777K) genotypes in Holstein cattle.

SNPID	Chromosome	Position	P-value* (nominal)	FDR_genome-wise
BovineHD1400000188	14	1588879	1.00E-300	0.01
BovineHD1400000199	14	1638045	1.00E-300	0.01
BovineHD1400000200	14	1640406	1.00E-300	0.01
ARS-BFGL-NGS-57820	14	1651311	1.00E-300	0.01
BovineHD1400000204	14	1667797	1.00E-300	0.01
BovineHD1400000206	14	1679844	1.00E-300	0.01
BovineHD1400000216	14	1736599	1.00E-300	0.01
ARS-BFGL-NGS-4939	14	1801116	1.00E-300	0.01
BovineHD1400000275	14	2019390	1.00E-300	0.01
BovineHD1400000281	14	2046297	1.00E-300	0.01
ARS-BFGL-NGS-107379	14	2054457	7.45E-256	0.01
BovineHD1400000262	14	1967325	6.45E-252	0.01
BovineHD1400000282	14	2051225	2.26E-245	0.01
BovineHD1400000283	14	2057629	2.26E-245	0.01
BovineHD1400000285	14	2066638	2.26E-245	0.01
BovineHD1400000286	14	2069181	2.26E-245	0.01
BovineHD1400000287	14	2076458	2.26E-245	0.01
BovineHD1400000290	14	2089613	2.26E-245	0.01

BovineHD1400000324	14	2257386	1.09E-150	0.01
Hapmap30383-BTC-005848	14	1489496	9.35E-147	0.01
BovineHD1400000443	14	2758369	2.47E-106	0.01
ARS-BFGL-NGS-34135	14	1675278	6.44E-104	0.01
BovineHD1400000207	14	1683767	1.40E-101	0.01
BovineHD4100010518	14	1672047	3.94E-99	0.01
BovineHD1400000205	14	1672913	3.94E-99	0.01
ARS-BFGL-NGS-94706	14	1696470	4.68E-99	0.01
BovineHD1400000462	14	2857000	8.68E-94	0.01
BovineHD1400000326	14	2273502	1.26E-86	0.01
BovineHD1400000482	14	2940147	9.19E-86	0.01
BovineHD1400000457	14	2833552	1.82E-85	0.01
BovineHD1400000463	14	2875999	5.11E-78	0.01
BovineHD1400000464	14	2883623	5.11E-78	0.01
BovineHD1400000465	14	2885790	5.11E-78	0.01
BovineHD1400000467	14	2898515	5.11E-78	0.01
BovineHD1400000468	14	2901016	5.11E-78	0.01
BovineHD1400000469	14	2902668	5.11E-78	0.01
BovineHD1400000470	14	2904752	5.11E-78	0.01
ARS-BFGL-NGS-18858	14	2909929	5.11E-78	0.01
BovineHD1400000471	14	2910531	5.11E-78	0.01

BovineHD1400000288	14	2084067	2.16E-76	0.01
BovineHD1400000445	14	2763040	1.23E-57	0.01
BovineHD1400000446	14	2763621	1.23E-57	0.01
BovineHD1400000442	14	2756339	1.97E-56	0.01
BovineHD1400000480	14	2936478	1.25E-47	0.01
BTA-34956-no-rs	14	1514056	4.89E-47	0.01
BovineHD1400000167	14	1493001	1.58E-46	0.01
BovineHD4100010510	14	1494482	1.58E-46	0.01
BovineHD4100010511	14	1497296	1.58E-46	0.01
BovineHD1400000170	14	1504519	1.58E-46	0.01
BovineHD4100010512	14	1509091	1.58E-46	0.01
BovineHD1400000172	14	1514738	2.04E-46	0.01
BovineHD4100010513	14	1516428	2.04E-46	0.01
BovineHD1400000173	14	1517553	2.04E-46	0.01
BovineHD4100010514	14	1518492	2.04E-46	0.01
BovineHD1400000174	14	1522098	2.04E-46	0.01
BovineHD4100010515	14	1524573	2.04E-46	0.01
BovineHD4100010516	14	1527842	2.04E-46	0.01
BovineHD4100010517	14	1529914	2.04E-46	0.01
BovineHD1400000166	14	1487447	5.89E-46	0.01
BovineHD1400000164	14	1486102	1.37E-45	0.01
BovineHD1400000165	14	1486682	1.37E-45	0.01
BovineHD1400000143	14	1427669	4.48E-44	0.01
BovineHD1400000429	14	2676321	7.35E-40	0.01
BovineHD1400000145	14	1429389	8.84E-36	0.01

BovineHD1400000152	14	1439476	8.84E-36	0.01
BovineHD4100010506	14	1442028	2.44E-35	0.01
BovineHD1400000154	14	1444759	2.44E-35	0.01
BovineHD1400000277	14	2027812	8.52E-25	0.01
BovineHD1400000273	14	2006872	1.28E-19	0.01
BovineHD1400000258	14	1949077	1.71E-19	0.01
BovineHD1400000259	14	1951389	1.71E-19	0.01
ARS-BFGL-NGS-71749	14	1954317	1.71E-19	0.01
BovineHD1400000267	14	1988704	4.14E-19	0.01
BovineHD1400000270	14	1999406	4.14E-19	0.01

* P-value (nominal): The most significant P-value for each SNP among all milk production traits

Table 3. 3: Common significant SNPs (genome-wise 1% FDR) identified among all fertility related traits (CTFS and DO) using single SNP regression mixed linear model on the imputed BovineHD (777K) genotypes in Holstein cattle.

SNPID	Chromosome	Position	P-value* (nominal)	FDR_genome-wise
BovineHD2100017054	21	59118105	4.97E-11	0.01
BovineHD2100017056	21	59123690	8.58E-11	0.01
ARS-BFGL-NGS-22156	21	55059767	1.09E-10	0.01
BovineHD2100015321	21	53432378	8.52E-10	0.01
BTB-00649131	21	53457352	8.52E-10	0.01
BTB-00649148	21	53500339	8.52E-10	0.01
BovineHD2100015609	21	54594098	2.11E-07	0.05
BovineHD2100015610	21	54600298	2.11E-07	0.05

* P-value (nominal): The most significant P-value for each SNP among all milk production traits

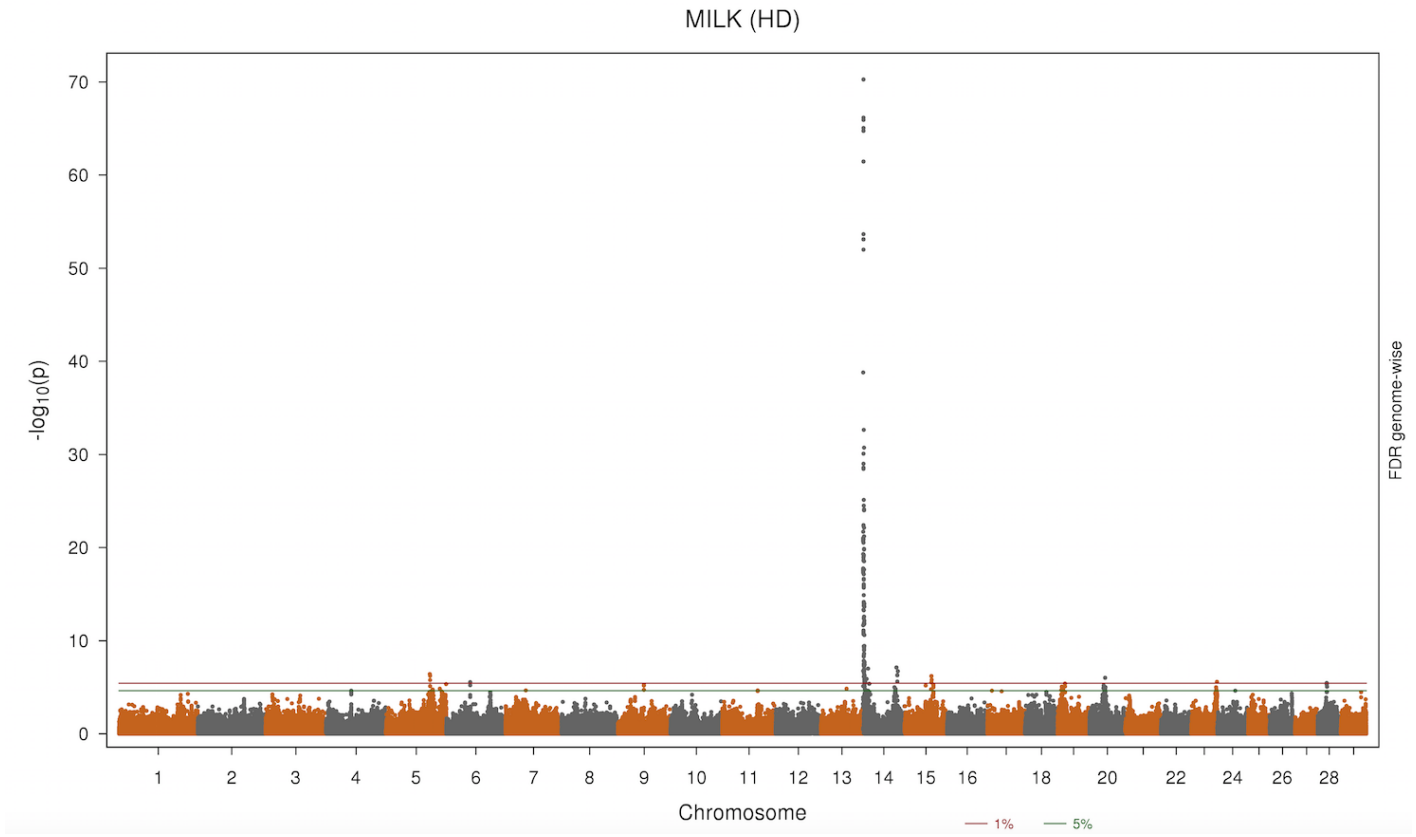


Figure 3. 1: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for milk production (MILK). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

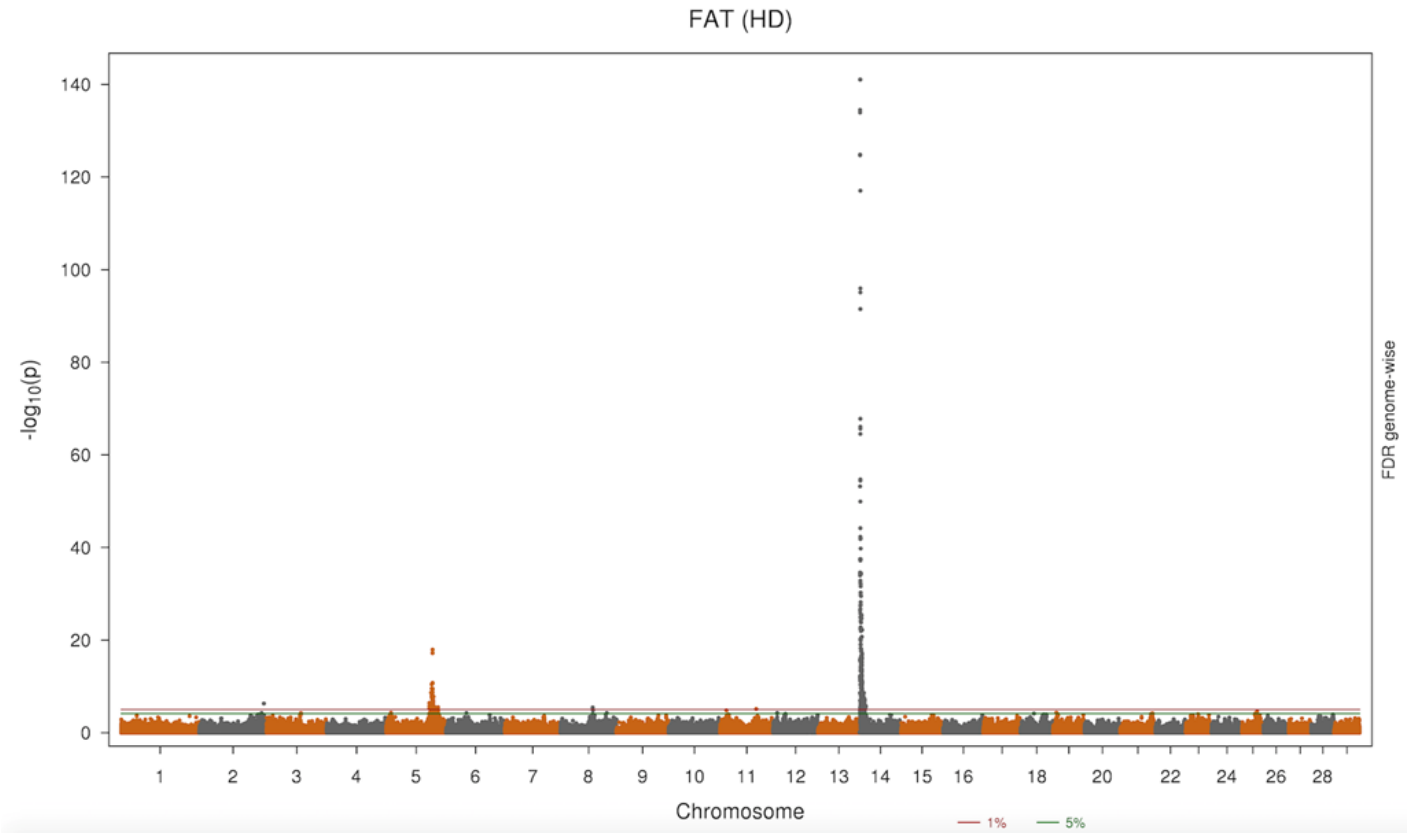


Figure 3. 2: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for fat production (FAT). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

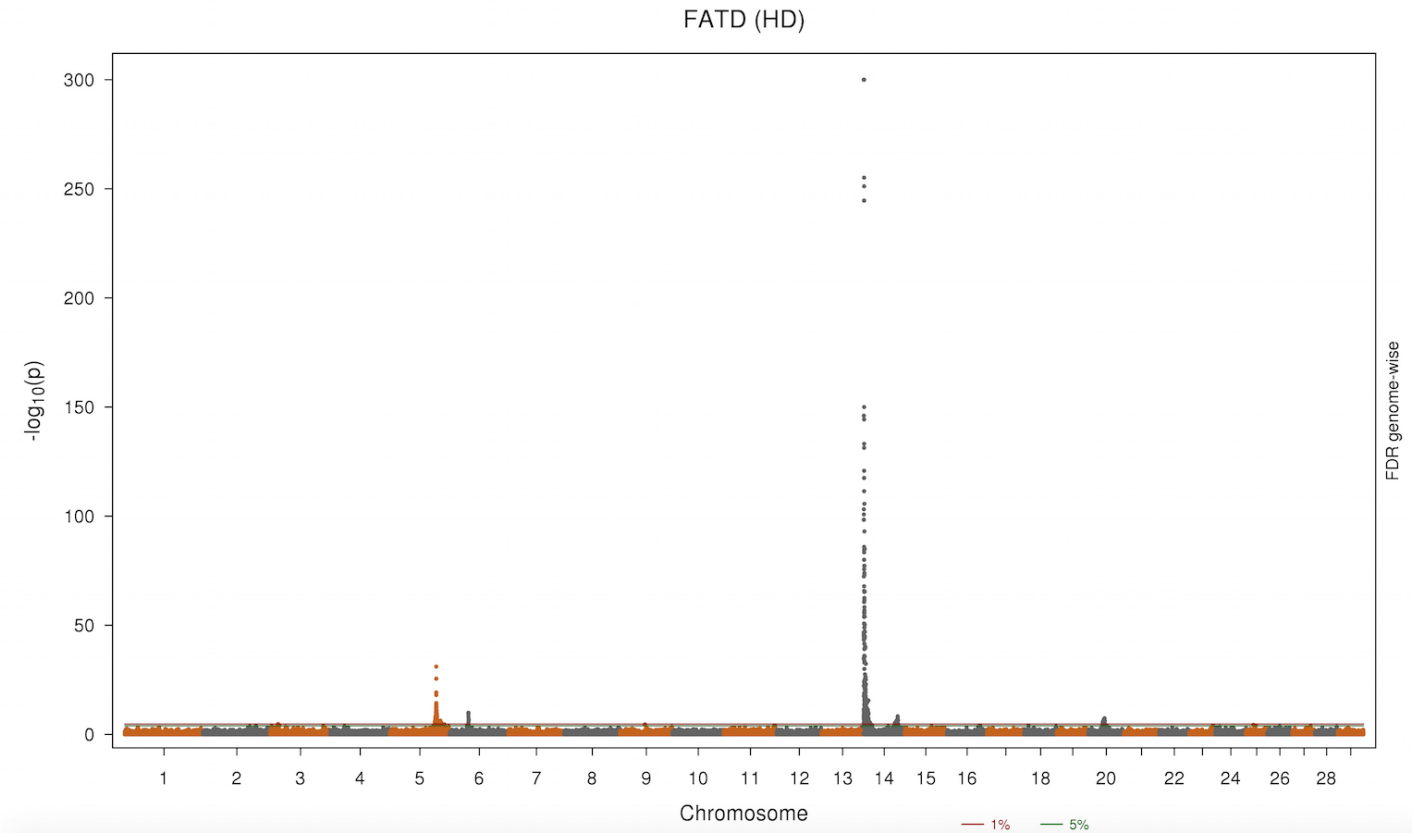


Figure 3. 3: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for fat deviation (FATD). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

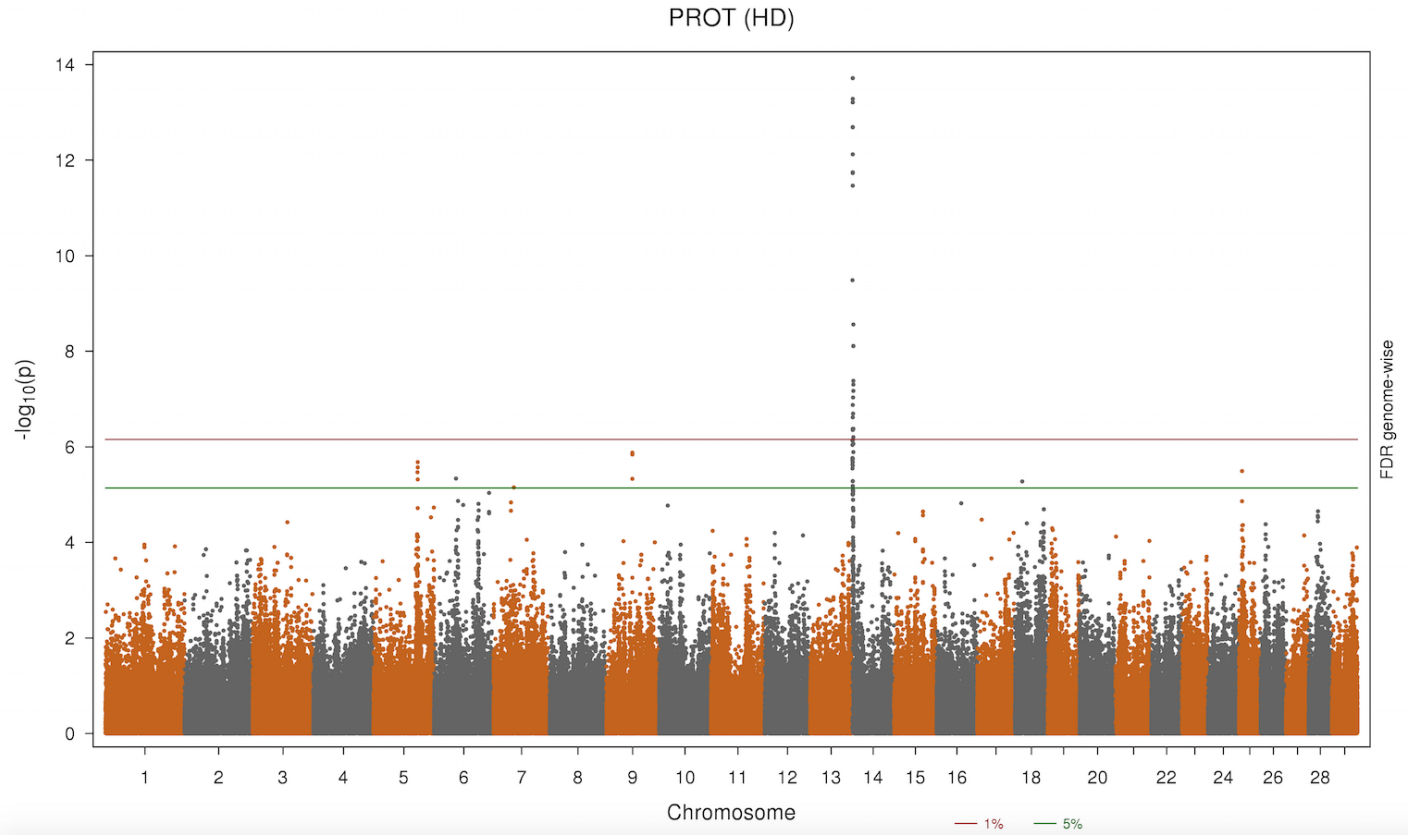


Figure 3. 4: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for protein production (PROT). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

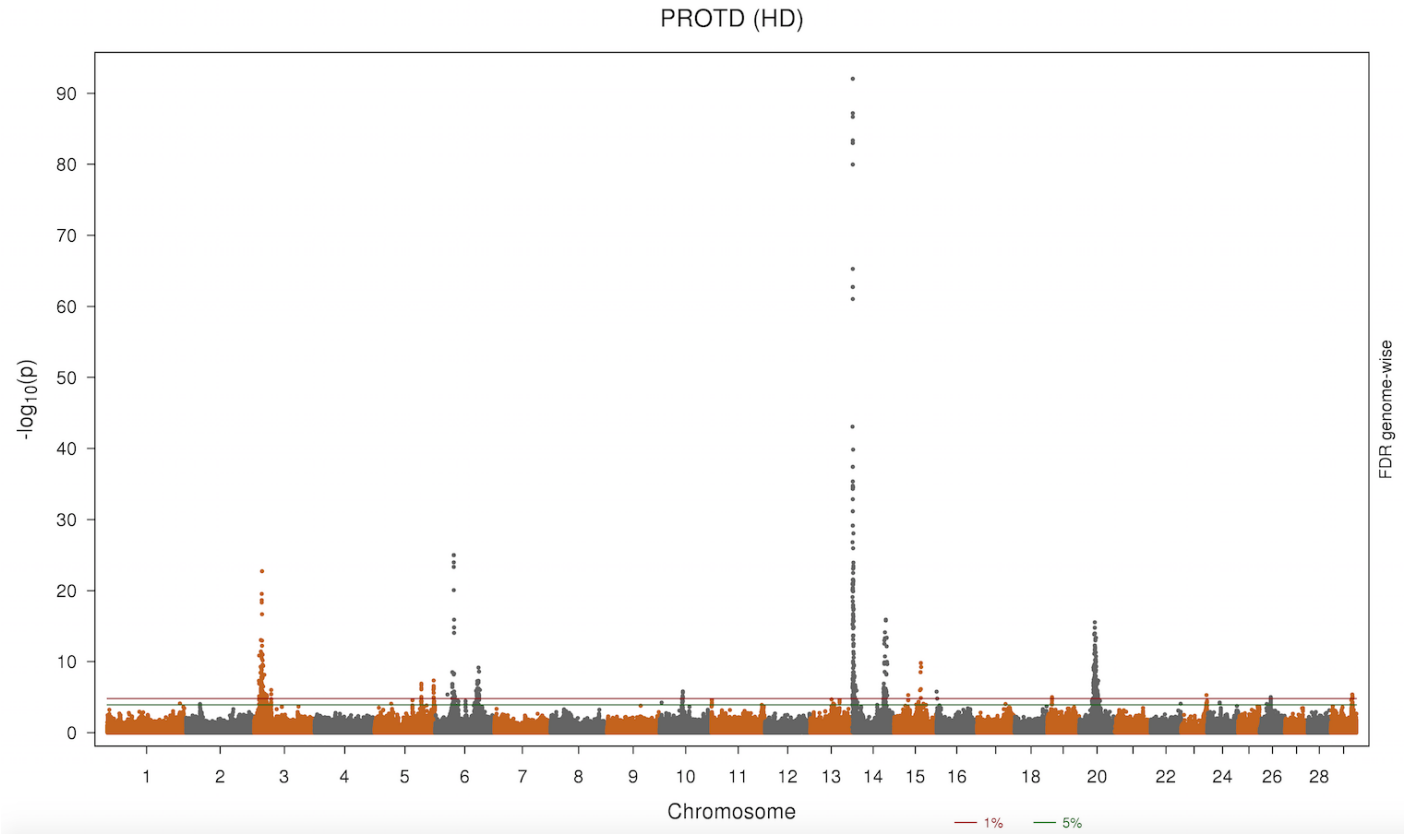


Figure 3. 5: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for protein deviation (PROTD). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

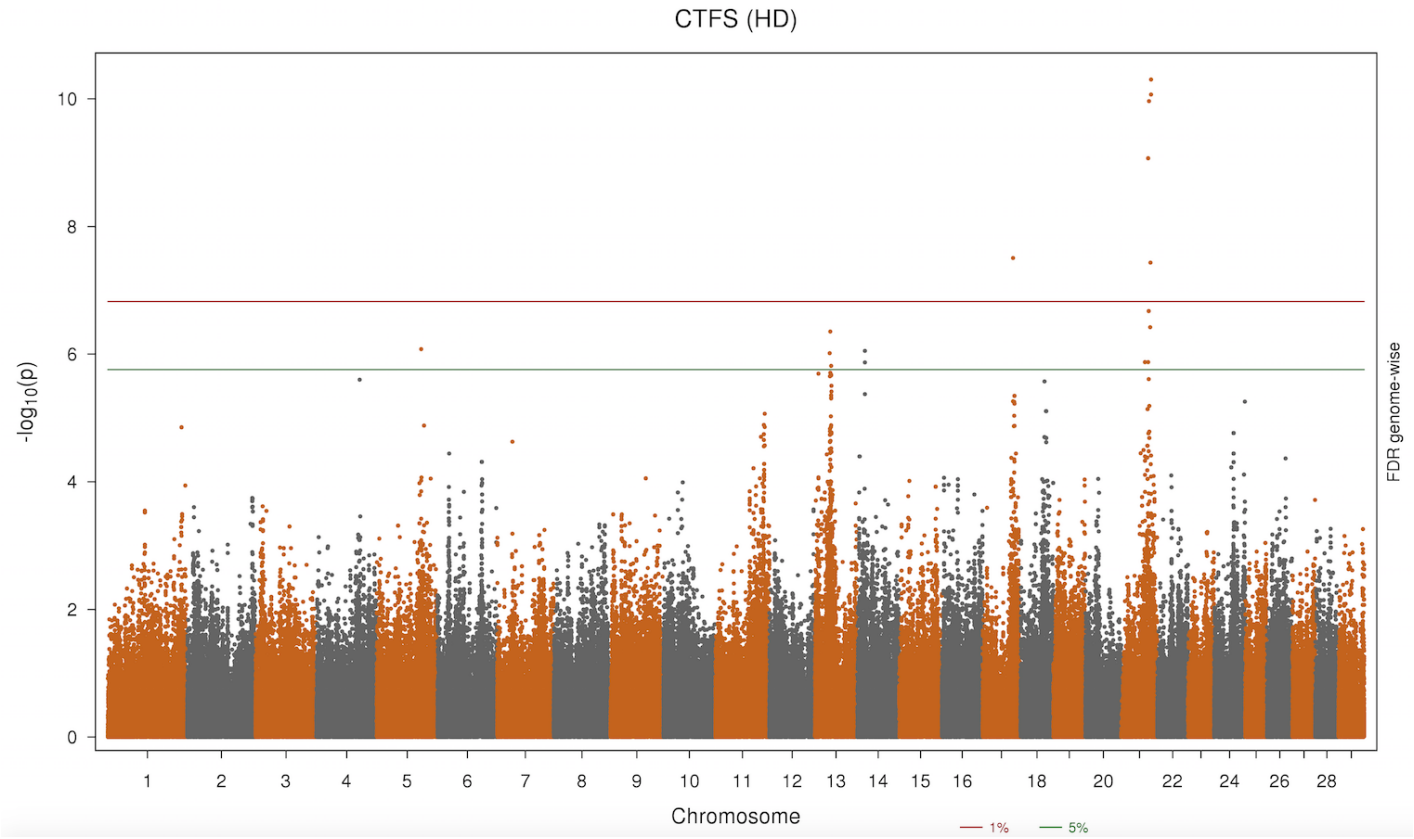


Figure 3. 6: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for calving to first service interval (CTFS). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

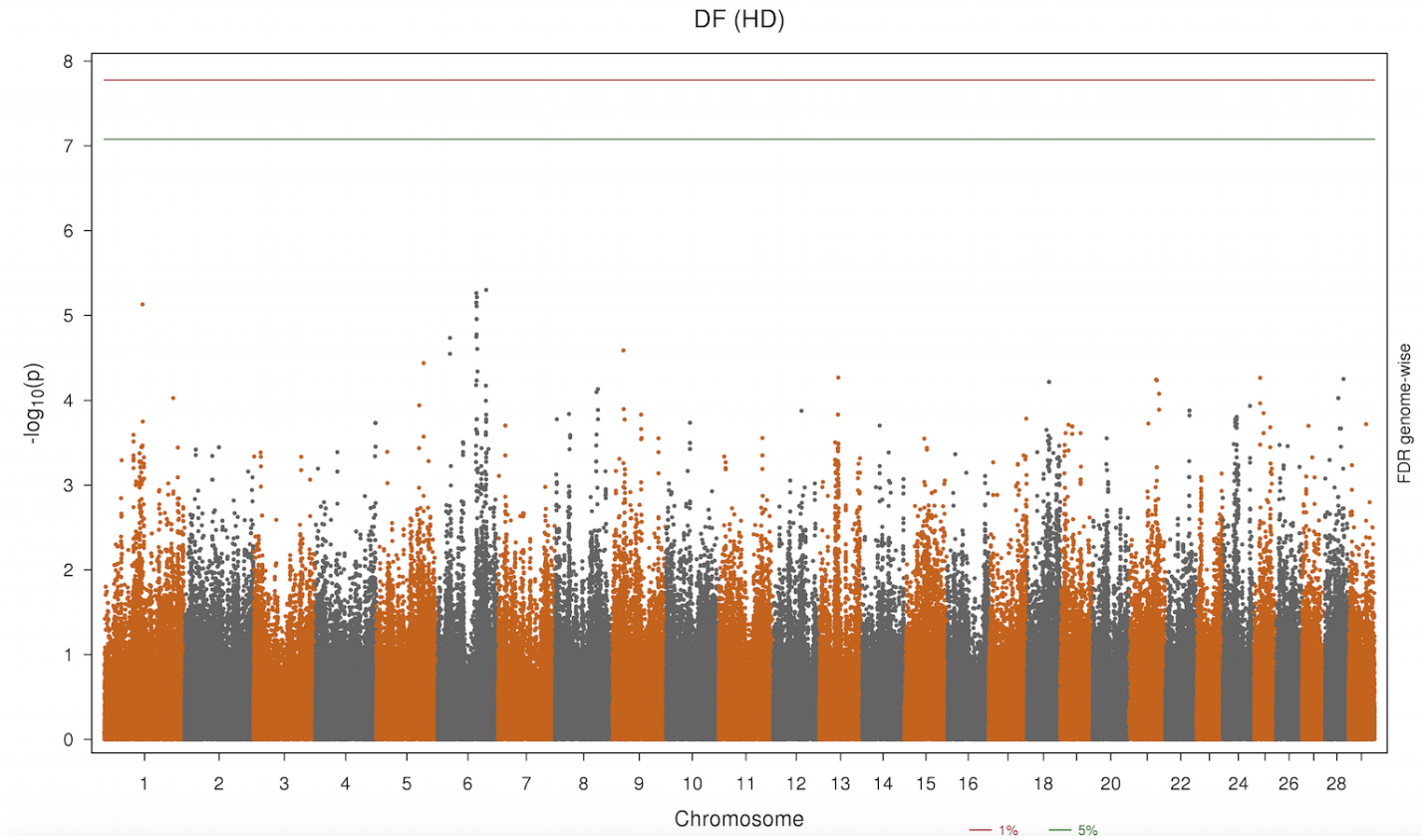


Figure 3. 7: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for daughter fertility (DF). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

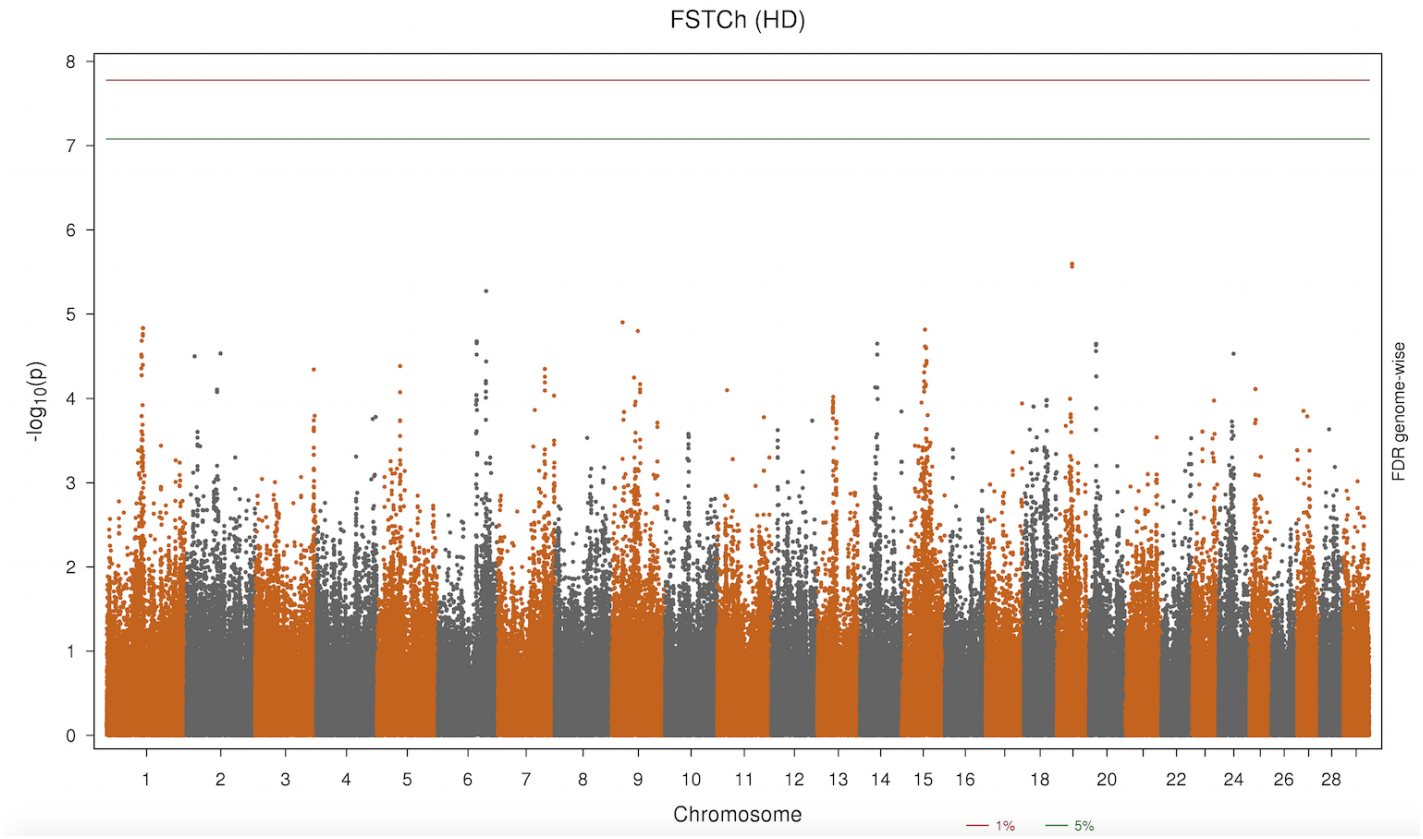


Figure 3. 8: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for heifer first service to calving interval (FSTCh). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

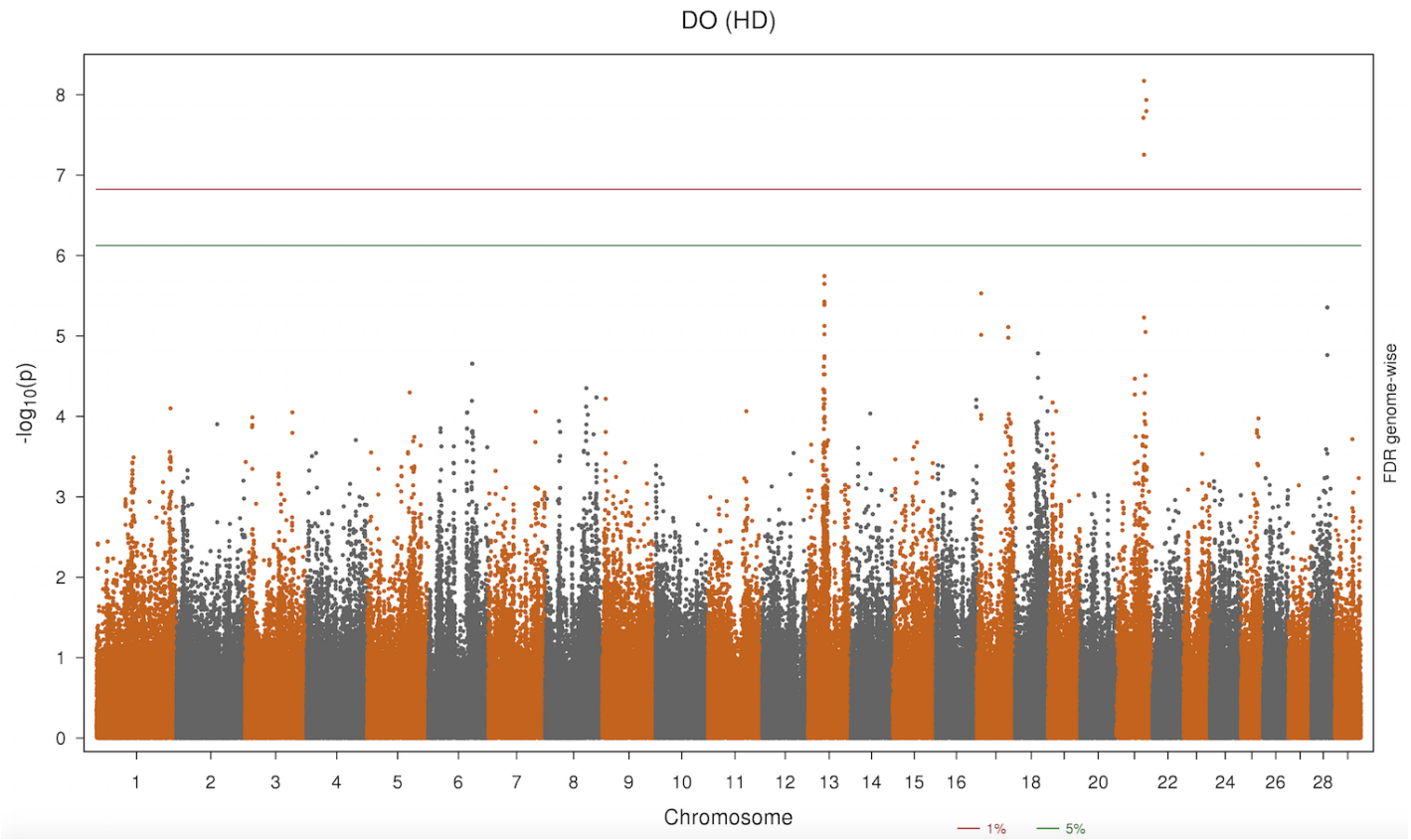


Figure 3. 9: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for days open (DO). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

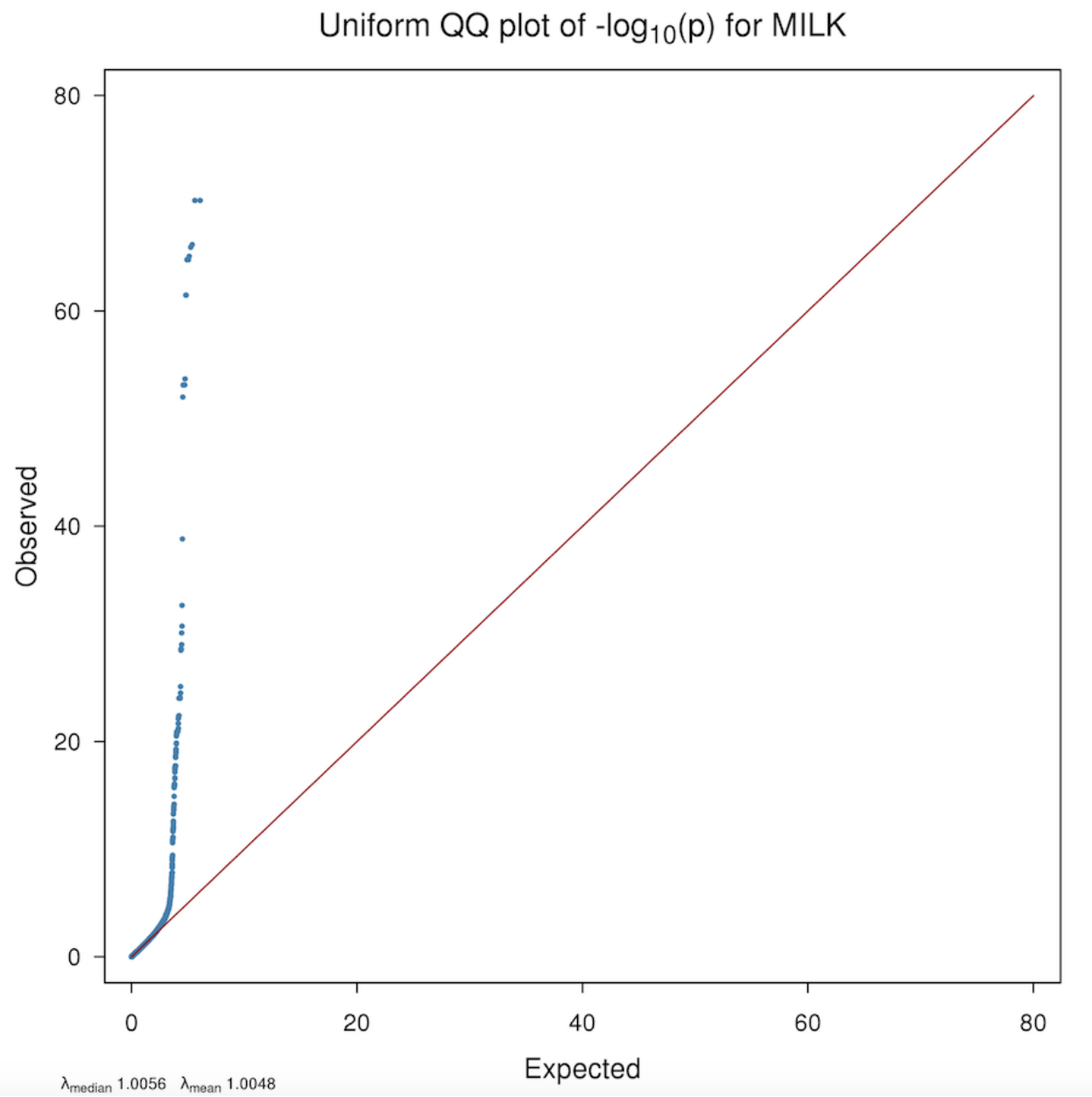


Figure 3. 10: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for milk production (MILK). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

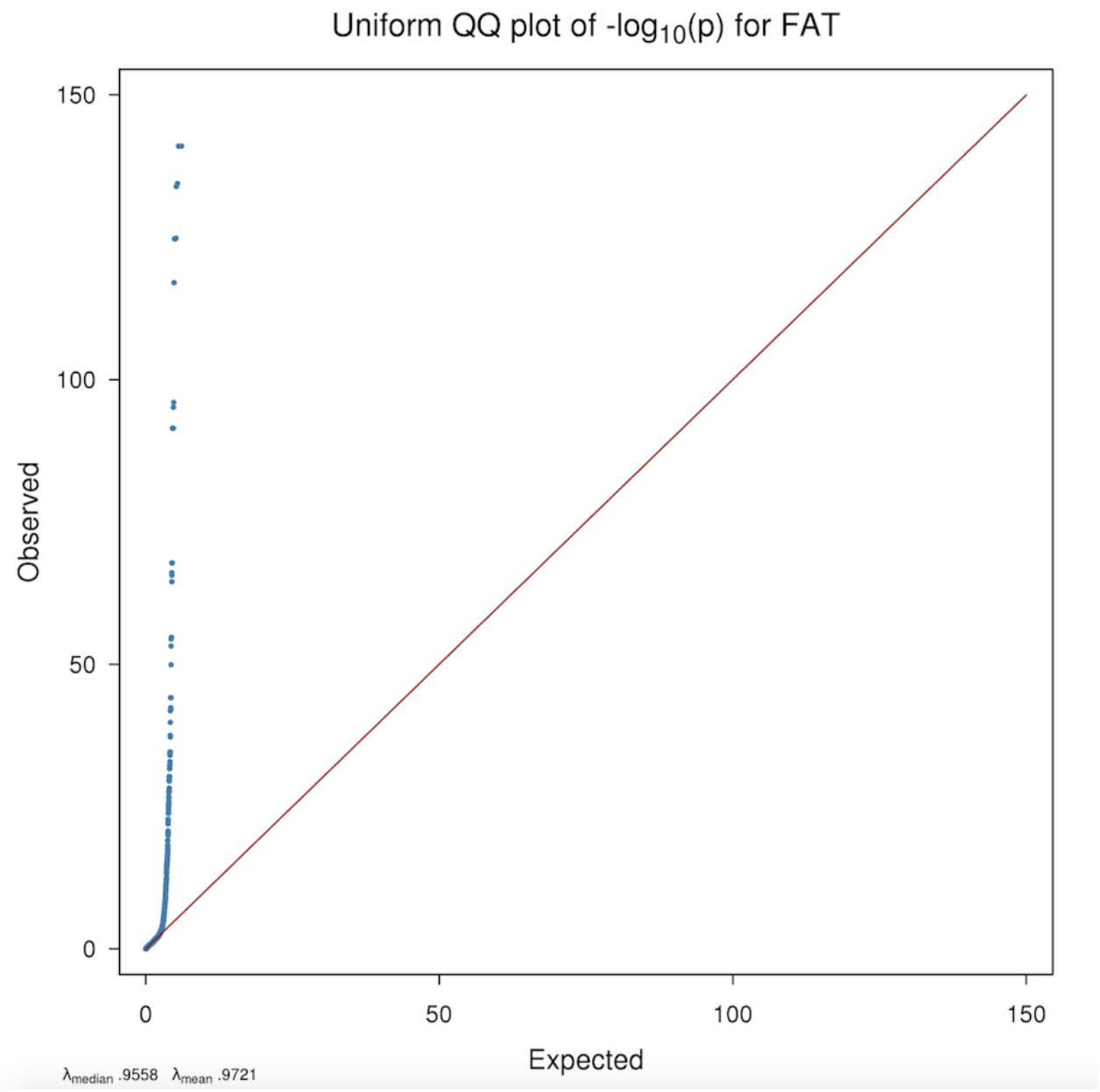


Figure 3. 11: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for fat production (FAT). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

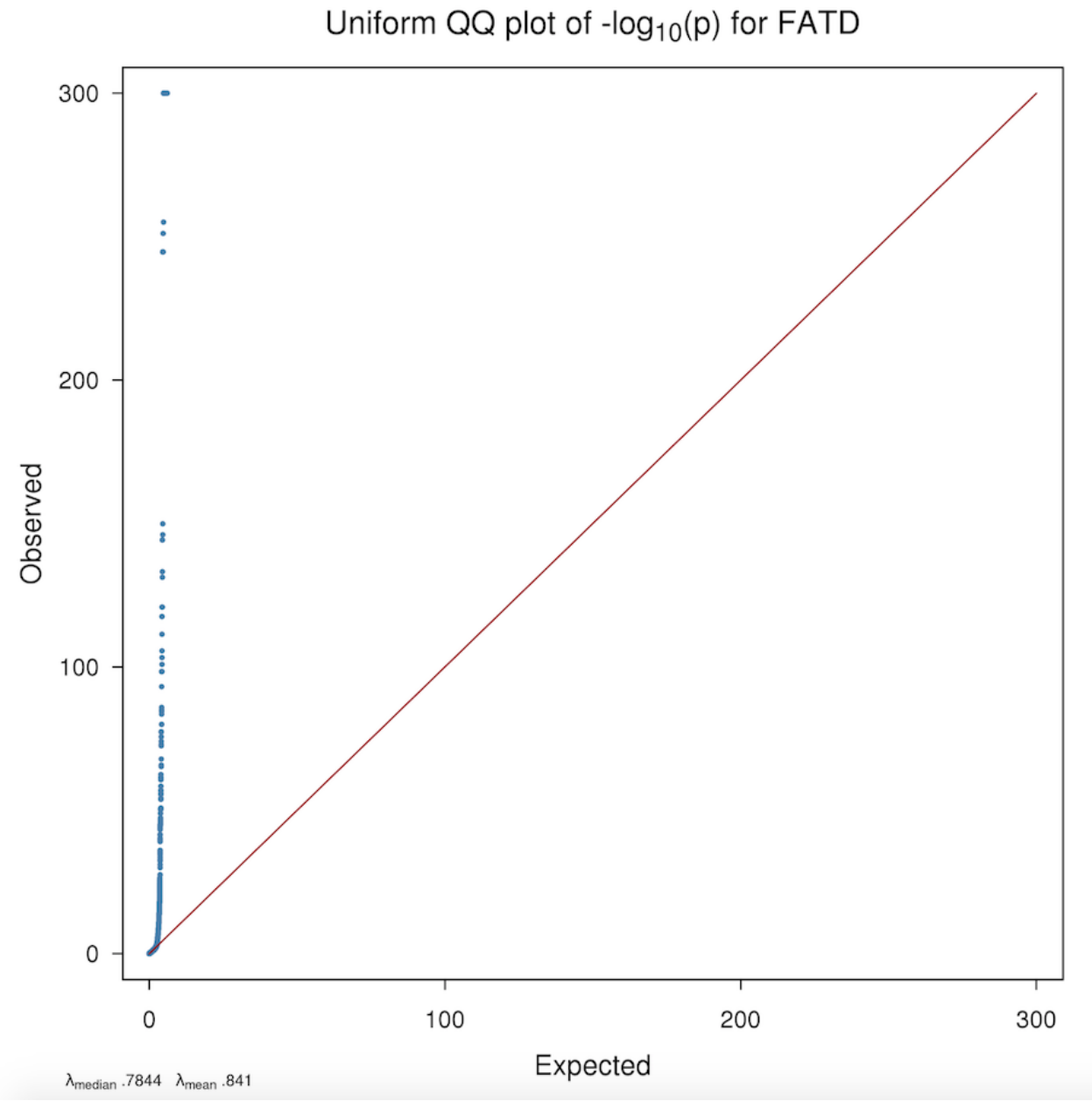


Figure 3. 12: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for fat deviation (FATD). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

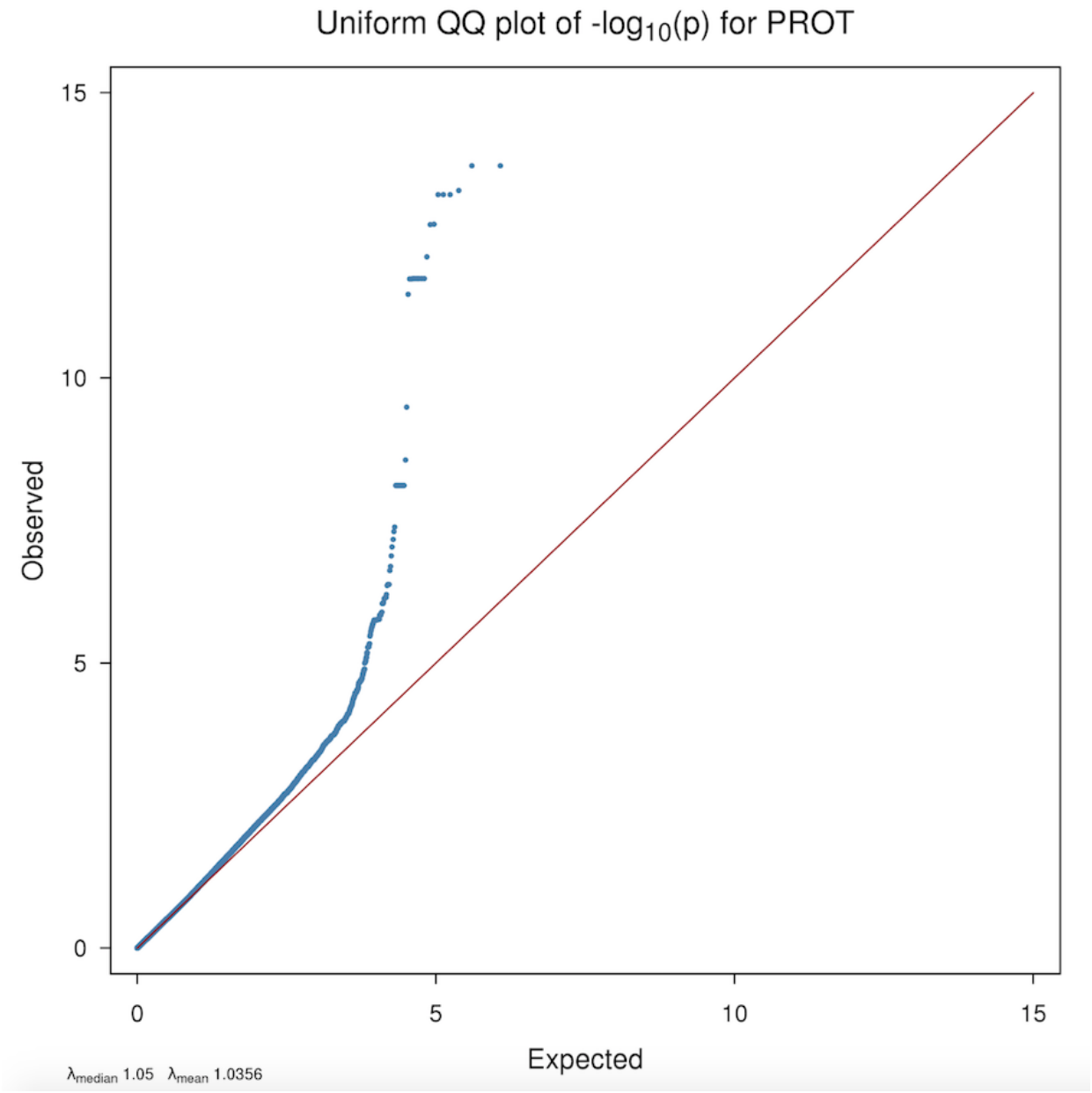


Figure 3. 13: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for protein production (PROT). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

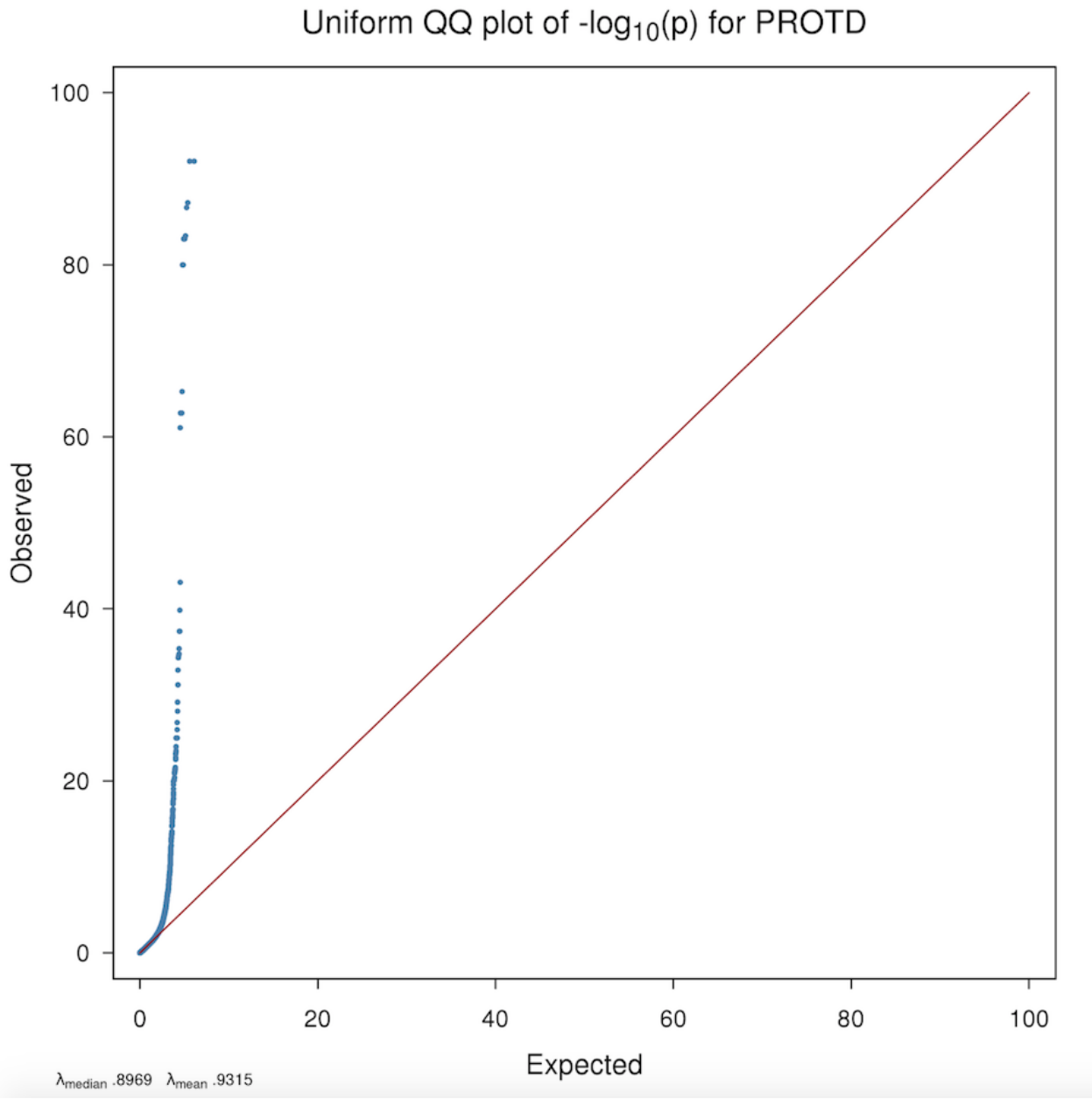


Figure 3. 14: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for protein deviation (PROTD). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

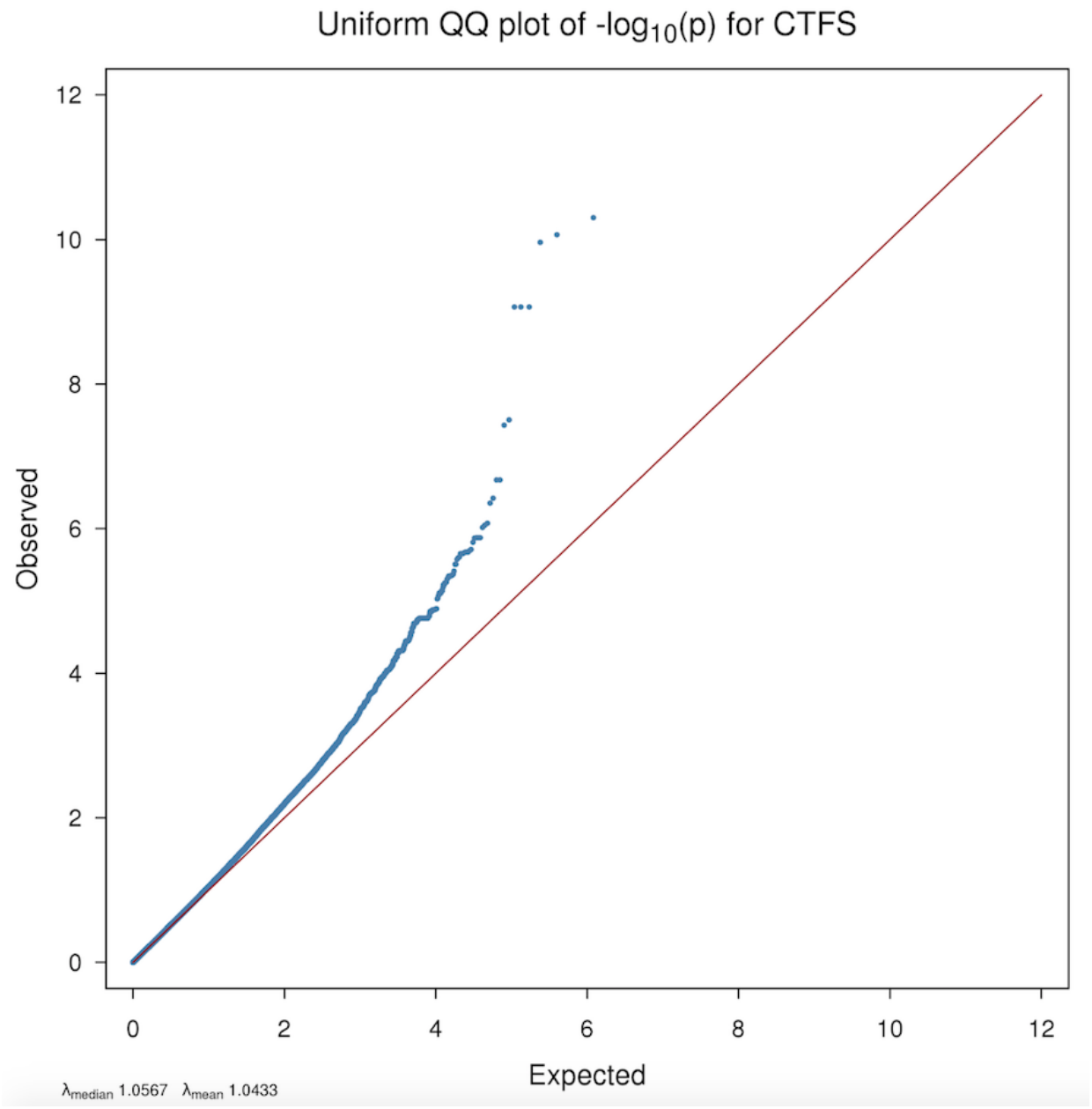


Figure 3. 15: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for calving to first service interval (CTFS). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

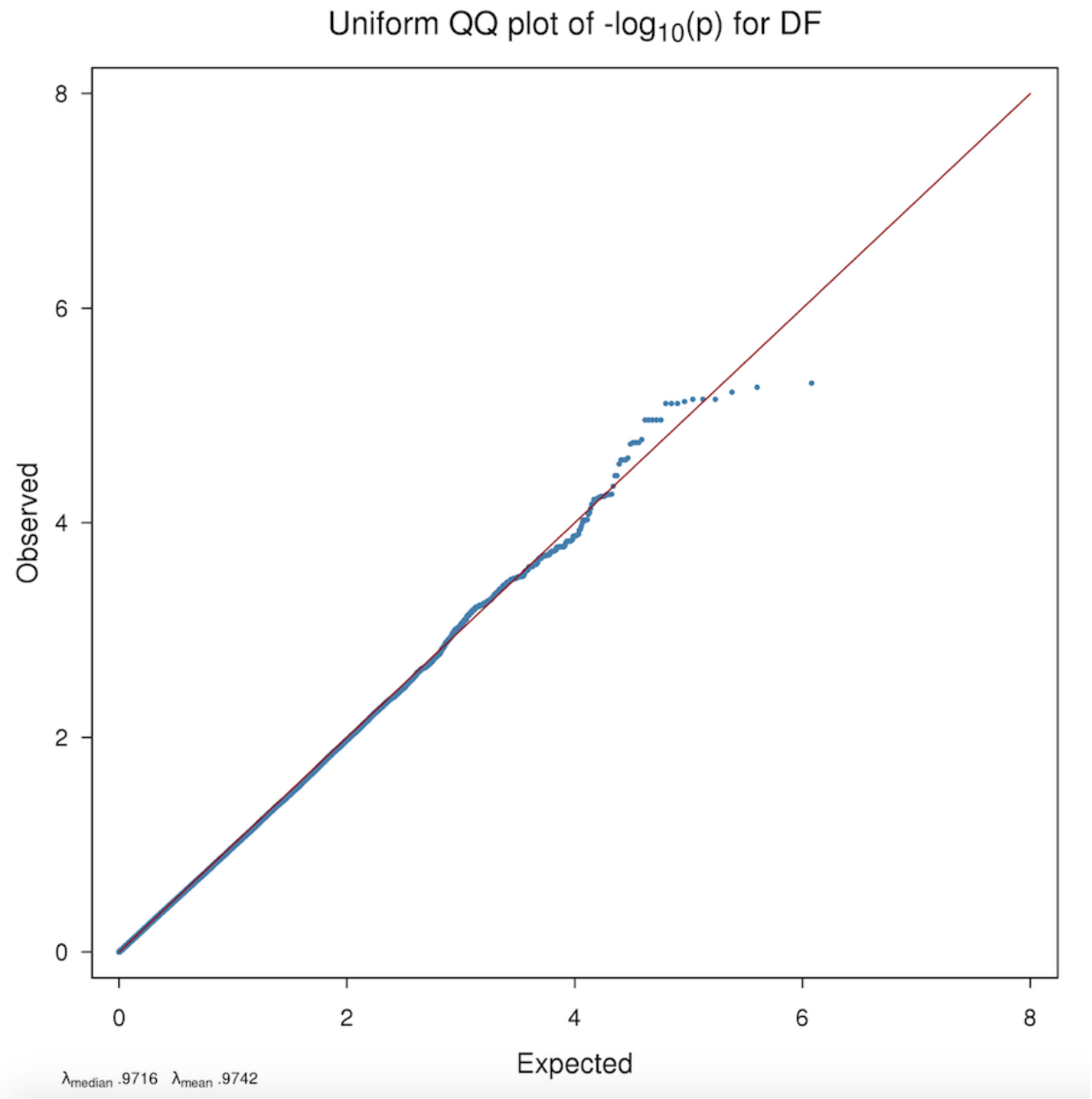


Figure 3. 16: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for daughter fertility (DF). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

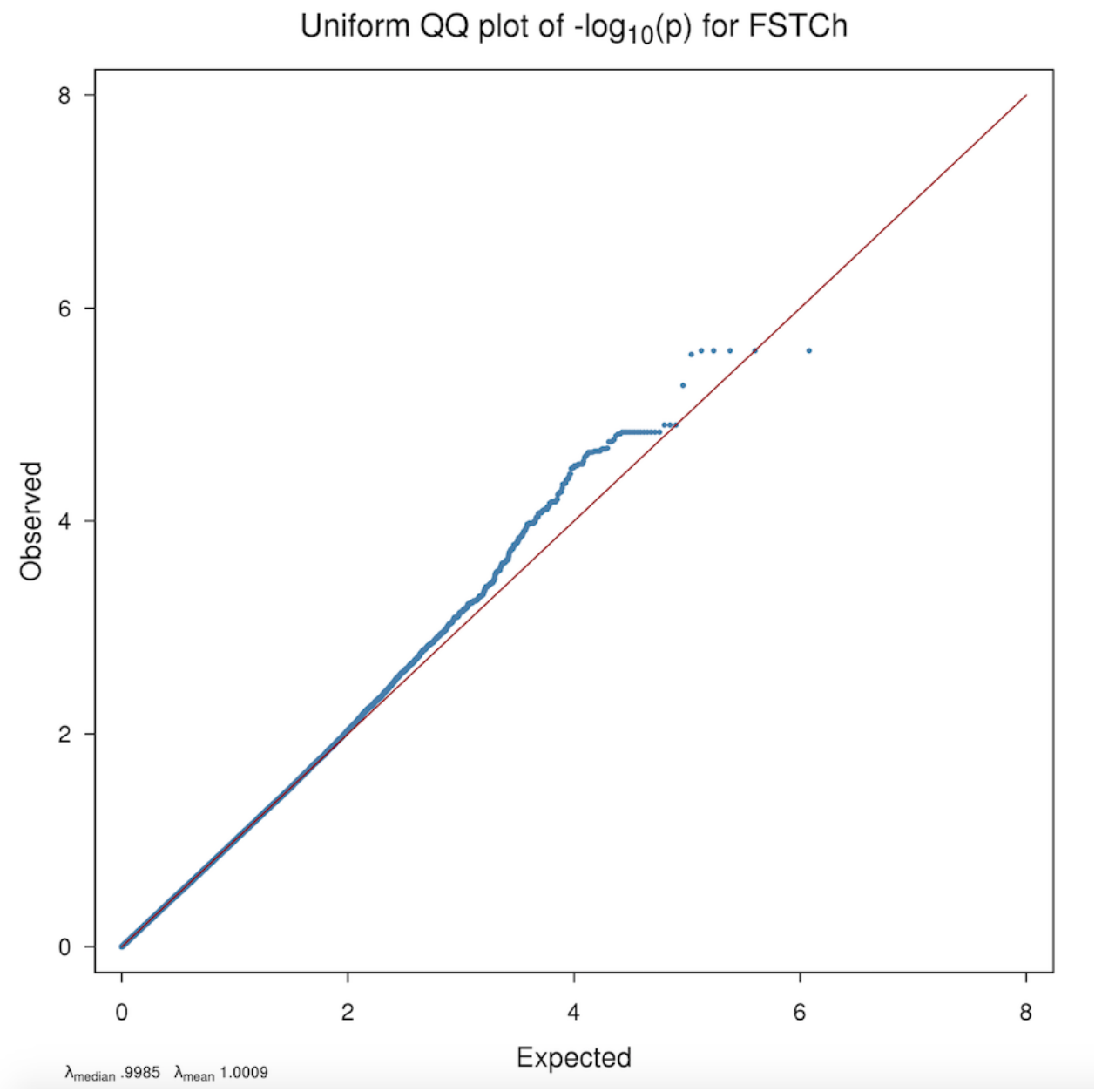


Figure 3. 17: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for heifer first service to calving interval (FSTCh). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

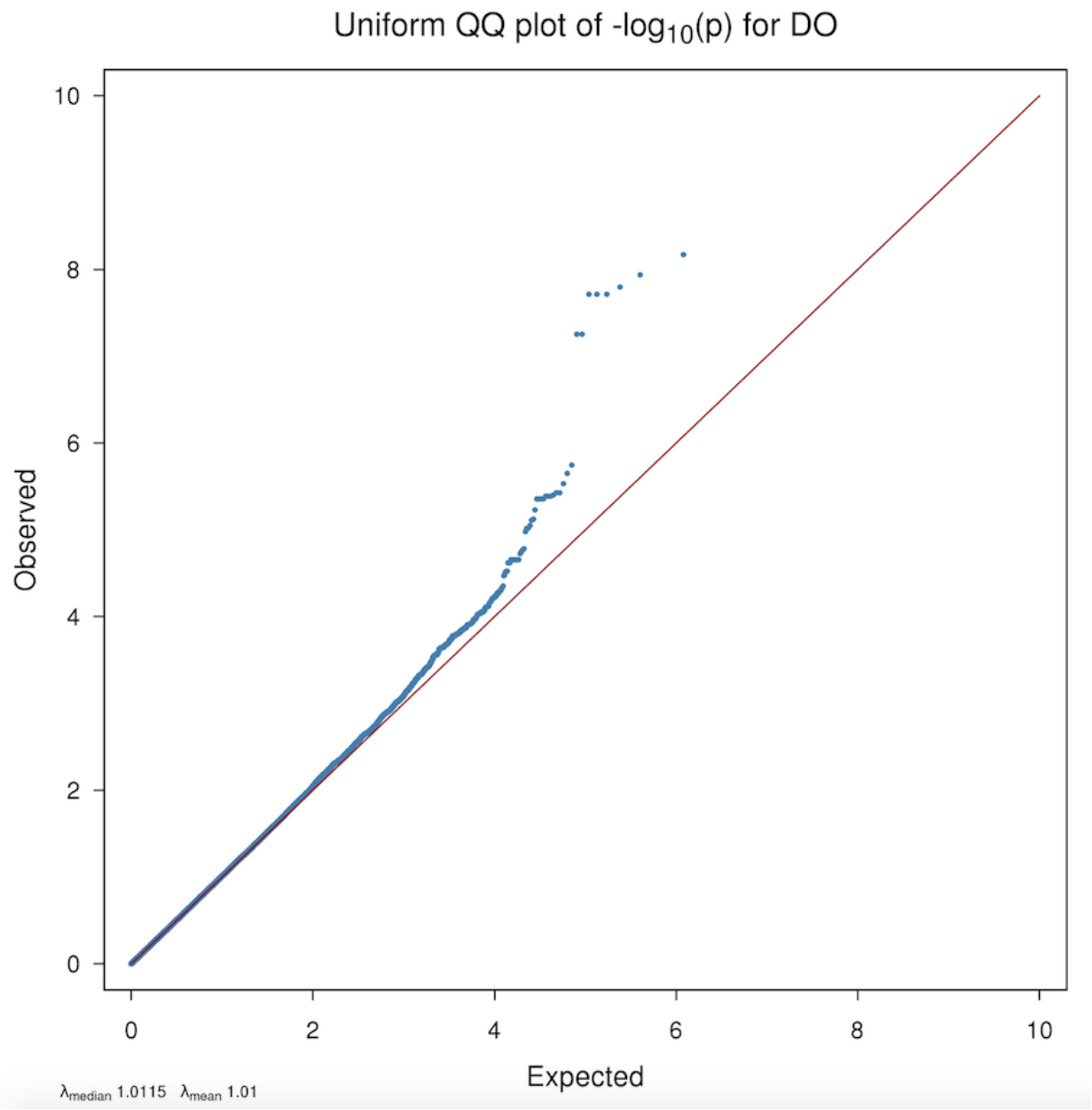


Figure 3. 18: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for days open (DO). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

CHAPTER 4. Genome wide association study for female fertility, longevity and lifetime profitability index traits in North American dairy Holstein cattle.

[4]

4.1. Abstract

Female fertility in Holstein cattle has declined through intense genetic selection for milk production over the last 20 years. One approach to improving fertility is to identify the genomic regions and variants affecting fertility traits and incorporate this knowledge into selection decisions. The objectives of this study were to identify or refine position of the genomic regions affecting lifetime profitability index, female fertility traits (age at first service, cow first service to conception, heifer and cow non-return rate) and longevity (herd life, direct and indirect herd life) in the North American Holstein dairy cattle population. A genome-wide association study (GWAS) was performed for each trait, using a single SNP regression mixed linear model and imputed HD panel (777k) genotypes. Several peak regions were detected for lifetime profitability index, lactation persistency and longevity. The results overlap with previous findings as well as identify some novel regions for lactation persistency. Previously proposed causative and candidate genes supported by this work include *DGAT1*, *GRINA*, *CPSF1* while new candidates are *SLC2A4RG* and *THRB*. Thus the chromosomal regions identified in this study not only confirm several previous findings but also highlight new regions that may contribute to genetic variation in lactation persistency and longevity associated traits in dairy cattle.

[4] A version of this chapter has been published in the *Journal of Dairy Science*. Nayeri et al. 2016. Genome-wide association study for lactation persistency, female fertility, longevity and lifetime profit index traits in Holstein dairy cattle. DOI: http://dx.doi.org/10.3168/jds.2016-11770.

4.1. Introduction

Milk production and female fertility are two important traits that contribute to the profitability of the dairy industry (Boichard, 1990). Increasing milk production in dairy cattle has been a primary focus of genetic selection (Oltenacu and Algers, 2005). This selection has caused a decline in cow fertility because of the negative genetic correlation between fertility and milk production (Kadarmideen et al., 2000; Royal et al., 2002). It has been concluded that a combination of physiology, nutrition, genetic and management strategies should be considered to provide a long-term improvement in fertility of high producing dairy cows (Shook, 2006). Dairy breeding programs stand to improve the overall profitability of the industry through an emphasis on durability, health and fertility of cows alongside the increase in milk production (Kulak et al., 1997).

Genomic regions explaining variation in female fertility traits in cattle have been identified in several genome-wide association studies (GWAS) within a variety of breeds (Höglund et al., 2009; Pryce et al., 2010; Schulman et al., 2011; Sahana et al., 2011; Hawken et al., 2012; Peñagaricano et al., 2012; Minozzi et al., 2013; Höglund et al., 2015). Significant associations have been identified on several chromosomes for age at puberty (Hawken et al., 2012), cow non-return rate (Holmberg and Andersson-Eklund, 2006), pregnancy rate (Ashwell et al., 2004) and calving performance (Holmberg and Andersson-Eklund, 2006). More in-depth analyses have identified several candidate genes affecting fertility traits such as pregnancy-associated plasma protein-A2 (*PAPP2-A2*) on chromosome 16 (associated with calving ease) (Wickramasinghe et al., 2011) and calpastatin (*CAST*) on chromosome 7 (Garcia et al., 2006) (associated with fertility and longevity) in dairy cattle (Minozzi et al., 2013).

The objective of this study was to detect genome regions affecting first service to conception cow (FSTCc), age at first service (AFS), heifer non-return rate (NNRh) and cow non-return rate (NRRc) as well as longevity and productivity related traits such as lactation persistency (LP), herd life (HL), indirect herd life (IHL) and direct herd life (DHL) in the North American Holstein population. Lifetime profitability Index (LPI) was also studied and consists of three main components: production (yield traits and milk components); durability (herd life, mammary system, feet and leg, dairy strength); health and fertility (daughter fertility). The LPI reflects the relative profitability that can be expected during the lifetime of future daughters.

4.2. Materials and methods

4.2.1. Animals and Data

A population of North American Holstein bulls was used in this study of nine fertility and profitability related traits including lifetime profitability index (LPI), lactation persistency (LP), herd life and indirect herd life (HL, IHL respectively), direct herd life (DHL), cow first service to conception (FSTCc), age at first service (AFS), heifer 56-day non-return rate (NNRh) and cow 56-day non-return rate (NRRc). The Canadian Dairy Network (CDN) provided genotypes, available pedigree information and official evaluations for proven bulls born between 1956 and 2009. Herd life was measured as the survival of each cow at five specific points during their productive life including: 1) survival from first calving to 120 days in milk in first lactation, 2) survival from 120 to 240 days in milk in first lactation, 3) survival from 240 days in milk in first lactation to second calving, 4) survival from second calving to third calving, and 5) survival from third calving to fourth calving (www.cdn.ca). Indirect herd life was evaluated based on a combination of conformation traits, reproduction traits and udder health (www.cdn.ca). Age at

first service was considered as the age in days at which a heifer was inseminated for the first time. Individuals were genotyped using the BovineSNP50K (50k) panel (3,729 bulls) or the high density (HD, 777k) SNP panel (2,387 bulls), respectively (Illumina, San Diego, CA). The 50k panel SNP genotypes were subjected to the standard quality control measures that are used by CDN (Wiggans et al., 2009). Quality control was performed on the HD genotyping data using the snp1101 (Sargolzaei, 2014) software. This step excluded 116,619 SNPs including 46,433 SNPs from sex chromosome or misplaced SNPs, 3,566 SNPs with high Mendelian error rate (> 0.05), 6,446 SNPs with low call rate (< 0.9), 61,577 SNPs with low Minor Allele Frequency (MAF < 0.000001) and 90 SNPs with the excess of heterozygosity (> 0.15). The number of SNPs remaining for downstream imputation was 40,666 SNPs for the 50k panel and 657,986 SNPs for the HD panel.

The 3,729 50k genotypes were imputed to the HD panel, using the 2,387 HD panel genotypes as the reference and the FImpute V2.2 software (Sargolzaei et al., 2014). After imputation, an additional quality control step was performed on the imputed data. A total of 55,817 SNPs with MAF of less than 1% and 74 SNPs with a Mendelian error rate of more than 5% (74 SNPs) were excluded. After quality control, 602,095 SNPs remained for use in the subsequent association analysis.

Animal Care and Use Committee approval was not obtained for this study because analyses were performed on existing data obtained under standard farm management from commercial dairy farmers and breeders. All dairy farmers in Canada must follow “The Code of Practice for the Care and Handling of Dairy Cattle” developed by the National Farm Animal Care Council of Canada (http://www.nfacc.ca/).

4.2.2. De-regressed Proofs Calculation

In this study, we used de-regressed genetic evaluations of Holstein bulls as independent variables to test the association with the HD panel. In this genetic evaluation, a bull's published estimated breeding value (EBV) is a weighted mean of his daughters deviations (DD) and his parental average (PA) (VanRaden et al., 2009). The de-regressed bull proofs were computed by CDN as shown below (VanRaden et al., 2009):

$$DE_{prg} = \frac{Rel_{EBV}}{1 - Rel_{EBV}} - \frac{Rel_{PA}}{1 - Rel_{PA}}$$

$$Rel_{DD} = \frac{DE_{prg}}{DE_{prg} + 1}$$

$$DEBV = PA + \frac{(EBV - PA)}{Rel_{DD}}$$

DE_{prg} is the daughter equivalent from progeny information, and Rel_{EBV} and Rel_{PA} are the reliabilities of EBV and PA, respectively. Rel_{DD} is the reliability of DD, and DEBV is the de-regressed bull proof.

4.2.3. Genome Wide Association Analysis (GWAS)

A single SNP regression mixed linear model implemented in the snp1101 software (Sargolzaei et al., 2014; VanRaden, 2008) was used for association analysis:

$$Y = \mu\mathbf{1} + \mathbf{x}\alpha_i + \mathbf{Z}a + e$$

Where y is a vector of de-regressed proofs of the trait of interest; μ is the fixed mean effect and **1** is a vector of 1s; α is the linear regression coefficient on the ith SNP; **x** is a vector of allele count for the ith SNP; a is a vector of the random additive polygenic effects; Z is an incidence matrix associated with additive polygenic effects and e is a vector of random error terms.

The model assumptions are a_i: $a \sim N\,(0, \mathbf{G}\sigma_a^2)$ where **G** is the genomic relationship matrix (VanRaden, 2008) and σ_a^2 is the polygenic additive genetic variance; e_i: $e \sim N\,(0, \mathbf{R}\sigma_e^2)$ where σ_e^2 is the residual variance. **R** is a diagonal matrix containing weights for the residual variance based on the reliabilities of the de-regressed bull proofs (VanRaden, 2008).

In order to account for multiple tests, the false discovery rate (FDR) was controlled at 5% and 1% genome-wise levels to identify significant and highly significant associations, respectively. The quantile-quantile (Q-Q) plots and inflation factor λ (Devlin and Roeder, 1999) were used to compare observed distributions of –log (P-value) to the expected distribution under the no association model for each trait.

4.3. Result and discussion

4.3.1. Statistical Analysis

GWAS identified strong associations for LPI on *Bos taurus* autosome (BTA) 14 and 18, for LP on BTA 20 and for IHL, HL and DHL on BTA 18 (Figure 4.1 to Figure 4.5). However, no significant association was found for AFS, FSTCc, NRRc and NRRh (Figure 4.6 to Figure 4.9). For those traits yielding significant associations, the number of significant SNPs identified (genome-wise FDR of 5% and 1%) is shown in Table 4.1.

The q-q plots showing the observed distribution of –log (P-value) to the expected distribution, under null hypothesis, are shown in Figure 4.10 to Figure 4.14. The inflation factor (λ_{median}) values for lifetime profitability index (LPI), lactation persistency (LP), herd life (HL), indirect herd life (IHL) and direct herd life (DHL) were close to the expected value of 1 are 1.0572, 1.065, 1.1221, 0.9763 and 1.3063, respectively.

4.3.2. Lifetime Profitability Index (LPI)

Selection for more profitable cows began with selection improvements in milk protein percentage in dairy cows (VanRaden, 2004). Later, traits were combined to form selection indexes including yield traits, service sire calving ease, daughter calving ease, daughter pregnancy rate and body conformation traits (VanRaden, 2004). LPI includes most of the traits that impact the profitability of a dairy cow. Genome-wide association analysis for LPI identified 53 SNPs significant at genome-wise FDR 5% (Additional file[5] S4.1). These significant SNPs were localized on BTA14 and 18 (Figure 4.1).

The most significant SNPs for LPI were identified on chromosome 14 at 1.6 to 1.8 Mb. Our result indicate that several SNPs within this region including BovineHD1400000271: rs136792973, BovineHD1400000204: rs137727465 (within the *CYHR1* gene), BovineHD1400000216: rs134432442 (within the *CPSF1* gene), ARS-BFGL-NGS-4939: rs109421300 (within the *DGAT1* gene), BovineHD1400000275: rs133271979 (within the *GRINA* gene) are located within the confidence intervals of previously identified QTL regions (Additional file S4.2). The significant influence of polymorphism (K232A polymorphism) in *DGAT1* gene on bovine milk production traits (milk yield, protein content, fat content and fatty

[5] Additional files are accessible through the link http://dx.doi.org/10.7939/DVN/10941.

acid composition) was previously shown in different studies (Grisart et al., 2002; Schennink et al., 2007, 2008; Conte et al., 2010). In addition, *CYHR1*, *CPSF1*, and *GRINA* genes were reported to be associated with milk fat component in Danish and Chinese Holstein cattle (Buitenhuis et al., 2014; Jiang et al., 2014). These SNPs also overlap with regions associated with clinical mastitis in Holstein dairy cattle (Additional file S4.2) (Sahana et al., 2013). We also identified a highly significant SNP (genome-wide FDR 1%) in this study, Hapmap52798-ss46526455: rs41256919, that overlaps with the *MAF1* gene on BTA14. The *MAF1* gene was shown to be associated with increased mammary gland milk protein synthetic capacity in lactating dairy cows (Sciascia et al., 2013).

Another highly significant SNP (BovineHD1400000188: rs134892687) on BTA14 was located within the Rho GTPase activating protein 39 (*ARHGAP39*) gene. This SNP and the assigned gene (*ARHGAP39*) along with other SNPs and genes (*CYHR1*, *CPSF1*, *DGAT1*, *GRINA*, *SMPD5*) were reported to be significantly associated with milk fat and fat component traits in Danish Holstein cattle (Buitenhuis et al., 2014). Additionally, in a GWAS study in Chinese Holstein cattle, several significant SNPs, including the ones identified significant in this study for LPI (ARS-BFGL-NGS-4939; ARS-BFGL-NGS-107379 and ARS-BFGL-NGS-57820) were reported to be associated with somatic cell score EBVs (Wang et al., 2015). The authors speculated that the linked gene to these SNPs (*ARHGAP39*) is a candidate gene for mastitis susceptibility in Holsteins (Wang et al., 2015). Mastitis is the most costly disease associated with loss of milk production and increase in culling rates in dairy cattle (Nemcova et al., 2007). The overlap in results between this previous mastitis work and our LPI results could be due to the influence of these variants on mastitis and consequently LPI.

4.3.3. Lactation Persistency (LP)

The GWAS analysis for lactation persistency identified several genomic regions with highly significant SNPs on BTA6, 13, 20 and 27 (Figure 4.2). The total number of significant SNPs was 83 after correcting for multiple testing at a genome-wise FDR of 5%, with the highest number of significant SNPs on BTA20 (Additional file S4.1).

Lactation persistency as a trait has been analyzed in several previous studies (Boichard et al., 2003; Harder et al., 2006; Kolbehdari et al., 2009). Some of the regions identified in our analysis overlap with the QTL regions that were reported previously. The most significant SNPs on BTA6 in our GWAS were located at 88 Mb. This region overlaps with a QTL region on BTA6 (between markers DIK082 at 48 Mb/57.55 cM and ILSTS097 at 64 Mb/72.43 cM) associated with persistency of milk yield (PMY), protein yield (PPY) and persistency of milk energy yield (PEY) in German Holstein cattle (Harder et al., 2006).

The region on BTA6 at 88Mb in the present study was previously reported to be associated with somatic cell score in Holstein cattle (Abdel-Shafy et al., 2014). Interestingly, we also assigned several significant SNPs including BovineHD0600024162: rs134391498 on this chromosome within the intronic region of the gene solute carrier family 4 (sodium bicarbonate cotransporter), member 4 (*SLCA4A4*). This gene has been reported as a candidate gene associated with mastitis susceptibility in Danish Holstein cattle (Wu et al., 2015). Cole et al. (2009) and Appuhamy et al. (2009) reported negative genetic correlations between clinical mastitis (after 100 days in milk) and persistency of yield traits (milk, fat and protein yield) in different cattle breeds. They indicated that an increase in persistency of milk, fat and protein yield traits is associated with decreases in persistency of somatic cell score (PSCS); somatic cell score is an indicator of

clinical and subclinical mastitis in dairy cattle (Appuhamy et al., 2009; Cole and Null, 2009). This result suggests that identified SNPs at 88Mb on BTA6 might affect udder disease susceptibility and subsequently lactation persistency in the dairy cattle.

Another major signal for lactation persistency in this study was identified on BTA13 at 54 to 55 Mb (Additional file S4.1). This region was not reported to be associated with lactation persistency in dairy cattle in previous investigations. The SNPs in this region (Additional file S4.1) are located within the genes *MYT1*, *SLC2A4RG*, *SLC17A9*, *LAMA5*, *ADRM1*, *OSBPL2* and *SS18L1*. Boinaz and Loor (2008) indicated that expression of the gene *OSBPL2* in the mammary gland increases along with *SREBF1* and *SREBF*2 genes during lactation. They suggested that this gene might be involved in the regulation of *SREBF1* and coordination of sphingolipid and cholesterol synthesis within the mammary gland in dairy cattle (Bionaz and Loor, 2008).

Kolbehdari et al. (2008) reported a significant SNP on BTA27 at 12 Mb affecting persistency of milk yield. Littlejohn et al. (2014) also reported significant variants on chromosome 27, affecting expression of *AGPAT6* gene (at 36 Mb), that are causally involved in milk fat synthesis and has pleiotropic effects on other milk components. The associated SNPs to the *AGPAT6* gene in Littlejohn et al. (2014) study were not significant in our results. However, we detected five significant SNPs on BTA27 at 41 Mb for lactation persistency (Additional file S4.1) overlapping with significant SNPs found on BTA 27 associated with milk fat percentage by Littlejohn et al. (2014). This region has not been reported to be associated with lactation persistency before, and we speculate this region to be a new region for this trait (Additional file S4.1). One significant SNP in this region (BovineHD2700012022: rs134277881) is located within the gene thyroid hormone receptor, beta (*THRB*). This gene encodes a nuclear hormone receptor for triiodothyronine and mutations in this gene are known to be the cause of generalized thyroid

hormone resistance (GTHR). Thyroid hormones are known for their importance in milk production through stimulation of metabolic rates with other hormones (including endocrine pancreas and adrenal gland cortex) (Blum et al., 1983). It has also been shown that there is a positive correlation between thyroid hormones in blood and energy metabolism (Reist et al., 2002). This proposed gene might be considered as a new potential candidate gene affecting production traits in the dairy cattle.

Significant SNPs on BTA20 for LP in this study were span from 31 to 32 Mb. Several of these SNPs including BovineHD4100014643: rs41639260, ARS-BFGL-NGS-118998: rs110482506 and UA-IFASA-7069: rs41639261 are intronic variants residing within the growth hormone receptor (*GHR*) gene. Association of significant SNPs close to *GHR* gene affecting milk fat content was reported in a study carried out on German Holstein cattle (Wang et al., 2012). This region on BTA20 was also in agreement with QTL regions previously identified to affect lean meat, carcass yield (Doran et al., 2014) and sire conception rate (Li et al., 2012) in Holstein cattle in general (Additional file S4.2).

The current study also revealed one significant SNP (5% FDR) on chromosome 17 at 72 Mb. Boichard et al. (2003) also reported a significant QTL associated with milk persistency (close to marker CSSM033) on chromosome 17. However, the position of this QTL was at 56 Mb (Boichard et al., 2003), which is different from the result found in the present study.

4.3.4. Longevity Related Traits

Longevity is one of the highly desirable traits in the dairy cattle industry that affects overall profitability (Sewalem et al., 2008). This trait is determined by voluntary and involuntary culling decisions based on production, health, fertility and other functional traits (VanRaden and

Wiggans, 1995; Sewalem et al., 2008). Genetic evaluation of longevity for proof bulls is based on herd-life evaluations that reflect the additional number of lactations which daughters of these bulls are expected to have due to reduced culling for non-production reasons (www.cdn.ca). Genetic evaluations and analysis of herd life for bulls are based on two evaluations; direct herd life (DHL) and indirect herd life (IHL) (www.cdn.ca). One of the components of direct herd life is daughter survival in the herd, at five specific points during their productive life. These five measurements of the daughter survival are inter-correlated and are used to compute an overall bull proof for direct herd life. Indirect herd life (IHL) value is calculated based on a function of proof for other non-production traits including conformation traits, reproduction traits and udder health traits (www.cdn.ca).

Herd Life (HL), Indirect Herd Life (IHL)

GWAS identified several significant regions on BTA5, 6, 7, 14, 18, 20 and 21 for HL with the highest number of significant SNPs on BTA6 and 18 (Figure 4.3; Additional file S4.1). Likewise, genome wide association analysis detected significant SNPs (genome-wise FDR 5%) on BTA5, 6 and 18 associated with IHL (Figure 4.4; Additional file S4.1).

The most significant SNPs for HL and IHL were located on BTA18 at 42 to 65 Mb (Additional file S4.1). The location of several highly significant SNPs on BTA18 in this study (BovineHD1800016754: rs135253383, BovineHD1800016606: rs137554975 and BovineHD1800016612: rs136113894) is in agreement with QTLs affecting direct calving (fertility associated trait) on BTA18 at 57.12 Mb, for both HL and IHL (Sahana et al., 2011; Höglund et al., 2012). Additionally, our study identified a significant intronic SNP (BovineHD1800016754: rs135253383) in the cytosolic thiouridylase subunit 1 (*CTU1*) gene. It

has been reported that a missense mutation in the *CTU1* gene is the most likely candidate variant for direct calving difficulty in the Holstein-Friesian cattle population (Purfield et al., 2015). Calving ease is a fertility-associated trait. Fertility and postpartum performance of replacement heifers are important traits to minimize the non-productive period of an animal's life (Akanno et al., 2015).

In another genome scan study carried out on economically important traits in Holstein cattle by Daetwyler et al. (2008) using a linkage disequilibrium single locus regression model, several QTL were identified on BTA6 (at 84 cM/82 Mb) and BTA13 (at 63 cM/50 Mb) associated with herd life. The SNPs identified in our study on BTA6 (for both HL and IHL), are located at 88 Mb. Furthermore, the significant SNPs on BTA13 in our analysis are mostly localized at 30 Mb. We also identified several SNPs on BTA6 at 88 Mb within the confidence intervals of QTLs associated with somatic cell score (Abdel-Shafy et al., 2014), milk yield and body condition score (Veerkamp et al., 2012) in the Holstein cattle breed (Additional file S4.2). One other highly significant SNP on this chromosome for both HL and IHL (BovineHD0600024403: rs110432804) was assigned to the intron of the gene neuropeptide FF receptor 2 (*NPFFR2*). This gene has been reported as a candidate gene for udder health and susceptibility to mastitis in Danish Holstein cattle (Wu et al., 2015).

Most of the significant SNPs that we found for these two traits were located within the QTL regions that have been previously reported to affect calving traits (calving ease, calf size, stillbirth), clinical mastitis, milk yield and conformation traits in Holstein and other cattle breeds (Additional file S4.2). According to the fact that indirect herd life evaluations are based on a combination of fertility, conformation and udder health traits, it is very possible that these regions and associated candidate genes, particularly on BTA6 and 18 influence this trait.

Direct Herd Life (DHL)

Genome wide association analysis for direct herd life (DHL) detected significant peaks on BTA6, 7, 14, 18, 20 and 21(Figure 4.5). The most significant SNPs (genome-wise FDR 1%) are shown in Additional file S4.1. One significant peak on BTA18 at 52 to 60 Mb, overlaps among DHL, HL and IHL in this study. This region on BTA18 with the highest number of significant SNPs (258 SNPs) for DHL was close to significant QTLs found by Cole et al. (2009) at 57 Mb and Kolbehdari et al. (2009) at 53 Mb associated with calf growth rate, conformation, calving ease and fertility traits in Canadian Holsteins. We further investigated the position of these significant SNPs relative to previously known QTLs. The result indicated several SNPs were located within confidence interval of QTLs affecting somatic cell score (Brand et al., 2009), disease susceptibility (M. paratuberculosis susceptibility) (Pant et al., 2011) and calving-associated traits (Schulman et al., 2011; Höglund et al., 2012) on BTA18 in Holstein cattle (Additional file S4.2).

The position of a significant SNP on BTA13 (BovineHD1300016998: rs110323780) at 59 Mb was in agreement with significant SNPs within a QTL identified in a previous study, associated with HL (Daetwyler et al., 2008). This location also overlaps with the position of a QTL found to be associated with calving ease, calf size and stillbirth in Holstein dairy cattle (Höglund et al., 2012). Additionally, the location of the highly significant SNPs on BTA21 in our study is consistent with a QTL associated with somatic cell score in Norwegian Red cattle (Sodeland et al., 2011) (Additional file S4.2).

Somatic cell score is an indicator of udder health and is highly correlated with clinical mastitis (Brand et al., 2009). Short et al. (1992) reported that udder traits have large absolute genetic

correlations with herd life traits (Short and Lawlor, 1992). Furthermore, increased susceptibility to disease is one of the main factors of reduced productivity, premature culling and mortality (Pant et al., 2011). Taken together with these previous results, our findings indicate that these regions and significant SNPs may potentially have an effect on longevity in dairy cattle and may be used for further genomic prediction analysis to identify causal candidate variants that affect herd life traits.

4.4. Conclusion

The present GWAS study identified SNPs associated with lactation persistency, longevity and lifetime profitability index in Holstein dairy cattle. However, no associations could be identified for female fertility traits. The strongest associations were detected on BTA14 and 18 for lifetime profitability index and BTA6 and 20 for lactation persistency. In addition, two novel regions associated with lactation persistency were identified on BTA13 and BTA27. These regions contain several candidate genes including *MYT1*, *SLC2A4RG* and *SLC17A9* on BTA13 and *THRB* on BTA27. The most significant SNPs associated with longevity traits were found on BTA18 and are mostly located within known QTL regions affecting mastitis and calving-associated traits. In summary, these results can be used for identifying candidate genes and casual mutations that might help to increase the accuracy of genomic selection and provide mechanistic insight into traits.

4.5. Legend

Additional files are accessible through the link http://dx.doi.org/10.7939/DVN/10941.

4.6. References

Abdel-Shafy H, Bortfeldt RH, Reissmann M, Brockmann GA (2014) Short communication: Validation of somatic cell score-associated loci identified in a genome-wide association study in German Holstein cattle. J Dairy Sci 97:2481–6. doi: 10.3168/jds.2013-7149

Akanno EC, Plastow G, Fitzsimmons C, et al (2015) Genome-wide association for heifer reproduction and calf performance traits in beef cattle. Genome 557:1–9. doi: 10.1139/gen-2015-0031

Appuhamy JADRN, Cassell BG, Cole JB (2009) Phenotypic and genetic relationships of common health disorders with milk and fat yield persistencies from producer-recorded health data and test-day yields. J Dairy Sci 92:1785–1795. doi: 10.3168/jds.2008-1591

Ashwell MS, Heyen DW, Sonstegard TS, et al (2004) Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. J Dairy Sci 87:468–75. doi: 10.3168/jds.S0022-0302(04)73186-0

Bionaz M, Loor JJ (2008) Gene networks driving bovine milk fat synthesis during the lactation cycle. 21:1–21. doi: 10.1186/1471-2164-9-366

Blum JW, Kunz P, Leuenberger H, et al (1983) Thyroid hormones, blood plasma metabolites and haematological parameters in relationship to milk yield in dairy cows. Anim Prod 36:93–104. doi: 10.1017/S0003356100039982

Boichard D (1990) Estimation of the economic value of conception rate in dairy cattle. Livest Prod Sci 24:187–204. doi: 10.1016/0301-6226(90)90001-M

Boichard DB, Grohs C, Bourgeois F, et al (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. Genet Sel Evol 35:77–101. doi: 10.1051/gse

Brand B, Baes C, Mayer M, et al (2009) Identification of a two-marker-haplotype on Bos taurus autosome 18 associated with somatic cell score in German Holstein cattle. BMC Genet 10:50. doi: 10.1186/1471-2156-10-50

Buitenhuis B, Janss LL, Poulsen NA, et al (2014) Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15:1112. doi: 10.1186/1471-2164-15-1112

Cole JB, Null DJ (2009) Genetic evaluation of lactation persistency for five breeds of dairy cattle. J Dairy Sci 92:2248–58. doi: 10.3168/jds.2008-1825

Conte G, Mele M, Chessa S, et al (2010) Diacylglycerol acyltransferase 1, stearoyl-CoA desaturase 1, and sterol regulatory element binding protein 1 gene polymorphisms and milk fatty acid composition in Italian Brown cattle. J Dairy Sci 93:753–63. doi: 10.3168/jds.2009-2581

Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JAB (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. J Dairy Sci 91:3225–36. doi: 10.3168/jds.2007-0333

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004. doi: 10.1111/j.0006-341X.1999.00997.x

Doran AG, Berry DP, Creevey CJ (2014) Whole genome association study identifies regions of the bovine genome and biological pathways involved in carcass trait performance in Holstein-Friesian cattle. BMC Genomics 15:837. doi: 10.1186/1471-2164-15-837

Garcia MD, Michal JJ, Gaskins CT, et al (2006) Significant association of the calpastatin gene with fertility and longevity in dairy cattle. Anim Genet 37:304–305. doi: 10.1111/j.1365-2052.2006.01443.x

Grisart B, Coppieters W, Farnir F, et al (2002) Positional candidate cloning of a QTL in dairy cattle: identi cation of a missense mutation in the bovine DGAT gene with major effect on milk yield and composition. Genome Res 12:222–231. doi: 10.1101/gr.224202

Harder B, Bennewitz J, Reinsch N, et al (2006) Mapping of quantitative trait loci for lactation persistency traits in German Holstein dairy cattle. J Anim Breed Genet 123:89–96. doi: 10.1111/j.1439-0388.2006.00577.x

Hawken RJ, Zhang YD, Fortes MRS, et al (2012) Genome - wide association studies of female reproduction in tropically adapted beef cattle. J Anim Sci 90:1398–1410. doi: 10.2527/jas2011-4410

Höglund JK, Buitenhuis B, Guldbrandtsen B, et al (2015) Genome-wide association study for female fertility in Nordic Red cattle. BMC Genet 16:110. doi: 10.1186/s12863-015-0269-x

Höglund JK, Guldbrandtsen B, Lund MS, Sahana G (2012) Analyzes of genome-wide association follow-up study for calving traits in dairy cattle. BMC Genet 13:71. doi: 10.1186/1471-2156-13-71

Höglund JK, Guldbrandtsen B, Su G, et al (2009) Genome scan detects quantitative trait loci affecting female fertility traits in Danish and Swedish Holstein cattle. J Dairy Sci 92:2136–43. doi: 10.3168/jds.2008-1104

Holmberg M, Andersson-Eklund L (2006) Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. J Dairy Sci 89:3664–71. doi: 10.3168/jds.S0022-0302(06)72406-7

Jiang L, Liu X, Yang J, et al (2014) Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. BMC Genomics 15:1105. doi: 10.1186/1471-2164-15-1105

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

Kolbehdari D, Wang Z, Grant JR, et al (2009) A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. J Anim Breed Genet 126:216–27. doi: 10.1111/j.1439-0388.2008.00793.x

Kulak KK, Dekkers JCM, McAllister AJ, Lee AJ (1997) Relationships of early performance traits to lifetime profitability in Holstein cows. Can J Anim Sci 77:617–624. doi: 10.4141/A96-129

Li G, Peñagaricano F, Weigel KA, et al (2012) Comparative genomics between fly, mouse, and cattle identifies genes associated with sire conception rate. J Dairy Sci 95:6122–9. doi: 10.3168/jds.2012-5591

Minozzi G, Nicolazzi EL, Stella A, et al (2013) Genome wide analysis of fertility and production traits in Italian Holstein cattle. PLoS One 8:e80219. doi: 10.1371/journal.pone.0080219

Nemcova E, Stipkova M, Zavadilova L, et al (2007) The relationship between somatic cell count, milk production and six linearly scored type traits in Holstein cows. Czech J Anim Sci 52:437–446.

Oltenacu PA, Algers B (2005) Selection for Increased Production and the Welfare of Dairy Cows : Are New Breeding Goals Needed ? Ambio A J Hum Environ 34:311–315.

Pant SD, Verschoor CP, Schenkel FS, et al (2011) Bovine PGLYRP1 polymorphisms and their association with resistance to Mycobacterium avium ssp. paratuberculosis. Anim Genet 42:354–360. doi: 10.1111/j.1365-2052.2010.02153.x

Peñagaricano F, Weigel KA, Khatib H (2012) Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. Anim Genet 43:65–71. doi: 10.1111/j.1365-2052.2012.02350.x

Pryce JE, Haile-Mariam M, Verbyla K, et al (2010) Genetic markers for lactation persistency in primiparous Australian dairy cows. J Dairy Sci 93:2202–14. doi: 10.3168/jds.2009-2666

Purfield DC, Bradley DG, Evans RD, et al (2015) Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. Genet Sel Evol 47:47. doi: 10.1186/s12711-015-0126-4

Reist M, Erdin D, von Euw D, et al (2002) Estimation of Energy Balance at the Individual and Herd Level Using Blood and Milk Traits in High-Yielding Dairy Cows,. J Dairy Sci

85:3314–3327. doi: 10.3168/jds.S0022-0302(02)74420-2

Royal MD, Flint APF, Woolliams JA (2002) Genetic and phenotypic relationships among endocrine and traditional fertility traits and production traits in Holstein-Friesian dairy cows. J Dairy Sci 85:958–67. doi: 10.3168/jds.S0022-0302(02)74155-6

Sahana G, Guldbrandtsen B, Lund MS (2011) Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. J Dairy Sci 94:479–486. doi: 10.3168/jds.2010-3381

Sahana G, Guldbrandtsen B, Thomsen B, Lund MS (2013) Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. Anim Genet 44:620–626. doi: 10.1111/age.12053

Sargolzaei M (2014) SNP1101 User ' s Guide. Version 1.0.

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. doi: 10.1186/1471-2164-15-478

Schennink A, Heck JML, Bovenhuis H, et al (2008) Milk Fatty Acid Unsaturation: Genetic Parameters and Effects of Stearoyl-CoA Desaturase (SCD1) and Acyl CoA: Diacylglycerol Acyltransferase 1 (DGAT1). J Dairy Sci 91:2135–2143. doi: 10.3168/jds.2007-0825

Schennink A, Stoop WM, Visker MHPW, et al (2007) DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Anim Genet 38:467–473. doi: 10.1111/j.1365-2052.2007.01635.x

Schulman NF, Sahana G, Iso-Touru T, et al (2011) Mapping of fertility traits in Finnish Ayrshire

by genome-wide association analysis. Anim Genet 42:263–9. doi: 10.1111/j.1365-2052.2010.02149.x

Sciascia Q, Pacheco D, McCoard SA (2013) Increased milk protein synthesis in response to exogenous growth hormone is associated with changes in mechanistic (mammalian) target of rapamycin (mTOR)C1-dependent and independent cell signaling. J Dairy Sci 96:2327–38. doi: 10.3168/jds.2012-6267

Sewalem A, Miglior F, Kistemaker GJ, et al (2008) Relationship between reproduction traits and functional longevity in canadian dairy cattle. J Dairy Sci 91:1660–1668. doi: 10.3168/jds.2007-0178

Shook GE (2006) Major advances in determining appropriate selection goals. J Dairy Sci 89:1349–1361. doi: 10.3168/jds.S0022-0302(06)72202-0

Short TH, Lawlor TJ (1992) Genetic parameters of conformation traits, milk yield, and herd life in Holsteins. J Dairy Sci 75:1987–98. doi: 10.3168/jds.S0022-0302(92)77958-2

Sodeland M, Kent MP, Olsen HG, et al (2011) Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. Anim Genet 42:457–465. doi: 10.1111/j.1365-2052.2010.02165.x

VanRaden PM (2004) Invited Review: Selection on Net Merit to Improve Lifetime Profit. J Dairy Sci 87:3125–3131. doi: 10.3168/jds.S0022-0302(04)73447-5

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423. doi: 10.3168/jds.2007-0980

VanRaden PM, Van Tassell CP, Wiggans GR, et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24. doi: 10.3168/jds.2008-1514

VanRaden PM, Wiggans GR (1995) Productive Life Evaluations: Calculation, Accuracy, and Economic Value. J Dairy Sci 78:631–638. doi: 10.3168/jds.S0022-0302(95)76674-7

Veerkamp RF, Coffey MP, Berry DP, et al (2012) Genome-wide associations for feed utilisation complex in primiparous Holstein–Friesian dairy cows from experimental research herds in four European countries. Animal 6:1738–1749. doi: 10.1017/S1751731112001152

Wang X, Ma P, Liu J, et al (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 16:111. doi: 10.1186/s12863-015-0263-3

Wang X, Wurmser C, Pausch H, et al (2012) Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. PLoS One 7:e40711. doi: 10.1371/journal.pone.0040711

Wickramasinghe S, Hua S, Rincon G, et al (2011) Transcriptome profiling of bovine milk oligosaccharide metabolism genes using RNA-sequencing. PLoS One 6:e18895. doi: 10.1371/journal.pone.0018895

Wiggans GR, Sonstegard TS, VanRaden PM, et al (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. doi: 10.3168/jds.2008-1758

Wu X, Lund MS, Sahana G, et al (2015) Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. Genet Sel Evol 47:50. doi: 10.1186/s12711-015-0129-1

Table 4. 1: The number of significant SNPs (5% and 1% genome-wise FDR) using single SNP mixed linear model on imputed BovineHD (777k) genotypes in Holstein dairy cattle.

Trait	Heritability	No. of sig.** SNPs (1% FDR*)	No. of sig. SNPs (5% FDR)
Lifetime profitability index	0.370	43	53
Lactation persistency	0.363	25	83
Herd life	0.097	197	510
Indirect herd life	0.097	59	140
Direct herd life	0.097	197	662

* FDR: false discovery rate

** Sig. SNPs: Significant SNPs

Heritability is calculated by CDN and used in national genetic evaluation.

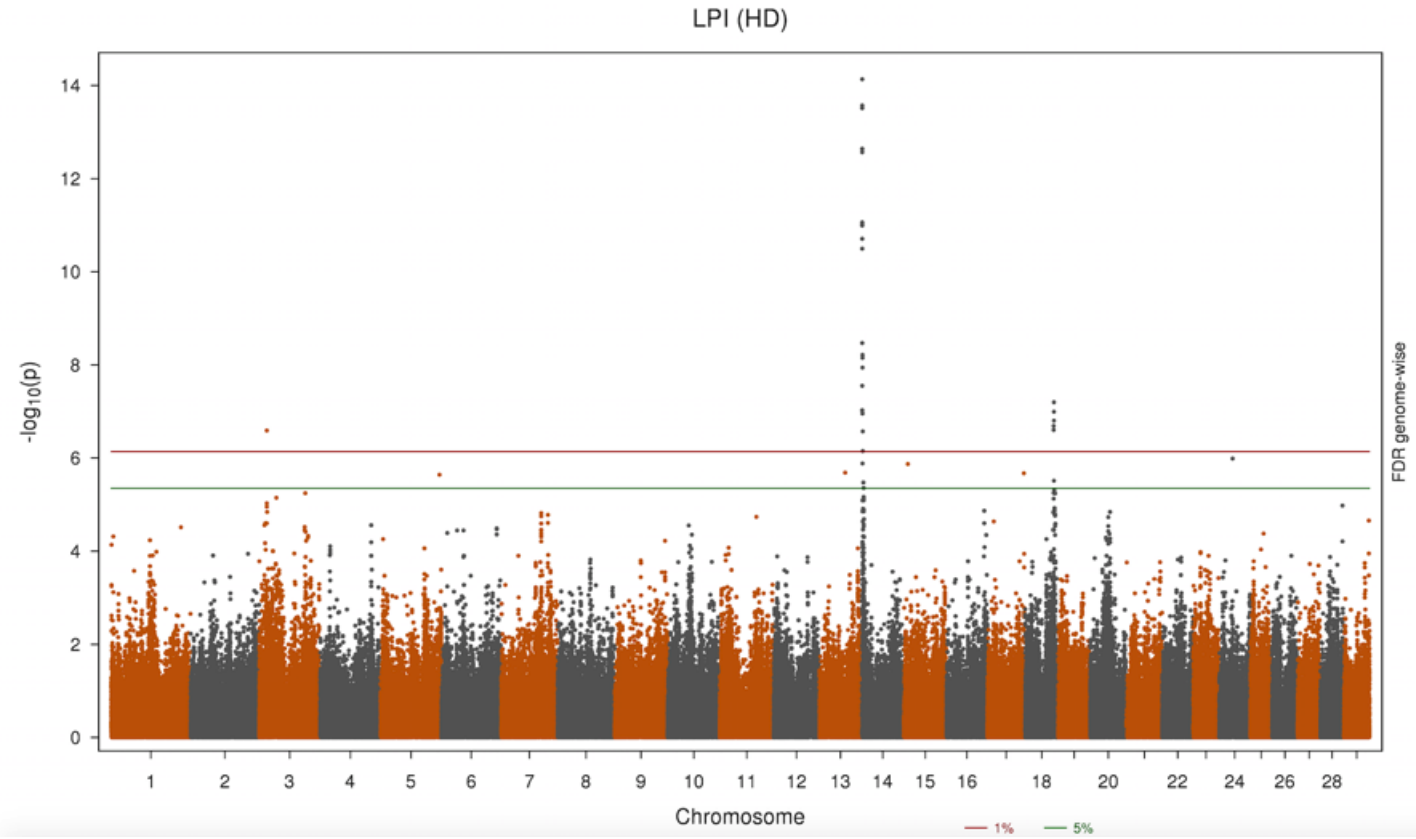


Figure 4. 1: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for lifetime profitability index (LPI). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

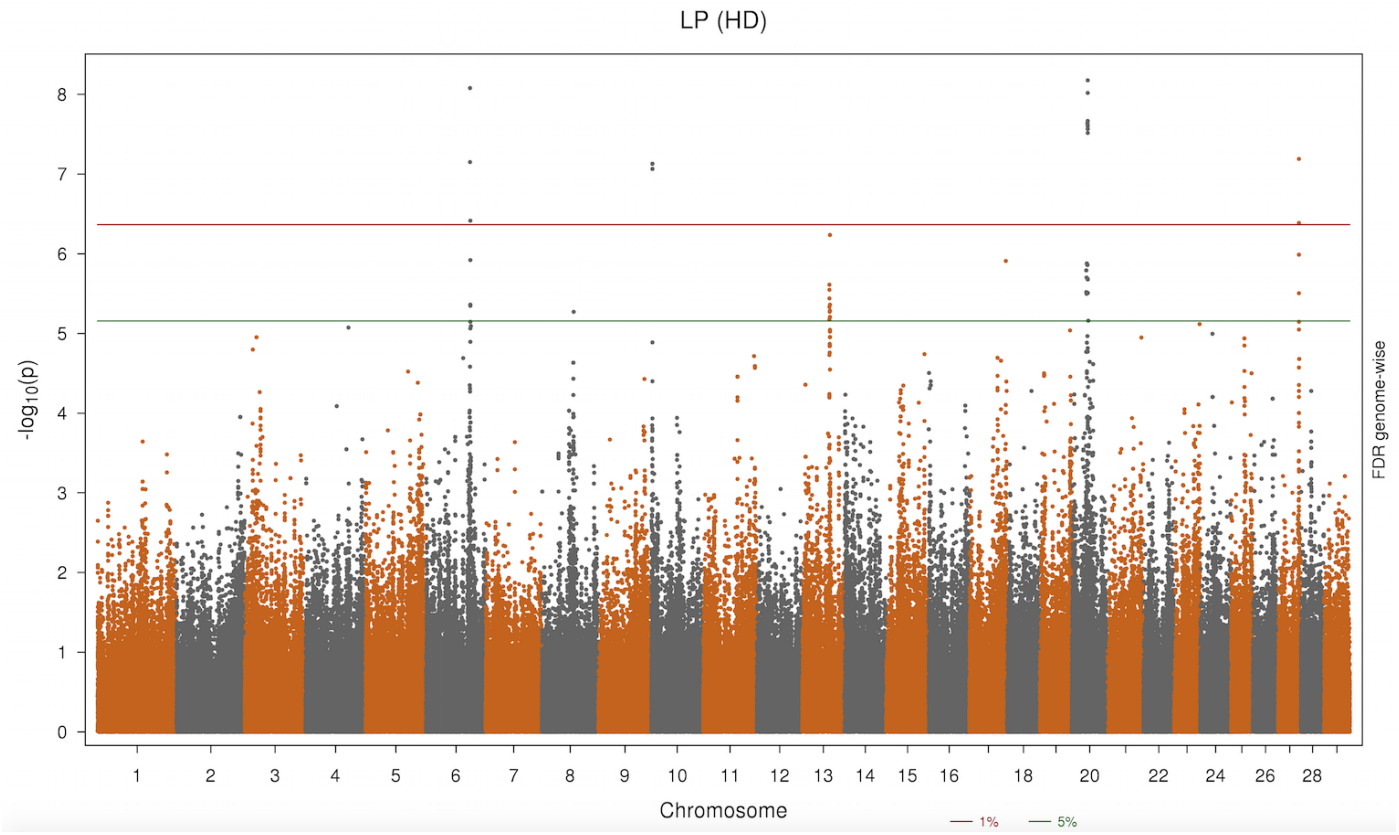


Figure 4. 2: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for lactation persistency (LP). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

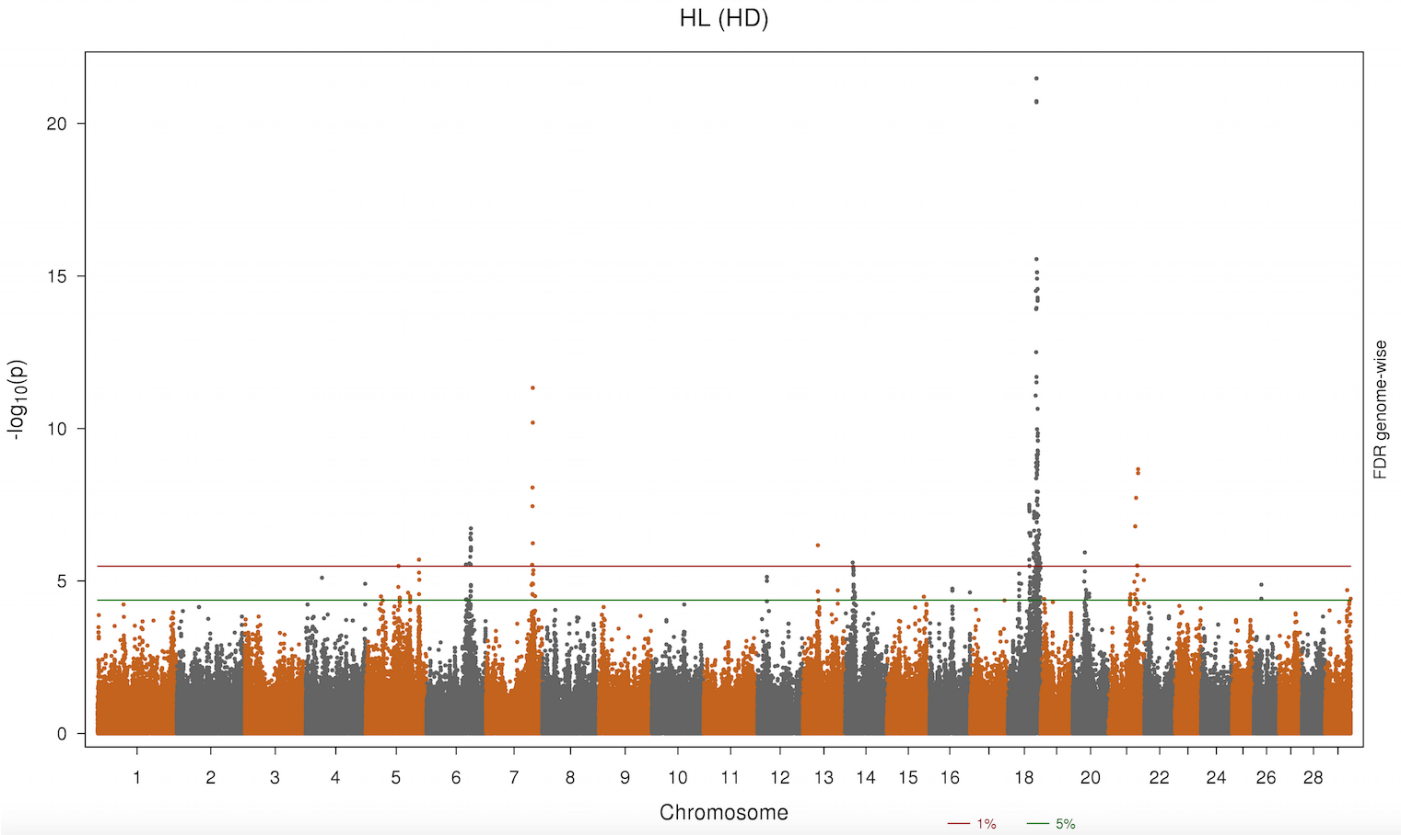


Figure 4. 3: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for herd-life (HL). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

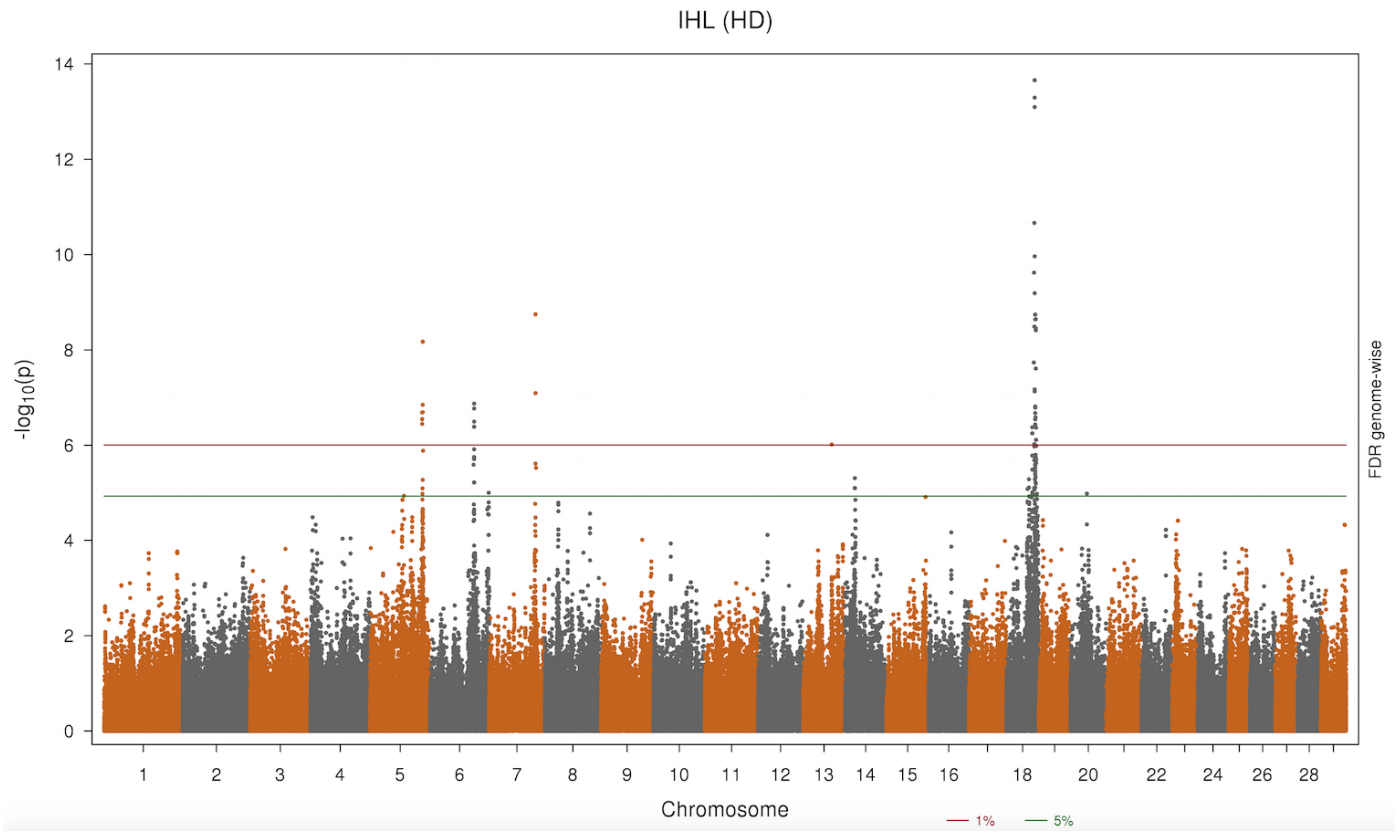


Figure 4. 4: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for indirect herd-life (IHL). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

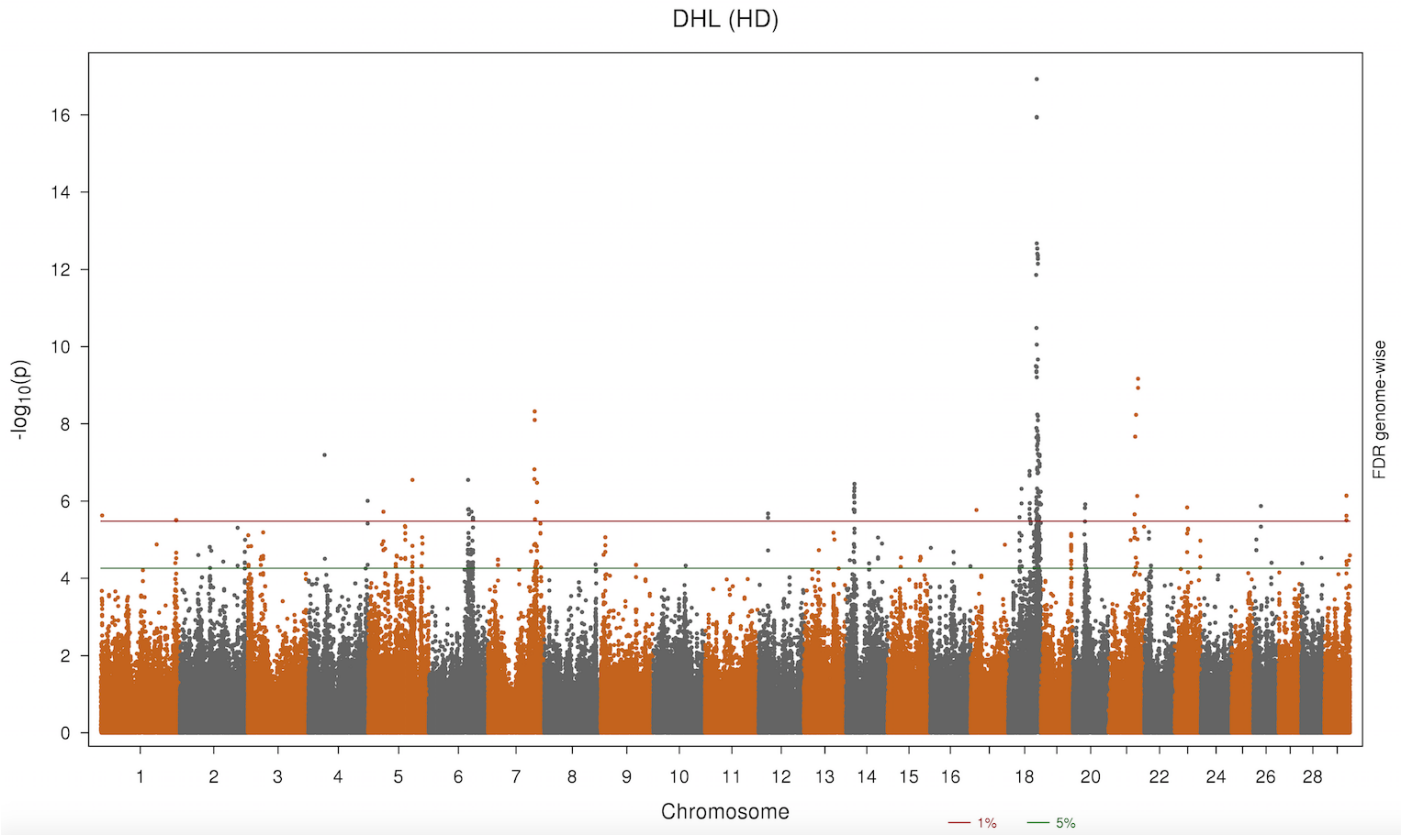


Figure 4. 5: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for direct herd-life (DHL). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

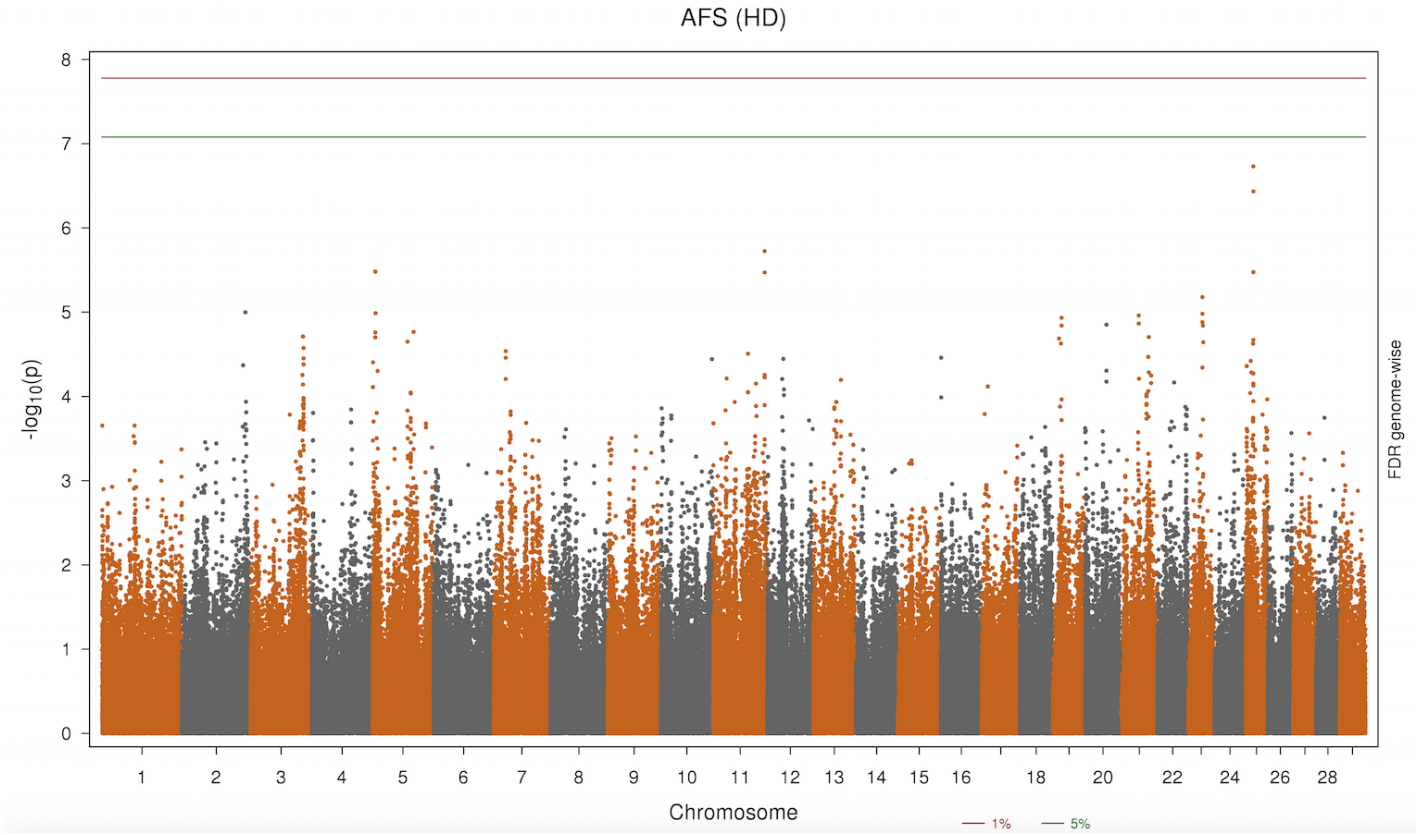


Figure 4. 6: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for age at first service (AFS). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

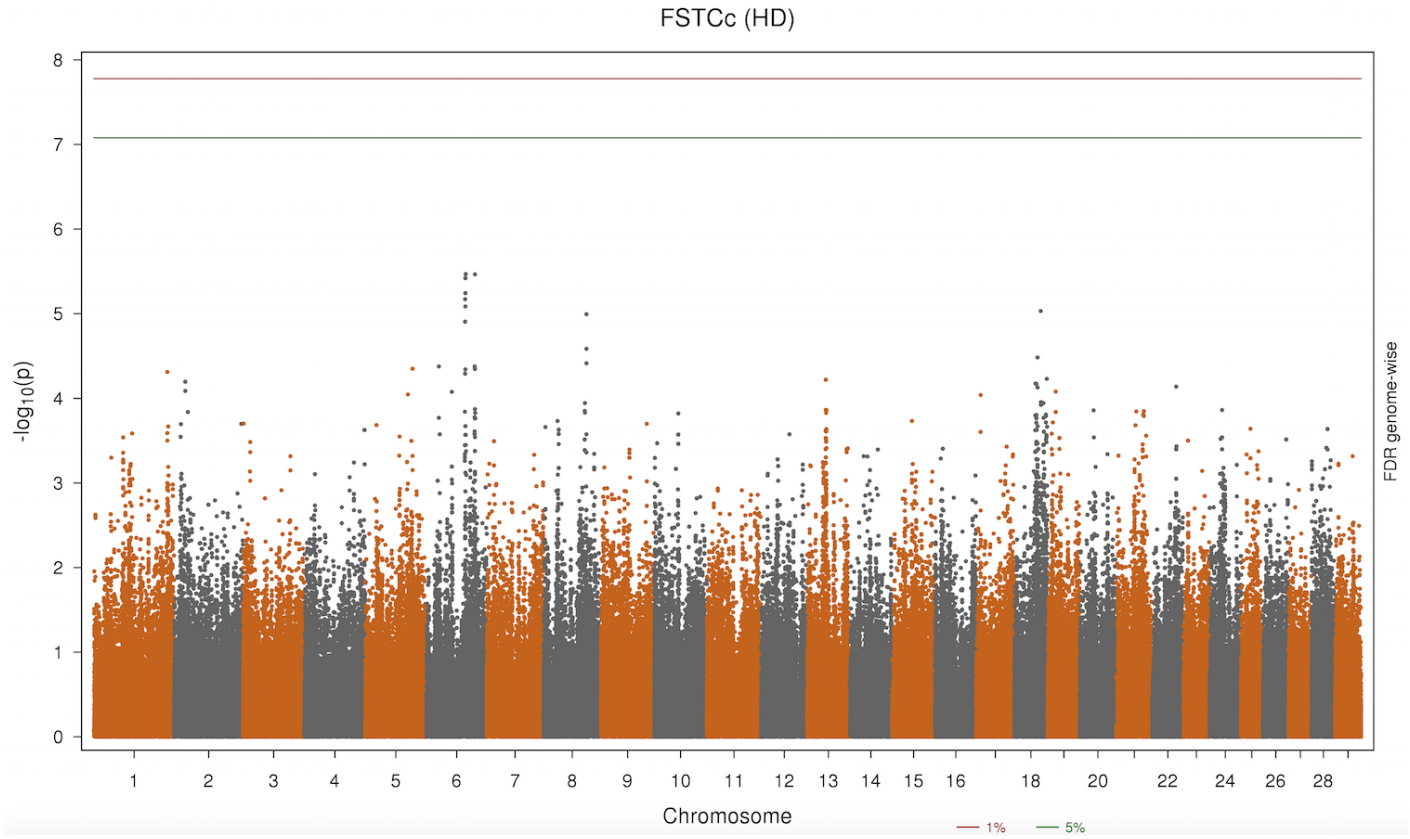


Figure 4. 7: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for first service to conception (FSTCc). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

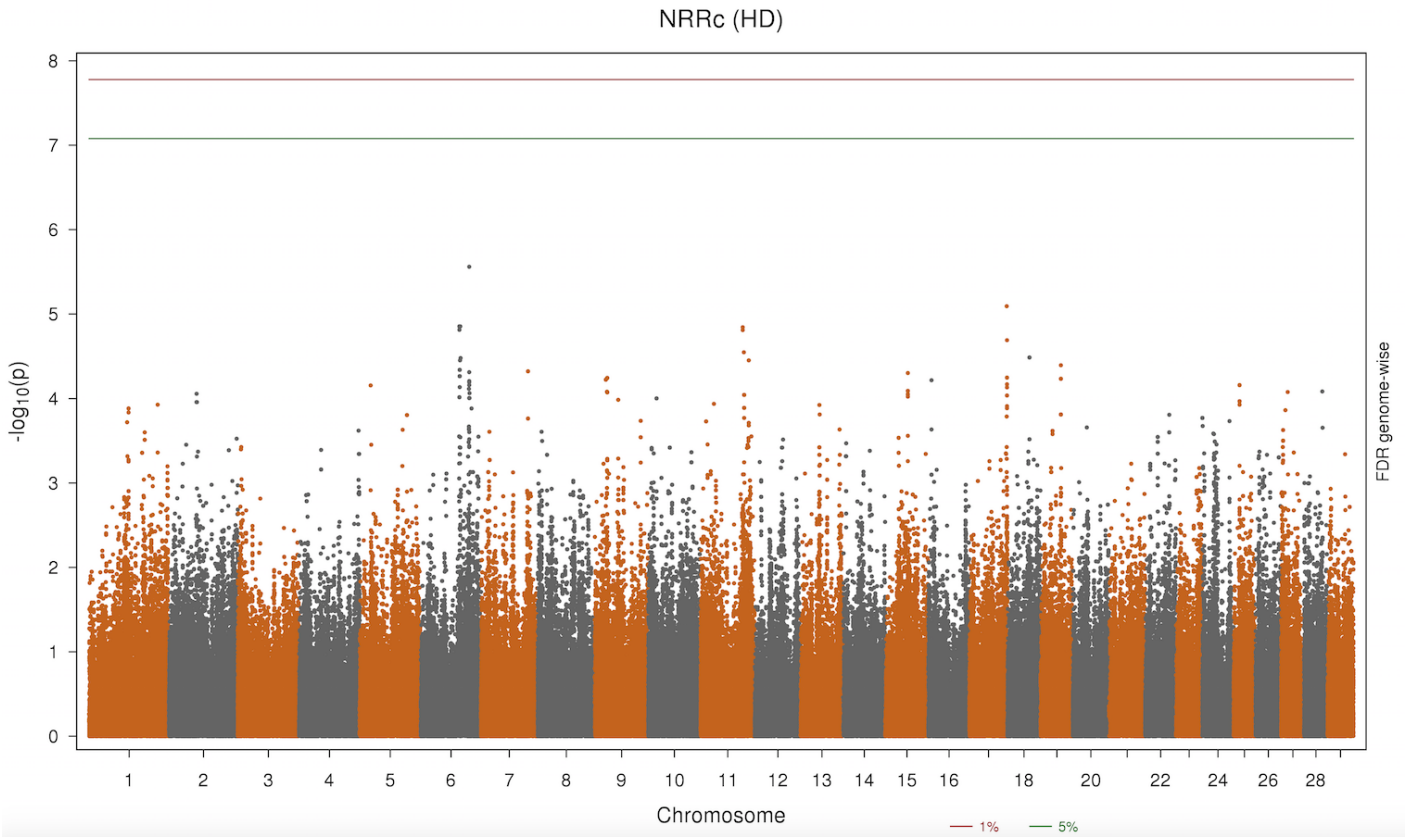


Figure 4. 8: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for cow non-return rate (NRRc). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

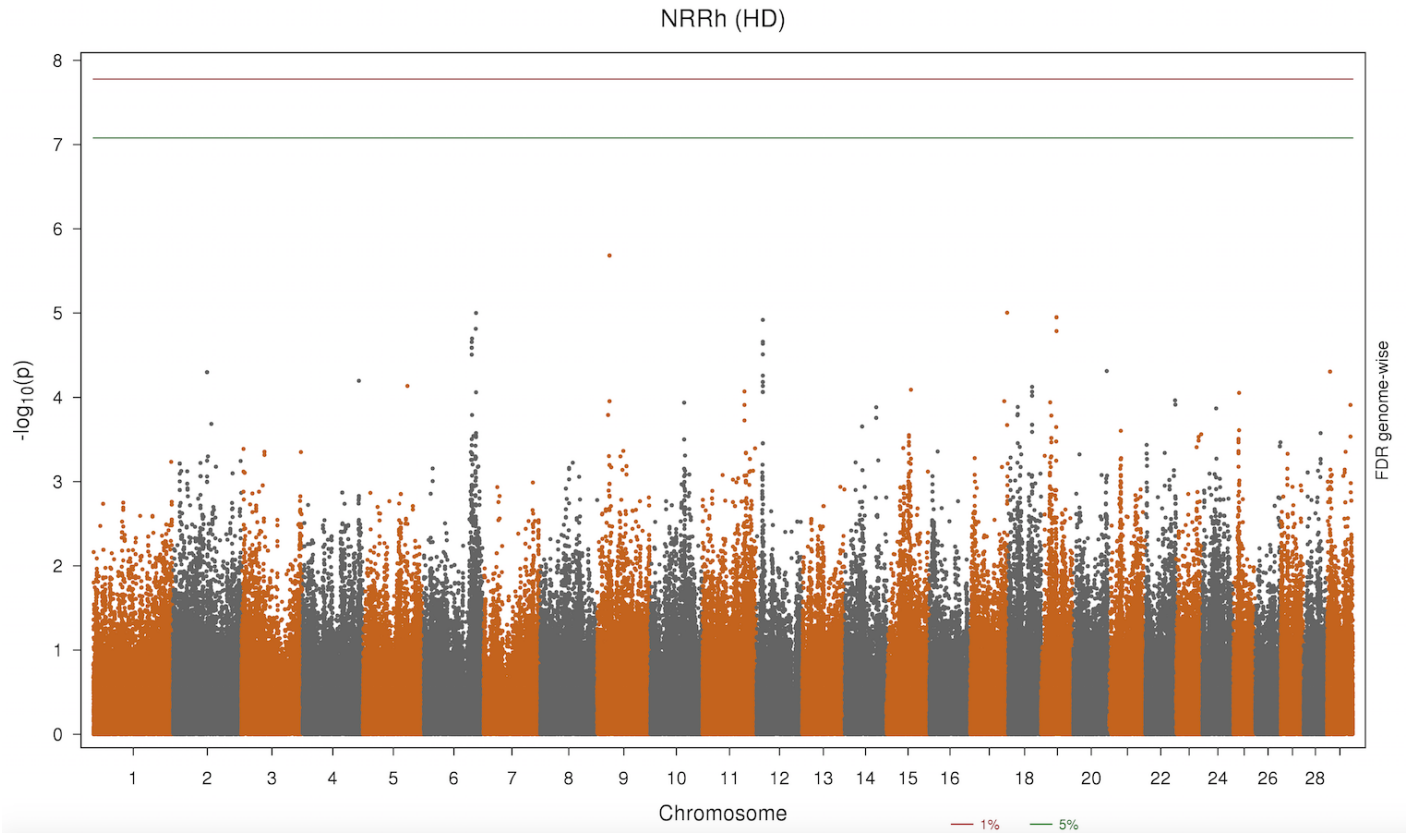


Figure 4. 9: Genome-wide association analysis of P-values of SNPs from single SNP regression mixed linear model for heifer non-return rate (NRRh). The $-\log_{10}$ of the P-value for association with SNPs is plotted. Chromosome number is shown on the horizontal axis. The red line is the threshold for significant SNPs at 1% FDR. The green line is the threshold for significant SNPs at 5% FDR.

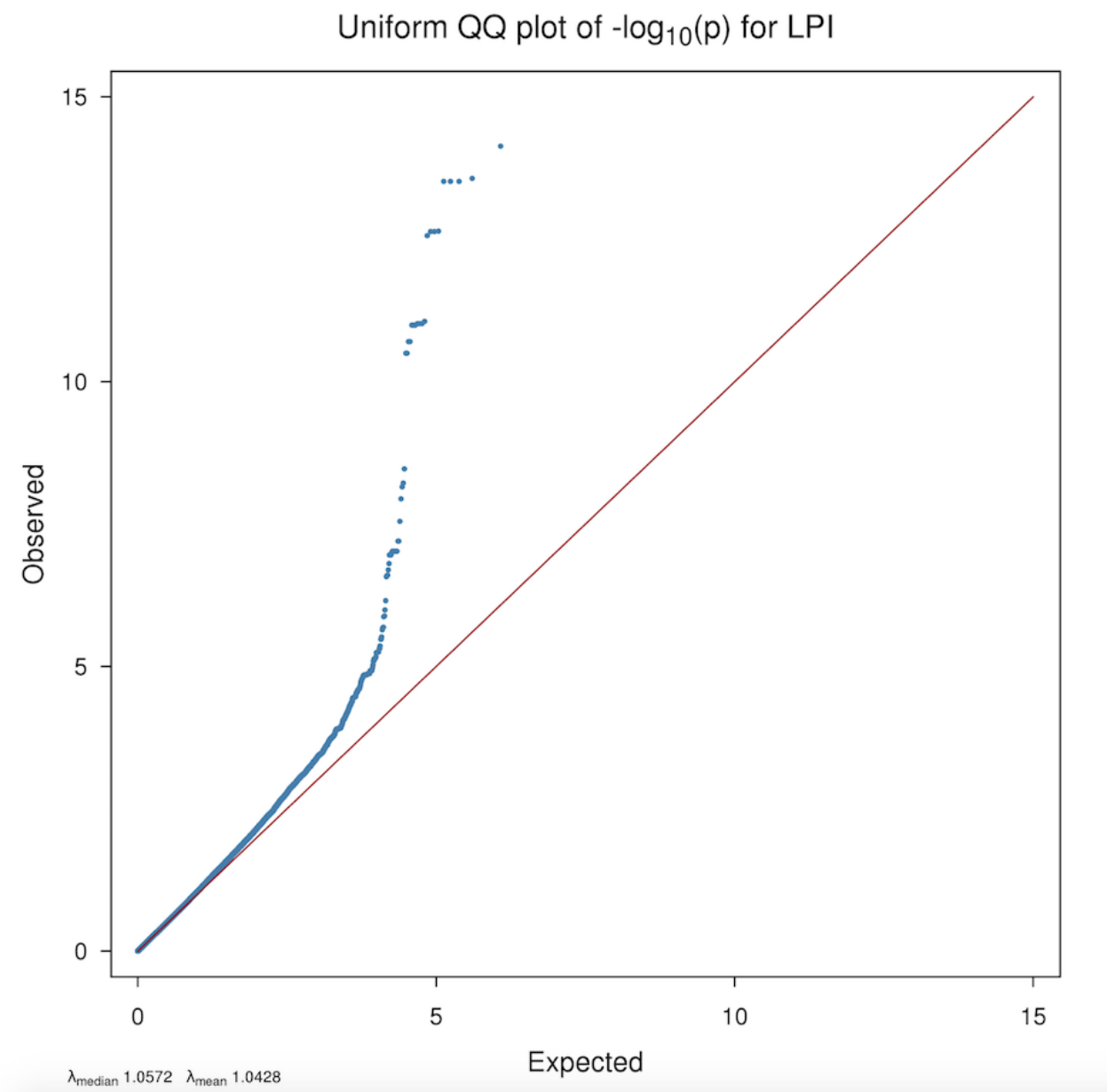


Figure 4. 10: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for lifetime profitability index (LPI). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

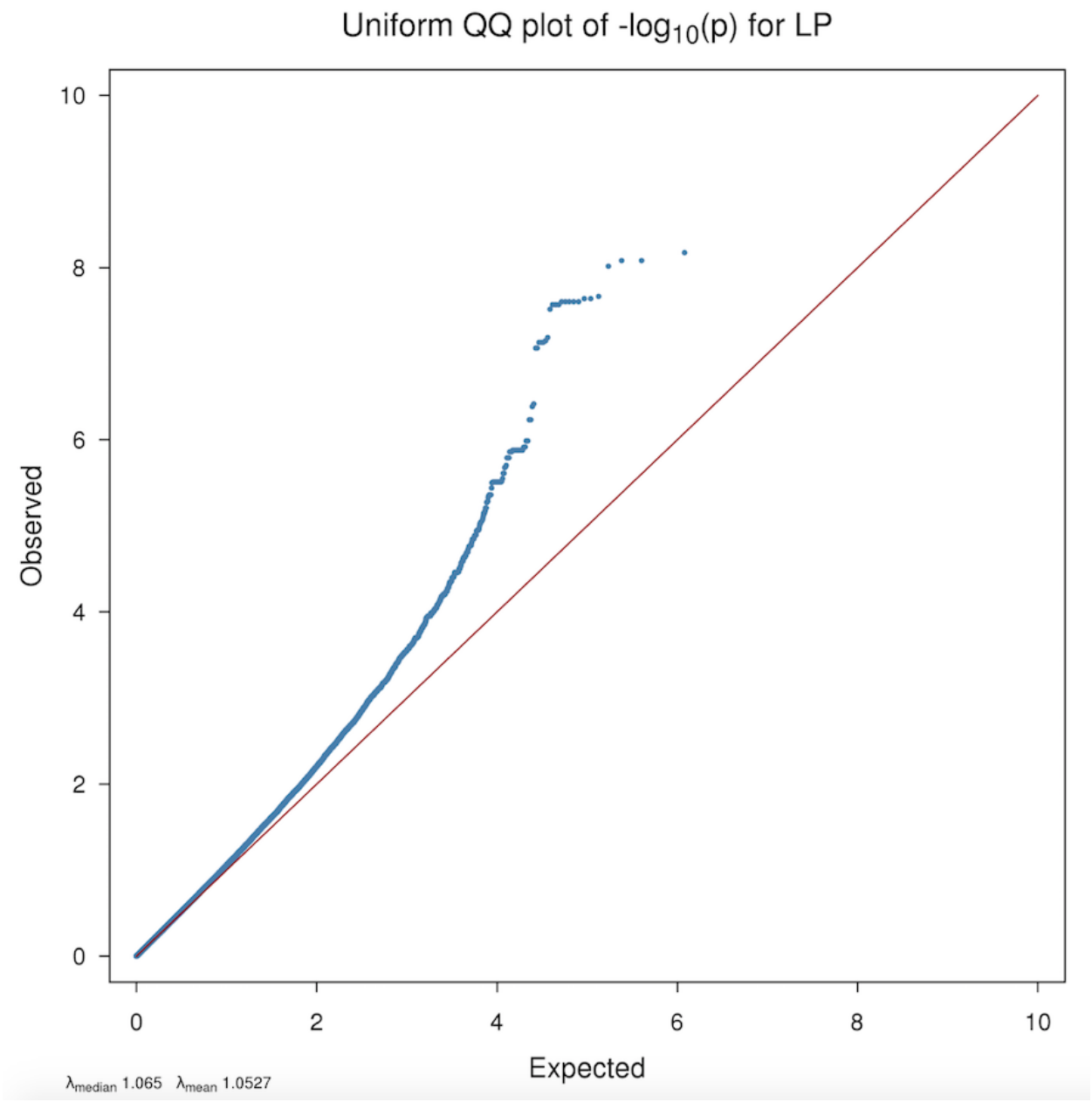


Figure 4. 11: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for lactation persistency (LP). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

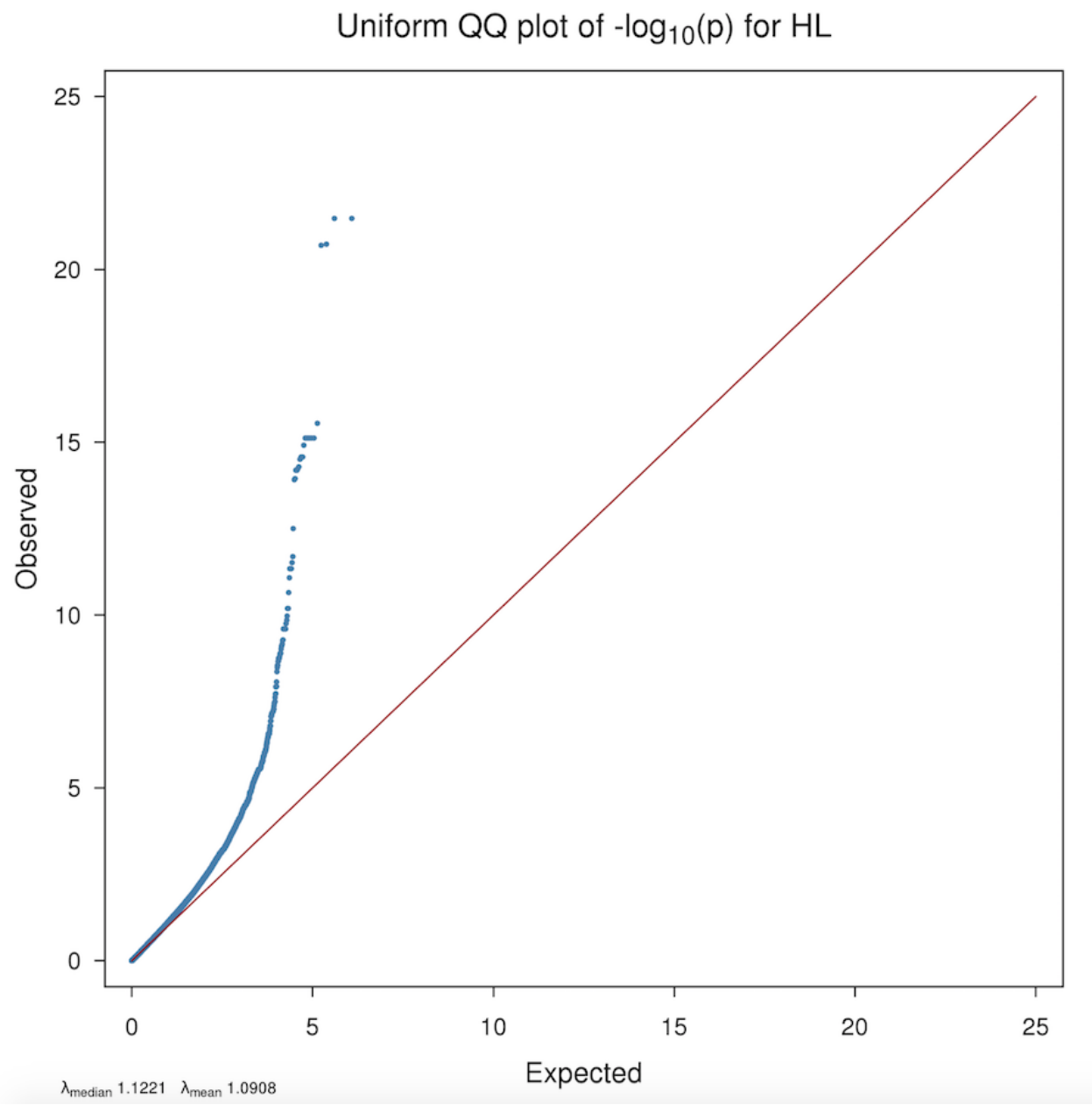


Figure 4. 12: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for herd-life (HL). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

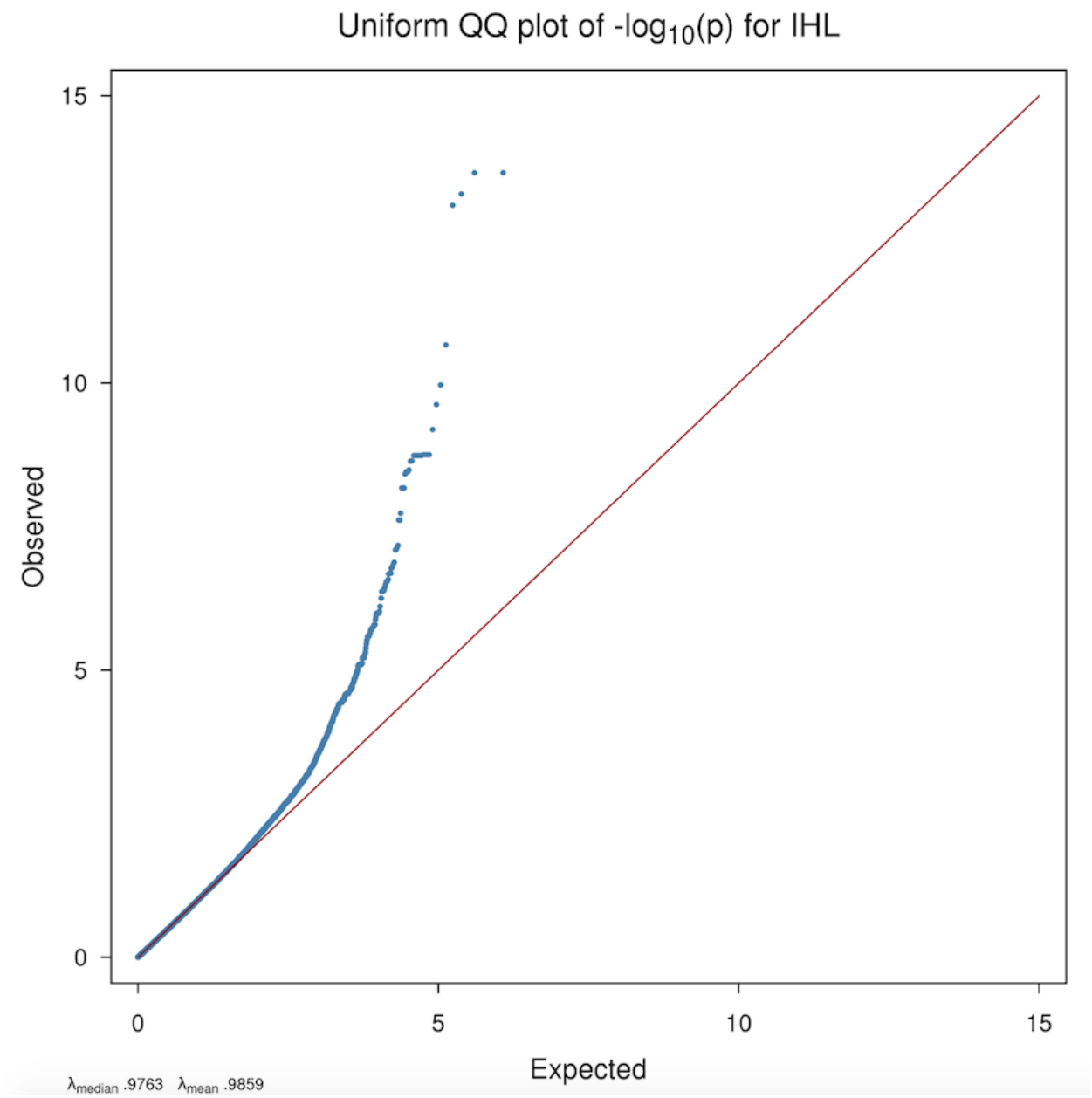


Figure 4. 13: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for indirect herd-life (IHL). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

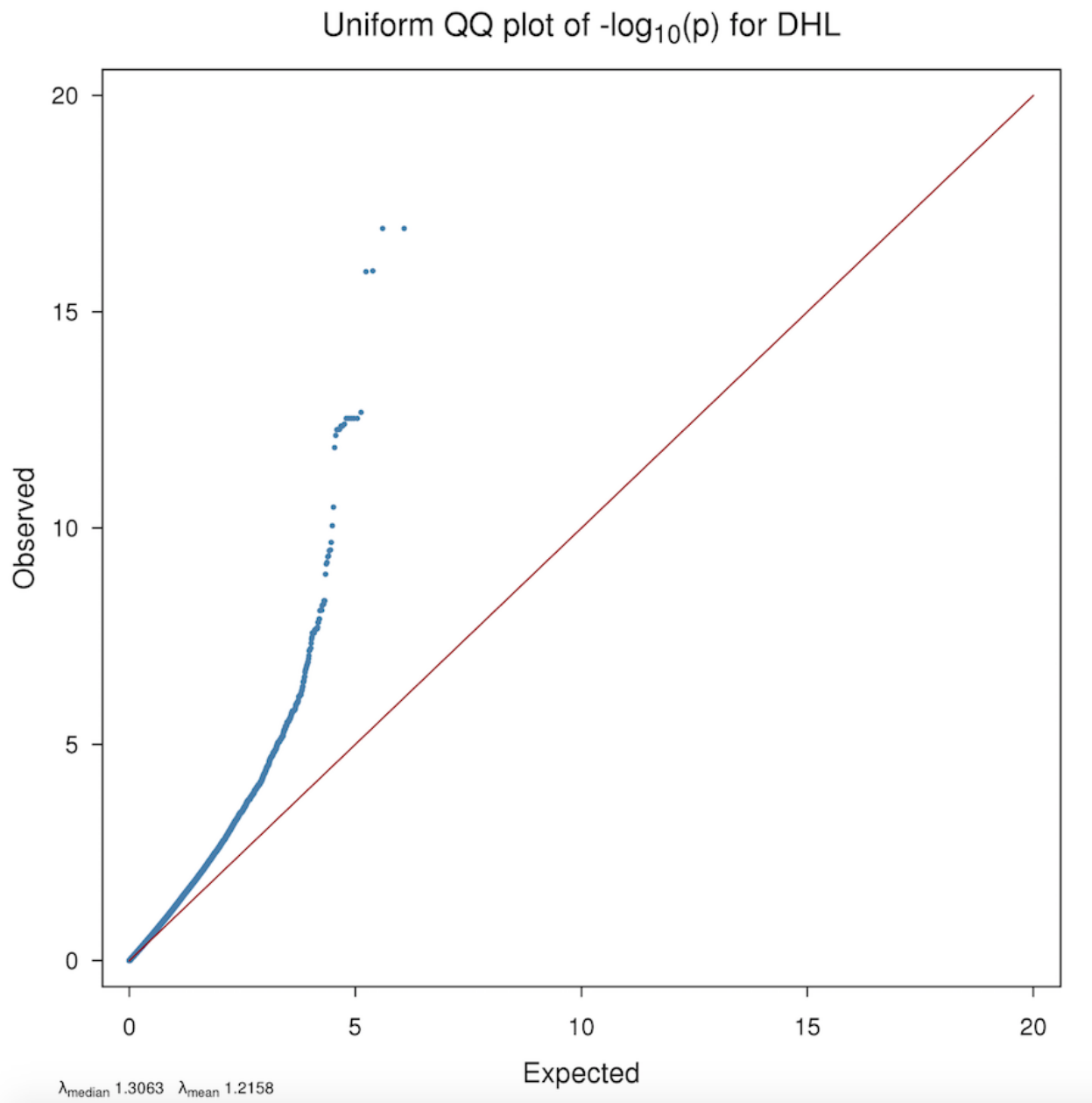


Figure 4. 14: Quantile-quantile (Q-Q) of P-values of SNPs from single SNP regression mixed linear model for direct herd-life (DHL). In the Q-Q plots the blue dots represent the $-\log_{10}$(P-values) to the expected distribution under the null hypothesis of no association. The red line denotes the expected pattern under the null hypothesis. Deviations between the red line and blue dots indicate how the test statistics of loci deviate from the null hypothesis.

CHAPTER 5. Adding new markers to genomic evaluations for milk production and fertility traits in Canadian Holstein dairy cattle through whole-genome sequencing, variant annotation, and candidate gene identification

5.1. Introduction

Milk production and fertility traits are influenced by many genes with small individual effects (Glazier et al., 2002; Schork et al., 2013). Selection for these traits was previously based on phenotypic and pedigree information and the knowledge of genetic parameters for the traits of interest (Dekkers and Hospital, 2002). However, recent advances in molecular genetic technology (Meuwissen et al., 2001; Matukumalli et al., 2009; Dekkers, 2012) in conjunction with bovine SNP discovery and sequencing projects (The Bovine Hapmap Consortium, 2009; Stothard et al., 2011; Daetwyler et al., 2014) have led to the of use genomic selection (GS) to identify genetically superior animals (Schaeffer, 2006; The Bovine Hapmap Consortium, 2009; Stothard et al., 2011; Daetwyler et al., 2014).

Selection decisions in GS are based on genomic estimated breeding values (GEBVs) obtained using genome-wide dense markers (Meuwissen et al., 2001). To calculate GEBVs, the effects of markers (mostly in the form of SNPs) are first estimated in a reference population (consisting of animals that are genotyped and phenotyped for the traits of interest). These effects are used to build a prediction equation that is then applied to a second population, consisting of selection candidates, for which genotype information is available but not necessarily phenotype information. The estimated effects of the markers that each animal carries are summed across the

whole-genome to calculate the GEBV (Meuwissen et al., 2001; Hayes et al., 2009) representing the genetic merit of the individual (Meuwissen and Goddard, 2007). The GS approach can increase the accuracy of selection (Meuwissen et al., 2001; Schaeffer, 2006; Poland and Rutoski, 2015), and can reduce generation intervals. In the dairy industry generation intervals declined from 5 to 6 years under traditional selection to 1.5 years with GS (Pryce and Daetwyler, 2012). This is because rather than waiting until a bull has daughters with phenotypic records, GS can predict breeding value of selection candidates based solely on genotype (Meuwissen et al., 2001; Pryce and Daetwyler, 2012), although caution is warranted when making decisions based only on genotype (Littlejohn et al., 2014). Genomic selection can be applied without knowledge of the underlying genes and pathways involved, however an understanding of the corresponding quantitative trait nucleotides (QTNs) may be valuable (Meuwissen and Goddard, 2007; Goddard and Hayes, 2009). Knowledge of the QTNs or the markers in high LD with the causal mutations can a) provide a better understanding of the physiology and biological pathways underlying variations in traits; b) allow for the use of a smaller DNA marker panel (containing QTN) for selection purposes, c) further increase the accuracy of genome-based breeding values, and d) potentially allow these panels to be applied across breeds due to a more direct link between genotype and phenotype (Snelling et al., 2013).

Potential QTNs can be identified using information from variety of sources. Positional candidate genes can be identified based on the results of association studies, for example a GWAS (genome-wide association study) (Veerkamp and Beerda, 2007). Existing knowledge of gene function and expression can be used to further refine the list of positional candidates (Snelling et al., 2013). DNA variation can then be identified within the candidate genes. Recently developed next-generation sequencing (NGS) technologies are helpful in this regard as large numbers of

animals can be characterized through genome sequencing or RNA-seq (Snelling et al., 2013; Koufariotis et al., 2014). The potential impacts/effects of variants can be predicted using bioinformatics tools (Grant et al., 2011), and further association studies and functional studies can be carried out to verify the importance of the putative QTNs (Ron and Weller, 2007).

The objective of this study was to identify candidate genes and mutations for production and fertility traits in Holstein dairy cattle by integrating information from GWAS, whole-genome sequencing and a variety of bioinformatics tools. A new genotyping panel was then constructed and evaluated to test whether it could lead to improved genomic predictions for eighteen production and fertility traits in Canadian Holstein dairy cattle.

5.2. Materials and method

5.2.1. Animals, phenotype and genotype data

Eighteen milk production and fertility associated traits were used in this study including milk production (MILK), fat production (FAT), protein production (PROT), fat deviation (FATD), protein deviation (PROTD), herd life (HL), lactation persistency (LP), lifetime profitability index (LPI), direct herd life (DHL), indirect herd life (IHL), daughter fertility (DF), age at first service (AFS), heifer non-return rate (NRRh), heifer first service to conception (FSTCh), cow non-return rate (NRRc), calving to first service interval (CTFS), cow first service to conception (FSTCc) and days open (DO). The Canadian Dairy Network (CDN) provided data for Holstein cattle with official evaluation for proven bulls in 2008. SNP genotypes for 16,054 Holstein bulls were obtained using the Illumina BovineSNP50K BeadChip (Illumina, San Diego, CA) and imputed to Illumina Bovine HD (777k) BeadChip genotypes using the FImpute V.2.2 software (Sargolzaei, 2012) and a reference set of Illumina Bovine HD from 1,659 animals. Quality

control (QC) was performed on genotyping data using the Sleuth software (Mehdi Sargolzaei) to remove SNPs with less than a 90% call rate, a minor allele frequency (MAF) of less than 1%, or heterozygosity more than 15%. After quality control, a total number of 311,725 SNPs across the bovine autosomal chromosomes remained for use in association analysis.

5.2.2. Genome-wide association analysis

Association analysis was performed by means of the Generalized Quasi-Likelihood Score (GQLS) method developed by Feng et al. (Feng et al., 2011), and implemented in the Sleuth software package. The GQLS method can handle the correlation structure among related subjects and is based on the logistic regression model to link the trait to the distribution of allelic frequencies (Feng et al., 2011). In this model the observed phenotype of each individual is treated as a covariate and the proportion of a specified allele in the genotype is the response (Feng et al., 2011). Breeding values were estimated for each trait and then association analysis was performed using both BovineSNP50 and the high-density (HD) imputed SNP datasets. A genome-wise false discovery threshold (FDR) of 5% was used to correct for multiple testing (Benjamini and Hochberg, 1995).

5.2.3. Complementary DNA (cDNA) library construction and sequencing

Tissue samples were collected from adipose (8 animals in pool, 4 pre-partum and 4 post-partum) (collected under Washington State University Institutional Animal Care and Use Committee, IACUC; project 3478), mammary gland (6 animals in pool, 3 non-lactating and 3 lactating, University of Vermont, USA) and liver (8 animals in pool, 4 non-lactating and 4 lactating; University of Alberta, Animal Care protocol number 143/10/11). Libraries were prepared using reagents and protocols provided by Applied Biosystems (ABI) and sequenced using the ABI

SOLiD 3 sequencer (Life Technologies Corporation, CA, USA). Sequence reads were mapped to the UMD 3.1 bovine genome assembly using the LifeScope 2.0 software with default settings (Life Technologies Corporation, CA, USA). Putative SNPs and indels were identified using SAMtools (Li et al., 2009). Variants were removed if they had a quality score less than 20, more than two alleles, or less than four reads supporting the alternative allele.

5.2.4. Candidate gene identification

Gene mapping and positional candidate gene identification

Significant QTL were detected at 5% FDR genome-wise from GWAS results. Then, for each trait considered in this study, a set of most significant regions that were common between the GWAS results/analyses with the BovineSNP50K and imputed HD (777K) panels were selected. We also selected new peak regions that were only detected in high density GWAS result/analysis. A list of significant markers from the peak regions was generated for all the traits. The position of each significant SNP was used to identify overlapping and nearby genes (within a distance of 1000 bp), using the Ensembl (release 67) Perl API. A list of these positional candidate genes was built and used for functional candidate gene identification.

Functional candidate gene identification

The positional candidate genes were prioritized based on their involvement in biological processes and relationships with other genes or gene products, using ingenuity pathway analysis (IPA). The IPA system uses records from the Ingenuity Pathway Knowledge Base (IPKB), which is the largest curated database of previously published findings on genes and their interactions in mammalians such as human, mouse and rat (Ficenec et al., 2003; Calvano et al.,

2005). For this study the network analyses were performed by filtering out the tissues that are of less interest and specifying tissues and organs known to have key roles in fat, protein, carbohydrate and reproductive system hormone metabolism including mammary gland, liver, ovary, pancreas, adrenal, pituitary and hypothalamus gland. These tissues are known to have important roles early in lactation, pre and post-partum in dairy cattle. The results were then used to identify the positional candidate genes that were directly or indirectly inter-connected and represented in significant pathways and biological processes.

5.2.5. Marker selection for inclusion in a custom genotyping panel

The positional candidate genes deemed by *in silico* analysis to represent functional candidates formed the basis of the SNP and indel identification step. For each candidate gene, overlapping SNPs and indels were obtained from three sources: Ensembl (release 67), whole-genome sequencing of two bulls (Stothard et al., 2011), and the aforementioned RNA-Seq. NGS-SNP (Grant et al., 2011) was used to assign a functional class to each SNP and indel and to provide flanking sequences extracted from the UMD3.1 reference assembly for use in panel design. This software uses information from Ensembl, NCBI and UniProt to place SNPs into functional categories (Grant et al., 2011). Next, variants that were close to another variant (within 20 bp) or with flanking sequences that yielded multiple BLAST hits (E-value = 0.003 using 100 nt before and 100 nt after variant site) when compared back to the UMD3.1 assembly were removed from consideration, unless they were predicted to have a dramatic functional consequence (stop gained, frame-shift, stop lost, initiator codon variant, inframe insertion, inframe deletion, missense variant, or mature miRNA variant). Lastly, for each gene the remaining set of candidate markers was sorted based on functional consequence, followed by source (variants observed in multiple data sets were ranked higher than those observed in a single data set). This list of

refined variants was submitted to Affymetrix for further evaluation using their probe design system. In this step, variants were excluded if they were already available on the Illumina HD panel, or if they were flagged as unsuitable by the panel design software used in-house by Affymetrix. Finally, a custom Affymetrix panel was ordered, consisting of 135,256 markers.

5.2.6. Genotyping

A total of 1,326 Holstein bulls (poof bulls from 2008) were genotyped using the custom panel. Quality control (QC) was performed on the genotyping results and variants exhibiting a MAF < 0.005 (n=39,582 SNPs), excess of heterozygosity > 0.15 (n=9 SNPs), or departure from Hardy-Weinberg equilibrium P < 0.00001 (n=350 SNPs) were excluded. Also, variants with low call rates (< 90%) were removed. Removing these variants from the genotyping results produced a final set of 80,294 SNPs, which we refer to as the "Affymetrix 80K" SNP panel. Combining these 80K SNPs with those genotyped using the 50K panel (44,369 SNPs remaining after quality control) yields what we refer to as the "124K SNP panel" (124,663 SNPs). Lastly, a "transcriptome panel" (with 74,884 SNPs) was prepared *in silico* by removing SNPs from the 124K SNP panel that were located in the intergenic, upstream, downstream or intronic regions of the bovine genome.

5.2.7. Imputation

50K panel genotypes from 53,022 Holstein bulls were imputed to the 124K panel using FImpute V.2.2 (Sargolzaei, 2012) and a reference set of 124K genotypes obtained from 1,326 Holstein bulls. The accuracy of imputation was evaluated by calculating the concordance rate between the imputed genotype and the true genotype (Browning and Browning, 2008). To calculate the accuracy, from 1,326 animals that were genotyped with 124K panel 319 younger animals (born

on 2006 or after) were considered as target group considering just their 50K genotypes (this group of animals used to validate the accuracy of imputation) and the remaining 1007 animals were considered as a reference group. Allelic r^2 was also calculated as the squared correlation between imputed and true genotypes (Browning and Browning, 2008).

5.2.8. Genomic prediction and validation

Genomic evaluation to assess the utility of the new custom SNP panel and other subsets of SNPs as tools for selection was performed using the genomic best linear unbiased prediction (GBLUP) method (VanRaden, 2008; VanRaden et al., 2009) implemented in the gebv genomic evaluation software (Sargolzaei et al., 2013). A total of 11,461 bulls (with domestic or MACE EBVs in 2008) were used as the reference population for genomic evaluation. The selection candidate population (used here as a population to validate prediction accuracy) consisted of animals born between years 2004 to 2008 without proofs in April 2008 but with domestic proofs in December 2012. Direct genomic breeding values (DGV) were obtained and the accuracy of genomic prediction was measured for each trait as the correlation between the DGVs of animals in 2008 and the estimated breeding values of domestic proofs in 2012. The SNP subsets evaluated were 50K, imputed Affymetrix 80K, imputed 124K genotypes (124K SNP panel) and imputed Transcriptome genotypes (Transcriptome panel).

5.2.9. GBLUP statistical model

A linear genomic model was computed for genomic predictions as described in VanRaden (2008). In this prediction method, the traditional additive genetic relationship matrix between individuals is replaced by a genomic relationship matrix derived from markers and equal genetic

variance is assigned to all markers (Nejati-Javaremi, 1997; VanRaden et al., 2009). The model is as below:

$$y = Xb + Zu + e$$

Where y is the vector of phenotypic records on individuals; Xb is the mean in which b is the vector of fixed effects of generations and X is an incidence matrix relating fixed effects of generations to the records; e is a random error vector with variance $R\delta_e^2$. Matrix R is diagonal with elements, $R_{ii} = \frac{1}{R_{dau}} - 1$, where R_{dau} is the bull's reliability obtained from daughters. u is the vector of additive genetic effects that correspond to allele substitution effects for each marker. The sum of Zu over all marker loci is assumed to equal the vector of breeding values (a). GEBVs are obtained using the selection index equation (Nejati-Javaremi, 1997; VanRaden et al., 2009), which is used to predict $\hat{a}$ directly using the genomic relationship matrix G. The selection index equation is constructed as below:

$$\hat{a} = G\left[G + R\left(\frac{\sigma_e^2}{\sigma_a^2}\right)\right]^{-1}(y - X\hat{b}).$$

In which G is genomic relationship matrix and is computed as,

$$\frac{ZZ'}{2\sum P_i(1 - P_i)}$$

In this equation P_i is the allele frequency of the i^{th} SNP.

In this approach allele frequencies are estimated in the base (founder) population with a linear model that solves for gene content of non-genotyped ancestors and descendants using pedigrees (Gengler et al., 2007; VanRaden et al., 2009). The known genotypes are treated as data and the unknown genotypes of relatives are estimated using the inverse of the traditional relationship

matrix and standard mixed model equations. Allele frequencies were obtained from the known genotypes.

In order to visualize and test for the effect of population stratification on the accuracy of genomic predictions, population stratification was investigated for 11,461 animals that passed quality control, using PLINK v1.07 (Purcell et al., 2007). Figures (Figure 5 – 1 and Figure 5 – 2) of the population structure were obtained using three dimensions of a classical multidimensional scaling (MDS).

5.3. Results

5.3.1. Association analysis

Association analysis identified strong associations for all of the production and fertility traits in this study. The number of significant SNPs varied depending on the trait between 22,879 (genome-wise FDR 1%) for protein deviation (PROTD) to 279 SNPs (genome-wise FDR 1%) for heifer non-return rate (Table 5.1). Strong peaks and associated significant SNPs (genome-wise FDR 1%) were used for candidate gene identification.

5.3.2. Analysis of gene networks and pathway analyses

A list of positional candidate genes was constructed for each trait under consideration, consisting of genes located within 1000 bases of a significant SNP from the GWAS analysis. In total 4,050 positional candidate genes were identified, most of which were shared among multiple traits.

To incorporate gene function criteria into gene selection, gene-network analysis was performed using IPA. For this purpose, IPA network analysis was performed on the 4,050 identified genes.

A summary of the five top significant gene networks (P-value < 0.05) identified by IPA is given in Table 5.2. Results from IPA indicated that the gene sterol regulatory element binding transcription factor 1 (*SREBF1*) was directly and indirectly interacting with several other genes in our list including *PCK1*, *FADS1*, *ACACB*, *FASN*, *SCD*, *INSIG2*, *ELOVL5*. This result shows components of a lipid metabolism gene network overrepresented in the list of GWAS genes and the SNPs within these genes (from our GWAS analysis) were reported to be associated with multiple production and fertility traits including fat production (FAT), protein production (PROT), fat deviation (FATD), age at first service (AFS), heifer first service to conception (FSTCh) and lifetime profitability index (LPI). The SREBF1 protein is known to activate the expression of lipogenic genes such as acetyl-CoA carboxylase α, fatty acid synthase (*FASN*), *ELOVL* fatty acid elongase 6 and stearoyl-CoA desaturase (Horton et al., 2002; Nafikov et al., 2014). Mutations in the *SREBF1* gene have also been reported to be associated with milk production and variations in lauric (12:0) and myristic acid concentrations in milk (Nafikov et al., 2013). The significant SNP within *SREBF1* gene identified in the GWAS study, BovineHD1900010279, was associated with traits fat deviation (FATD) and protein deviation (PROTD).

Components of an IPA gene network associated with embryonic development, organismal and cellular development were also over-represented in the GWAS results. This network includes the candidate genes insulin receptor (*INSR*), fibronectin 1 (*FN1*), insulin like growth factor 2 (IGF2), CD83 molecule (CD83), activating transcription factor 3 (ATF3), LDL receptor related protein 5 (LRP5) and protein tyrosine kinase 2 (*PTK2*). The *PTK2* gene is located on BTA14 (2 Mb upstream from the *DGAT1* gene). It has been shown that several SNPs within this gene were significantly associated with involution pathways, milk production and milk composition in

Holstein cattle (Wang et al., 2012; Raven et al., 2014). The *IGF2* gene has been reported to be related to fetal growth in human (Frost and Moore, 2010) and placenta development and tissue differentiation in bovine (Gebert et al., 2006). The SNPs assigned to these genes from our GWAS analysis were significantly associated to lifetime profitability index (LPI), milk production (MILK), fat production (FAT), protein production (PROT), fat and protein deviation (FATD, PROTD) and direct herd life (DHL).

The IPA gene network analysis results were used to select 2,500 positional and functional candidate genes from the list of 4,050 positional candidate genes.

5.3.3. Affymetrix panel design

For each functional candidate gene, overlapping SNPs and indels were obtained from Ensembl (release 67), whole-genome sequencing and the RNA-Seq data. RNA-seq was built from tissues samples, liver, adipose and mammary gland and sequencing was performed. From the total number of 921,197 SNPs (annotated with NGS-SNP), 210,189 variants were fully known and 711,008 were identified as novel variants (Table 5.3). Following guidelines from Affymetrix, a custom panel was built.

5.3.4. Genomic prediction result using new custom panel

The utility of subsets of SNPs from the custom genotyping panel were evaluated in terms of their performance for genomic prediction. For each SNP set the correlation [r(DGVs/TBV)] and accuracy of genomic breeding values (the squared correlation of the DGVs for the validation group and the corresponding proofs in December 2012; r^2) were calculated (Table 5.4).

Imputation accuracy was evaluated (concordance rate of 99.47% and allelic r^2 of 99.18%) and then genomic prediction accuracy accuracies obtained with the imputed 124K SNP panel were compared with those obtained with the 50K panel. The prediction accuracy of GEBVs from the imputed 124K panel were generally slightly higher than those obtained using the 50K SNP panel (Table 5.4).

In order to decrease redundancy as a result of imputation in calculating genomic breeding values, an LD analysis (SNP pruning) was performed on the imputed 124K genotypes. In this analysis if a pair of SNP exhibited high LD ($r^2 > 0.95$), then the SNP that was not on the 50K panel was removed. If both SNPs were on the Affymetrix 80K or both on the 50K panel, then the SNP that had the smaller base pair, position was removed. Performing LD-based SNP pruning led to the removal of 16,461 SNPs from the 124K panel and the accuracy between true breeding values and predicted values in the validation dataset were calculated (Table 5.5). Removing redundant SNPs led to a small change in the average of the prediction accuracies, and was 0.366 (Table 5.5). This result was in agreement with Harris et al. (2010) which demonstrated that accuracy was slightly different when density of the panel increased from 20K to 1000K.

We also compared prediction accuracies resulting from different SNP densities (Table 5.5). For this purpose, we selected SNPs that were 30 kb (28K selected SNPs in the panel), 38 kb (26K selected SNPs in the panel) and 89 kb (15K selected SNPs in the panel) apart and genomic breeding values were predicted for the selected number of SNPs in the panel. As the number of markers increased in the panel (distance between variants decreased), the accuracy over the 50K panel tended to increase and was the greatest when SNPs were 38 kb apart (averaged accuracy across all traits was 0.3611) (Table 5.5) and was similar to the accuracy of the genomic prediction with 50K panel (0.360) (Table 5.5).

The prediction accuracies from the transcriptome panel, (Table 5.6) showed a higher increase (0.367 averaged across all the traits) than the averaged accuracy obtained by 50K (Table 5.4). The average gain in accuracy for all the traits with the transcriptome panel over 50K panel was 0.72%. This is while the gain in accuracy with the 124K over 50K panel was 0.57%

5.4. Discussion

Several studies have compared the accuracy of genomic predictions obtained using high-density marker panels (such as Illumina Bovine HD chip) to those from medium-density marker panels (such as Illumina Bovine 50K) (Erbe et al., 2012; Su et al., 2012b; Gao et al., 2013). Increasing the SNP density in genomic selection has the advantage of increasing LD between the SNP markers and the QTL (Harris and Johnson, 2010). An increased level of LD can provide a better QTL signal across and within families (Harris and Johnson, 2010). The results of previous studies, however, showed that the gains in accuracy of genomic predictions in Holstein dairy cattle moving from a lower density to a higher density marker panel were small (Harris and Johnson, 2010; VanRaden et al., 2011; Su et al., 2012a; Gao et al., 2013). For example, in a study carried out by Su et al. (2012), the reliability of direct genomic breeding values based on HD markers (averaged over the traits protein, udder health and fertility) were 0.5% higher than values obtained using 54K data in Nordic Holstein and 1.0% higher in Red dairy cattle. This result was in agreement with a simulation study by VanRaden et al. (2011), which reported gains of 0.9 to 1.2% using an imputed 500K marker genotypes. Similarly, Harris and Johnson (2010) reported a very small gain of 0.8% when moving from 20K to 1,000K in their simulation study. Erbe et al. (2012) also reported a small increase in the accuracy of genomic predictions (by 0.01 averaged across the traits milk, fat and protein yield) when an imputed 800K marker genotypes was used rather than 50K panel. There are several possible reasons for these results. First,

although increasing density of the marker panel has the advantage of increasing the LD, the number of unknown parameters to be estimated also increases. One way to reduce the number of unknown parameters is to reduce the non-informative markers by deleting markers in complete LD with the other marker in the dataset (Su et al., 2012a). In the present study however, removing such variants led to a small increase in the accuracy of genomic prediction (Table 5.2). Su et al. (2012) suggested that it might be necessary to further reduce the redundancy by removing markers that are nearly non-informative. Second, in modern dairy cattle populations, the effective population size is small. Therefore, LD between the potential QTL and SNP marker can be captured sufficiently even with medium density marker panels such as the 50K panel (Erbe et al., 2012). A third reason is that improvements with higher density may require a more sophisticated variable selection methods (rather than linear models) (Erbe et al., 2012; Su et al., 2012a; Gao et al., 2013). For example, Su et al. (2012) showed that a Bayesian mixed model performed slightly better than GBLUP model for HD data (777K). Erbe et al. (2012) and Gao et al. (2013) also reported better genomic predictions using a Bayesian model than GBLUP for high-density marker panels. Another driver behind the lack of improvement in accuracy of predictions could be the underlying genetic structure (Habier et al., 2007; Harris and Johnson, 2010). In dairy cattle populations, factors such as artificial insemination, genetic selection and changes in gene frequency, genomic sampling (drift) and genetic hitchhiking of selection (changes in allele frequency due to linkage disequilibrium with loci that is subjected to selection) are significantly contributing to population stratification and spurious associations (Barton, 2000; Ma and Corl, 2012). Yet another reason is that the additive genetic relationships between individuals can be captured by the markers that are used to estimate marker effects in genomic prediction methods (Fernando, 1998). Thus in related training and validating populations, even if

markers are not in LD with QTL, the accuracy of GEBVs is expected to be non-zero (Fernando, 1998; Habier et al., 2007). This could contribute to the lack of change in accuracy observed when using 26K vs 15K SNP subsets in this study (Table 5.5).

In this work we not only investigated the utility of adding more SNPs—we tested the hypothesis that using a higher density marker panel targeting candidate genes and candidate causal variants can increase the accuracy of estimated genomic breeding values. The accuracy of genomic prediction expressed as $[r(DGVs/TBV)]^2$ when using the imputed Affymetrix 80K was less than that of the 50K, showing a reduction of 0.34% averaged across all traits. When the Affymetrix 80K marker panel was used in conjunction with the 50K (yielding the imputed 124K panel genotypes) there was an increase over 50K alone, but a modest one of 0.57% averaged across all traits (Table 5.4). Thus based on our findings we found little benefit to adding candidate mutations, at least the ones we were able to identify using the sequence data sets and bioinformatics tools described in this work.

One explanation for the disappointing performance of the candidate SNP panel is simply that we did not include many true causal SNPs or SNPs in high LD with those causal SNPs on the new panel. Identifying candidate causal mutations from among the many variants identified in a region is challenging due to incomplete existing knowledge of the roles of genes and regulatory sequences. A further challenge in this study was that the limited sequence information made it likely that some causative mutations were simply not present in the data set. Another possibility is that the 50K panel already contains markers in high LD with many of the causal mutations. Yet another limitation / challenge relates to our reliance on imputation in order to obtain a higher accuracy in genomic predictions. Imputation of SNPs with low MAF is more difficult than SNPs with moderate or high allele frequencies (Erbe et al., 2012), and in our analysis we removed

SNPs with low MAF. Removing SNPs with low MAF can result in a part of the genetic variance being missed and consequently a decrease in accuracies of genomic predictions (Erbe et al., 2012).

The most promising result in our study was obtained from the imputed transcriptome panel genotypes (Table 5.4), which we built removing SNPs that were located in the intergenic, upstream, downstream or intronic regions of the bovine genome. This custom set of SNPs worked well for all the traits and on average led to a better accuracy (0.72%) than what was obtained using the imputed 124K or 50K genotypes. Erbe et al. (2012) performed a similar analysis, comparing the prediction accuracy obtained when using variants in transcribed regions (named as "TRANS" panel with 58,532 SNP) to that obtained when using 50K or imputed 800K genotypes. They showed that the accuracy of prediction was better than imputed 800K when the "TRANS" panel was used. In a study in *Drosophila melanogaster* using SnpEff program, variants in regulatory and coding regions were noted to be the most influential effectors of gene function (Cingolani et al., 2012) and are more likely to have an effect on any trait. In addition, a recent study examining the influence of various SNP types on traits in cattle reported that missense and synonymous variants explain significantly more variation per SNP than the intron, intergenic and non-coding conserved SNPs-(Koufariotis et al., 2014).

Continued studies into gene regulation through, for example, the ongoing FAANG project, as well as the collection of additional sequence information (genomic and expressed sequence), should allow for the construction of better transcriptome panels. In addition, imputation to the level of genome sequence could better ensure that the causative variants are included in the predictions. However, one challenge will continue to be rare causative variants. Yang et al., (2010) reported that cumulative effect of causal variants with small effects or rare alleles with

large effect (low MAF) can explain 45% of the phenotypic variation for human height (Yang et al., 2010). Assuming that the situation is similar for many complex traits, imputation to sequence will need to be done in a way that yields accurate genotypes for these rare markers if prediction accuracies are to be maximized.

5.5. Conclusion

In this study we examined whether the addition of markers from candidate causal genes and regions can improve the accuracy of genomic predictions. The result showed that the custom genotyping panel could not increase in the accuracy of genomic prediction, however, the accuracy of prediction, for a subset of variants that were selected from regions near transcribed regions was promising. This result supports the assumption of the good predictive ability of biologically significant SNPs that are located within or near transcribed regions.

5.6. References

Barton NH (2000) Genetic hitchhiking. Philos Trans R Soc Lond B Biol Sci 355:1553–1562. doi: 10.1098/rstb.2000.0716

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 57:289–300. doi: 10.2307/2346101

Browning BL, Browning SR (2008) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210–223. doi: 10.1016/j.ajhg.2009.01.005

Calvano SE, Xiao W, Richards DR, et al (2005) A network-based analysis of systemic inflammation in humans. Nature 437:1032–1037. doi: 10.1038/nature03985

Cingolani P, Platts A, Wang L, et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118 ; iso-2; iso-3. Landes Biosci 4:74–77. doi: 10.4161/mabs.4.1.18566

Daetwyler HD, Capitan A, Pausch H, et al (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet 46:858–865. doi: 10.1038/ng.3034

Dekkers J (2012) Application of Genomics Tools to Animal Breeding. Curr Genomics 13:207–212. doi: 10.2174/138920212800543057

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of

agricultural populations. Nat Rev Genet 3:22–32. doi: 10.1038/nrg701

Erbe M, Hayes BJ, Matukumalli LK, et al (2012) Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci 95:4114–29. doi: 10.3168/jds.2011-5019

Feng Z, Wong WWL, Gao X, Schenkel F (2011) Generalized genetic association study with samples of related individuals. Ann Appl Stat 5:2109–2130. doi: 10.1214/11-AOAS465

Fernando R. (1998) Genetic evaluation and selection using genotypic, phenotypic and pedigree information. In: Proceedings of the 6th World Congress on Genetics Applied on Livestock Production. Armidale, NSW, Australia, p Vol. 26, 329-336

Ficenec D, Osborne M, Pradines J, et al (2003) Computational knowledge integration in biopharmaceutical research. Brief Bioinform 4:260–278. doi: 10.1093/bib/4.3.260

Frost JM, Moore GE (2010) The importance of imprinting in the human placenta. PLoS Genet 6:1–9. doi: 10.1371/journal.pgen.1001015

Gao H, Su G, Janss L, et al (2013) Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. J Dairy Sci 96:4678–87. doi: 10.3168/jds.2012-6406

Gebert C, Wrenzycki C, Herrmann D, et al (2006) The bovine IGF2 gene is differentially methylated in oocyte and sperm DNA. Genomics 88:222–229. doi: 10.1016/j.ygeno.2006.03.011

Gengler N, Verkenne C, Unit AS (2007) Bayesian Approach to Integrate Molecular Data into

Genetic Evaluations.

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science 298:2345–2349. doi: 10.1126/science.1076641

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10:381–391. doi: 10.1038/nrg2575

Grant JR, Arantes AS, Liao X, Stothard P (2011) In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. Bioinformatics 27:2300–2301. doi: 10.1093/bioinformatics/btr372

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–97. doi: 10.1534/genetics.107.081190

Harris BL, Johnson DL (2010) The impact of high density SNP chips on genomic evaluation in dairy cattle. Interbull Bull 42:40–43.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–43. doi: 10.3168/jds.2008-1646

Horton JD, Goldstein JL, Brown MS (2002) SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. Most 109:1125–1131. doi: 10.1172/JCI200215593.Lipid

Koufariotis L, Chen Y-PP, Bolormaa S, Hayes BJ (2014) Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. BMC Genomics 15:436. doi: 10.1186/1471-2164-15-436

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Littlejohn MD, Henty KM, Tiplady K, et al (2014) Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. Nat Commun 5:5861. doi: 10.1038/ncomms6861

Ma L, Corl B a (2012) Transcriptional regulation of lipid synthesis in bovine mammary epithelial cells by sterol regulatory element binding protein-1. J Dairy Sci 95:3743–55. doi: 10.3168/jds.2011-5083

Matukumalli LK, Lawley CT, Schnabel RD, et al (2009) Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS One. doi: 10.1371/journal.pone.0005350

Meuwissen THE, Goddard ME (2007) Multipoint identity-by-descent prediction using dense markers to map quantitative trait loci and estimate effective population size. Genetics 176:2551–2560. doi: 10.1534/genetics.107.070953

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829.

Nafikov R a., Schoonmaker JP, Korn KT, et al (2014) Polymorphisms in lipogenic genes and milk fatty acid composition in Holstein dairy cattle. Genomics 104:572–581. doi: 10.1016/j.ygeno.2014.10.001

Nafikov R a, Schoonmaker JP, Korn KT, et al (2013) Sterol regulatory element binding

transcription factor 1 (SREBF1) polymorphism and milk fatty acid composition. J Dairy Sci 96:2605–16. doi: 10.3168/jds.2012-6075

Nejati-Javaremi a (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. J Anim Sci 75:1738–1745.

Poland J, Rutoski J (2015) Advances and Challenges in Genomic Selection for Disease Resistance. Annu Rev Phytopathol. doi: 10.1146/annurev-phyto-080615-100056

Pryce JE, Daetwyler HD (2012) Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim Prod Sci 52:107–114. doi: Doi 10.1071/An11098

Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. doi: 10.1086/519795

Raven L-A, Cocks BG, Goddard ME, et al (2014) Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genet Sel Evol 46:29. doi: 10.1186/1297-9686-46-29

Ron M, Weller JI (2007) From QTL to QTN identification in livestock - Winning by points rather than knock-out: A review. Anim Genet 38:429–439. doi: 10.1111/j.1365-2052.2007.01640.x

Sargolzaei M (2012) FImpute User ' s Guide.

Sargolzaei M, Schenkel FS, Vanraden PM (2013) gebv : Genomic breeding value estimator for livestock. L'alliance boviteq, sainthyacinthe, quebec Cent Genet Improv livestock, Univ

guelph, ontario 1–7.

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218–223. doi: 10.1111/j.1439-0388.2006.00595.x

Schork AJ, Thompson WK, Pham P, et al (2013) All SNPs Are Not Created Equal: Genome-Wide Association Studies Reveal a Consistent Pattern of Enrichment among Functionally Annotated SNPs. PLoS Genet. doi: 10.1371/journal.pgen.1003449

Snelling WM, Cushman R a., Keele JW, et al (2013) Networks and pathways to guide genomic selection. J Anim Sci 91:537–552. doi: 10.2527/jas2012-5784

Stothard P, Choi J-W, Basu U, et al (2011) Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics 12:559. doi: 10.1186/1471-2164-12-559

Su G, Brøndum RF, Ma P, et al (2012a) Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. J Dairy Sci 95:4657–65. doi: 10.3168/jds.2012-5379

Su G, Madsen P, Nielsen US, et al (2012b) Genomic prediction for Nordic Red Cattle using one-step and selection index blending. J Dairy Sci 95:909–17. doi: 10.3168/jds.2011-4804

The Bovine Hapmap Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528–532. doi: 10.1126/science.1167936

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–

4423. doi: 10.3168/jds.2007-0980

VanRaden PM, O'Connell JR, Wiggans GR, Weigel K a (2011) Genomic evaluations with many more genotypes. Genet Sel Evol 43:10. doi: 10.1186/1297-9686-43-10

VanRaden PM, Van Tassell CP, Wiggans GR, et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24. doi: 10.3168/jds.2008-1514

Veerkamp RF, Beerda B (2007) Genetics and genomics to improve fertility in high producing dairy cows. Theriogenology 68:266–273. doi: 10.1016/j.theriogenology.2007.04.034

Wang M, Moisá S, Khan MJ, et al (2012) MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation. J Dairy Sci 95:6529–35. doi: 10.3168/jds.2012-5748

Yang J, Goddard ME, Visscher PM (2010) A commentary on "common SNPs explain a large proportion of the heritability for human height" by Yang et al. (2010). Twin Res Hum Genet 13:517–524. doi: 10.1038/ng.608

Table 5. 1: The number of significant SNPs (5% and 1% genome-wise DSR) using generalized quasi-likelihood score (GQLS) study on Bovine 50K and Imputed HD (777K) genotypes in Holstein dairy cattle.

		Bovine 50K		Imputed HD (777 K)	
Trait	Heritability	No. Significant SNPs (1% FDR)	No. Significant SNPs (5% FDR)	No. Significant SNPs (1% FDR)	No. Significant SNPs (5% FDR)
Lifetime profitability Index	0.37	2,359	4,812	15,820	33,188
Milk Production	0.41	3,355	6,155	22,607	42,571
Fat production	0.34	2,983	5,444	20,325	38,310
Protein production	0.37	3,337	6,368	23,055	44,229
Fat deviation	0.37	1,937	3,474	12,981	24,075
Protein deviation	0.37	3,570	5,663	22,879	38,230
Lactation persistency	0.3637	243	734	1,596	4,806
Herd life	0.0975	700	1,659	4,802	11,013
Direct herd life	0.0975	313	749	1,989	4,686
Indirect herd life	0.0975	86	233	598	1,967
Daughter fertility	0.07	279	812	1,885	5,697
Age at first service	0.0942	77	348	576	2,142
Non-return rate (heifer)	0.03	23	95	279	805
First service to conception (heifer)	0.033	99	246	720	1,600
Non-return rate (cow)	0.0395	89	283	708	1,870
Calving to first service interval	0.0715	859	2,193	6,139	14,921
First service to conception (cow)	0.0772	778	2,063	5,473	14,124
Days open	0.1018	1,612	3,556	11,249	24,667

Table 5. 2: Network analysis and associated functions detected for 4,050 positional candidate genes for production and fertility traits.

Associated Network Functions	Score
1. Carbohydrate Metabolism, Molecular Transport, Small Molecule Biochemistry	23
2. Cardiovascular System Development and Function, Cellular Development, Cellular Growth and Proliferation	23
3. Post-Translational Modification, Cell Signaling, Cellular Assembly and Organization	23
4. Hereditary Disorder, Neurological Disease, Psychological Disorders	23
5. RNA Post-Translational Modification, RNA Damage and Repair, Cardiac Enlargement	23

Table 5. 3: Summary of the basic annotation stats from NGS-SNP

All the chromosomes	**Number**
3_prime_UTR_variant	11743
5_prime_UTR_variant	1224
coding_sequence_variant	30
downstream_gene_variant	37495
initiator_codon_variant	31
intergenic_variant	266074
intron_variant	528141
mature_miRNA_variant	16
missense_variant	16155
nc_transcript_variant	2
non_coding_exon_variant	1062
splice_acceptor_variant	6798
splice_donor_variant	2214
splice_region_variant	15155
stop_gained	1157
stop_lost	14
stop_retained_variant	5
synonymous_variant	11049
upstream_gene_variant	22832
Total	921197
fully_known	210189
novel	711008
partially_known	0
Total	921197

Table 5. 4: The squared correlation (r2) between direct genomic breeding values (DGVs) and the corresponding domestic proofs in validation Holstein bull cattle

Trait	Heritability (h^2)	No. Animals G1	No. Animals G2	Panels			
				50K	Imputed 80K	Imputed124K	Transcriptome
Trait01_LPI	0.37	8456	1528	0.408302176	0.394862231	0.407116174	0.411029903
Trait05_MILK	0.41	8851	1528	0.590000356	0.573155542	0.590050448	0.593670123
Trait06_FAT	0.34	8848	1528	0.570435178	0.572358661	0.579900807	0.579979915
Trait07_PROT	0.37	8849	1528	0.51514291	0.48982253	0.512535644	0.51644315
Trait08_FATD	0.37	8848	1528	0.64502109	0.656342315	0.661622635	0.664566068
Trait09_PROTD	0.37	8849	1528	0.647192391	0.655015714	0.665520757	0.666465311
Trait40_LP	0.3637	2164	1528	0.356945052	0.333173156	0.350579899	0.363491175
Trait43_HL	0.0975	8694	1528	0.380721652	0.373802426	0.385544394	0.384916927
Trait44_DHL	0.0975	8169	1528	0.274662224	0.265540543	0.274903901	0.27692122
Trait45_IHL	0.0975	7823	1528	0.32182322	0.333047774	0.338600257	0.33078926
Trait48_DF	0.07	8144	1541	0.221724197	0.223472393	0.228057206	0.229691257
Trait49_AFS	0.0942	2144	1541	0.164152156	0.164960938	0.173587301	0.170440451
Trait50_NRRh	0.03	2449	1541	0.156090048	0.154936295	0.158424726	0.15936246
Trait51_FSTCh	0.033	2144	1541	0.174618646	0.169432616	0.175435854	0.181155125
Trait54_NRRc	0.0395	4440	1541	0.170010329	0.167029561	0.172379142	0.174989896
Trait55_CTFS	0.0715	8246	1541	0.282159799	0.28471646	0.290725804	0.287665322
Trait56_FSTCc	0.0772	8195	1541	0.27881113	0.281876261	0.286668785	0.287894279
Trait59_DO	0.1018	8184	1541	0.33288478	0.33521586	0.341728314	0.342045733
Average of the accuracy for all the traits				0.360594296	0.357153404	0.366299003	0.367862088
Average gain (x100) in accuracy over 50K					-0.3440892	0.570470649	0.726779127

50K. Number of total SNPs in the 50K SNP panel after quality control (44,369 SNPs)

Imputed124K. Number of SNPs in the Imputed124K panel (124,507 SNPs)

Transcriptome. Set of SNPs within and near transcribed regions from Imputed124K panel (74,884 SNPs)

80K. Number of total SNPs in the 80K SNP panel after Imputation and quality control (80,138 SNPs); this panel consisted of all the newly discovered SNPs

No. Animals G1. Number of animals in the estimation group

No. Animals G2. Number of animals in the validation group

Table 5. 5: The squared correlation (r^2) between direct genomic breeding values (DGVs) and the corresponding domestic proofs with different SNP subset in validation Holstein bull cattle

Trait	No. Animals G1	No. Animals G2	50K	Panels			
				Scenario 1	Scenario 2	Scenario 3	Scenario 4
Trait01_LPI	8456	1528	0.408302176	0.40572909	0.400815871	0.400097035	0.391302353
Trait05_MILK	8851	1528	0.590000356	0.591076497	0.583328892	0.586329509	0.576068769
Trait06_FAT	8848	1528	0.570435178	0.575506099	0.558600838	0.571751966	0.535420501
Trait07_PROT	8849	1528	0.51514291	0.513272081	0.506990918	0.508755061	0.50217947
Trait08_FATD	8848	1528	0.64502109	0.658618318	0.624835798	0.651439661	0.602073342
Trait09_PROTD	8849	1528	0.647192391	0.665751471	0.63155435	0.654750595	0.632167238
Trait40_LP	2164	1528	0.356945052	0.35451512	0.348444668	0.351213661	0.336695527
Trait43_HL	8694	1528	0.380721652	0.386438089	0.372754167	0.382476962	0.37775971
Trait44_DHL	8169	1528	0.274662224	0.274737609	0.275894395	0.274920704	0.278465341
Trait45_IHL	7823	1528	0.32182322	0.338081673	0.298698934	0.329887584	0.30404424
Trait48_DF	8144	1541	0.221724197	0.230117435	0.213815705	0.22425554	0.212816027
Trait49_AFS	2144	1541	0.164152156	0.171808431	0.143778511	0.165478564	0.15768305
Trait50_NRRh	2449	1541	0.156090048	0.157140454	0.158075896	0.156237256	0.156638678
Trait51_FSTCh	2144	1541	0.174618646	0.17536077	0.170054193	0.174239794	0.171279296
Trait54_NRRc	4440	1541	0.170010329	0.173928812	0.165332868	0.169246203	0.165076936
Trait55_CTFS	8246	1541	0.282159799	0.288241891	0.274040405	0.277687967	0.271183442
Trait56_FSTCc	8195	1541	0.27881113	0.287368771	0.272655264	0.282603417	0.271851107
Trait59_DO	8184	1541	0.33288478	0.34179414	0.32310146	0.337563313	0.325366915
Average of the accuracy for all the traits			0.360594296	0.366082597	0.351265174	0.361051933	0.348226219
Average gain (x100) in accuracy over 50K				0.535872341	-0.932912229	0.045763658	-1.236807723

50K. Number of total SNPs in the 50K SNP panel after quality control (44,369 SNPs)

Scenario 1. Set of SNPs without redundant SNPs (SNPs that are in high LD ($r^2 > 0.95$)) from Imputed124K panel (108,046 SNPs)

Scenario 2. Set of SNPs (28K) that are 30kb far apart

Scenario 3. Set of SNPs (26K) that are 38kb far apart

Scenario 4. Set of SNPs (15K) that are 89kb far apart

No. Animals G1. Number of animals in the estimation group

No. Animals G2. Number of animals in the validation group

Table 5. 6: Correlations (r) between direct genomic breeding values (DGVs) and the corresponding domestic proofs in validation Holstein bull cattle

Trait	No. Animals G1	No. Animals G2	No. Animals G3	Panels			
				50K	124K	80K	Transcriptome
Trait01_LPI	8456	3111	1528	0.638985271	0.638056561	0.628380642	0.641116138
Trait05_MILK	8851	2717	1528	0.768114806	0.768147413	0.757070368	0.770499917
Trait06_FAT	8848	2720	1528	0.755271592	0.761512184	0.756543892	0.761564124
Trait07_PROT	8849	2719	1528	0.717734568	0.715915947	0.699873224	0.718639792
Trait08_FATD	8848	2720	1528	0.80313205	0.813401891	0.810149563	0.815209217
Trait09_PROTD	8849	2719	1528	0.804482685	0.815794555	0.809330411	0.816373267
Trait40_LP	2164	9357	1528	0.597448786	0.59209788	0.577211535	0.602902293
Trait43_HL	8694	2873	1528	0.61702646	0.620922212	0.611393838	0.620416737
Trait44_DHL	8169	3399	1528	0.524082268	0.524312789	0.515306262	0.526233047
Trait45_IHL	7823	3715	1528	0.56729465	0.581893682	0.577102914	0.575142817
Trait48_DF	8144	3422	1541	0.470875989	0.477553354	0.472728667	0.479261158
Trait49_AFS	2144	9384	1541	0.405156952	0.416638094	0.406153835	0.412844343
Trait50_NRRh	2449	9091	1541	0.39508233	0.398026037	0.39361948	0.39920228
Trait51_FSTCh	2144	9384	1541	0.417873959	0.418850635	0.411621933	0.425623219
Trait54_NRRc	4440	7100	1541	0.412323088	0.415185672	0.408692501	0.418317936
Trait55_CTFS	8246	3320	1541	0.53118716	0.539189951	0.533588287	0.536344406
Trait56_FSTCc	8195	3266	1541	0.528025691	0.535414592	0.530920202	0.536557806
Trait59_DO	8184	3382	1541	0.57696168	0.584575328	0.57897829	0.58484676
Average of the correlation for all the traits				0.585058888	0.589860488	0.582148103	0.591171959

50K. Number of total SNPs in the 50K SNP panel after quality control (44,369 SNPs)

Imputed124K. Number of SNPs in the Imputed124K panel (124,507 SNPs)

80K. Number of total SNPs in the 80K SNP panel after Imputation and quality control (80,138 SNPs); this panel consisted of all the newly discovered SNPs

Transcriptome. Set of SNPs within and near transcribed regions from Imputed124K panel (74,884 SNPs)

No. Animals G1. Number of animals in the estimation group

No. Animals G2. Number of animals in the prediction group

No. Animals G3. Number of animals in the validation group

Figure 5. 1: Population structure identified by first 3 dimensions of a classical multi-dimensional scaling approach. The population is Canadian Holstein cattle.

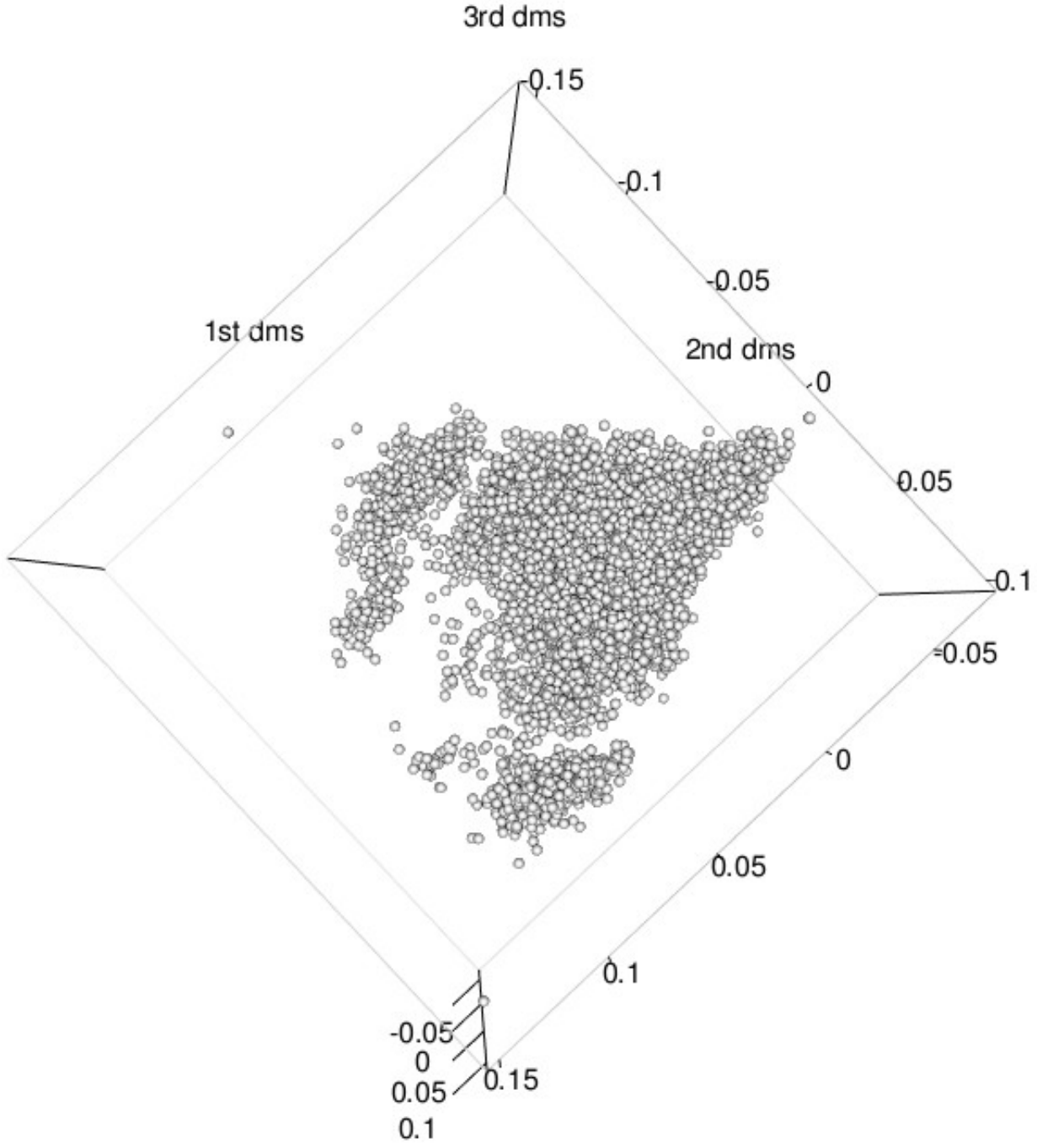
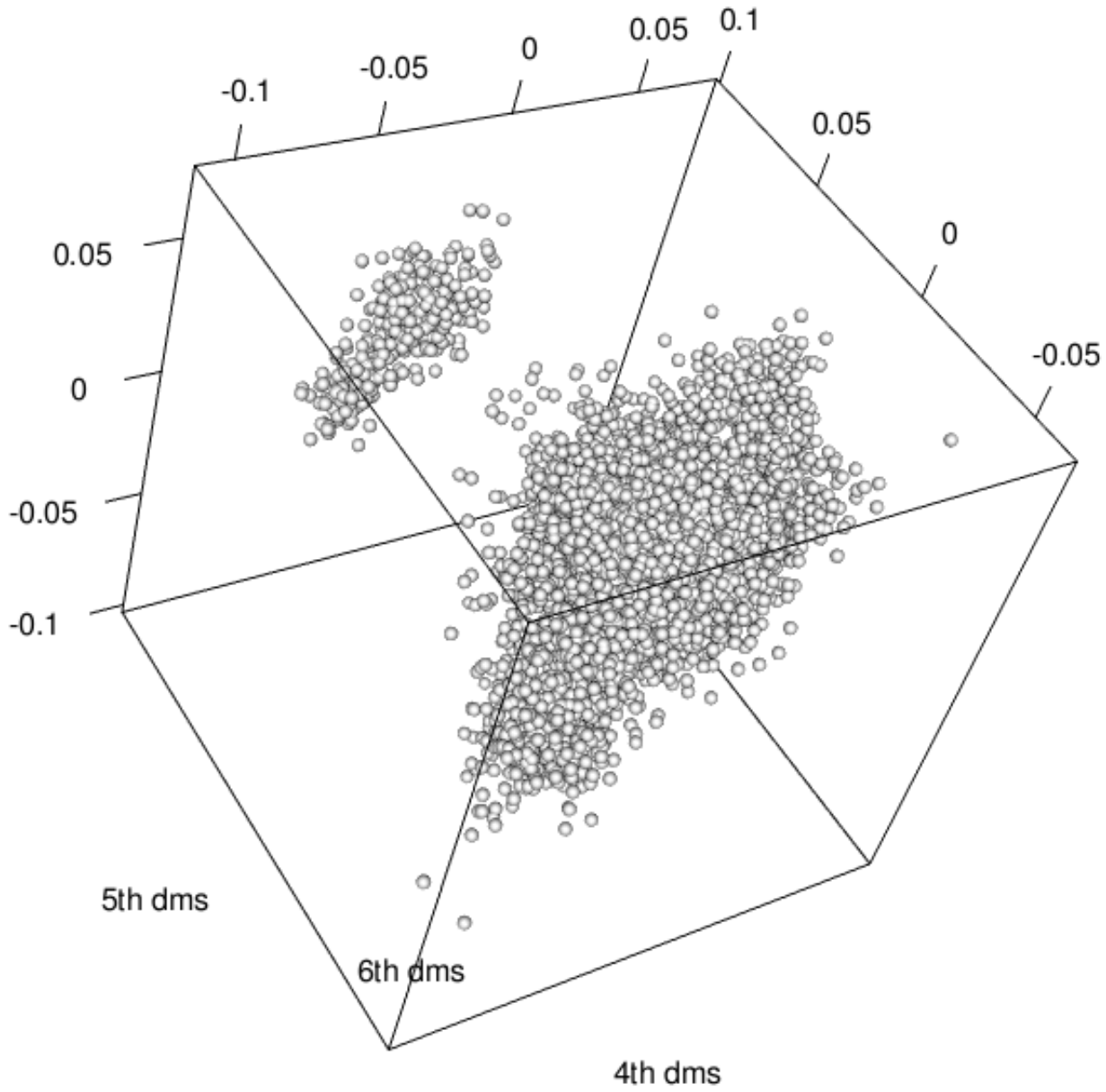


Figure 5. 2: Population structure identified by second 3 dimensions of a classical multi-dimensional scaling approach. The population is Canadian Holstein cattle.



CHAPTER 6: General conclusion

Selection for milk production, milk composition and fertility traits can be performed without consideration of the specific genes and genetic variation involved. However, the application of dense SNP genotyping panels through association analysis will help to identify genomic regions affecting milk production and fertility traits. Knowledge of QTNs or the markers that are in high LD with the causal mutations can allow us to: understand the molecular mechanisms underlying variations in traits; and to develop a smaller DNA marker panel for selection purposes that may be more effective across generations and populations due to a more direct link between genotype and phenotype (Snelling et al., 2013). The purpose of this study was to identify candidate causal mutations and genes for production and fertility traits integrating information from genome-wide association study (GWAS), whole genome sequencing and different bioinformatics tools in Holstein dairy cattle and to develop and evaluate a customized marker panel using new biologically important candidate variants identified in this study.

The objectives of the first study in Chapter 3 were to identify or refine the positions of genomic regions affecting milk production and milk components (milk production, fat production, fat deviation, protein production, protein deviation) and fertility traits (heifer first service to calving interval, daughter fertility, calving to first service interval, and days open) in Canadian Holstein dairy cattle, and to use these positions to identify genes and pathways that may influence these traits. Many of the identified QTL regions for production traits (including milk production) support previous findings and some overlap with genes with known relevant biological functions identified in earlier studies such as *DGAT1* and *CPSF1*. Significant SNPs identified for FAT and FATD were located on chromosomes 5, 6, 14 and 20. A common region (between 87 to 100 Mb) on chromosome 5 in this study for FAT and FATD was reported to be associated with a SNP in

the *GABARAPL1* gene with an antagonistic effect on milk yield and fat percentage in a previous study (Chmuzynska, 2006). Our study also detected strong associations on chromosomes 5, 9, 14 for PROT and 3, 6, 14 and 20 for PROTD. Several highly significant SNPs (FDR ≤ 5%) on BTA6 in this study, associated with PROT and PROTD are located within a known QTL region associated with milk whey protein in dairy cattle (Huang et al., 2012). Association analysis, however, did not detect any significant SNP for daughter fertility or heifer first service to calving interval. The identified chromosomes and significant regions for CTFS were mostly located on BTA13 and 21. The only significant region for DO was on BTA21. The significant region on chromosome 21 was overlapping with the gene *FAM181A* and was not previously reported linked to fertility traits in dairy cattle. A functional enrichment analysis of the GWAS results yielded GO terms associated with the specific phenotypes tested; for example GO terms GO:0043627 (response to estrogen) for milk production and GO:0051057 (positive regulation of small GTPase mediated signal transduction) for fat production, GO:0040019 (positive regulation of embryonic development) for calving to first service interval (CTFS) and GO:0043268 (positive regulation of potassium ion transport) for days open (DO). In other cases the connection between the enriched GO terms and the traits were less clear; for example GO:0003279 (cardiac septum development) for FAT and GO:0030903 (notochord development) for DO trait. The chromosomal regions and enriched pathways identified in this study confirm several previous findings and highlight new regions and pathways that may contribute to variation in production or fertility traits in dairy cattle.

Identifying genomic regions affecting longevity (herd life, indirect herd life, and daughter herd life), female fertility (age at first service, cow first service to conception, heifer non-return rate and cow non-return rate), and lifetime profitability index was carried out in the second study,

described in Chapter 4. For this purpose, GWAS analysis was performed using a single SNP regression mixed linear model implemented in the snp1101 software. The GWAS result identified strong associations for LPI on BTA14 and 18, for LP on BTA20 and for IHL, HL and DHL on BTA18. However, no significant association was found for AFS, FSTCc, NRRc and NRRh. Previously proposed causative and candidate genes supported by this work include *DGAT1, GRINA, CPSF1* for lactation persistency. The significant region on BTA6 identified for LP at 88 Mb in this study was reported to be associated with persistency of milk yield and milk composition traits in German Holstein cattle; this region was also shown to overlap with a QTL identified associated with somatic cell score in Holstein cattle. The suggested candidate gene for this region is *SCLA4A4*, which was shown to be related to mastitis susceptibility in Danish Holstein cattle. This result suggests that significant SNPs on BTA6 might affect udder disease susceptibility and subsequently lactation persistency in dairy cattle. A novel region for lactation persistency in this study is speculated on BTA13 at 54 to 55 Mb. The SNPs in this region are located within the genes *MYT1*, *SLC2A4RG*, *SLC17A9*, *LAMA5*, *ADRM1*, *OSBPL2* and *SS18L1*. A previous study reported that expression of the gene *OSBPL2* in the mammary gland increases along with *SREBF1* and *SREBF*2 genes during lactation (Bionaz and Loor, 2008). We also identified another region in our study on BTA27 associated with lactation persistency. Significant SNPs in this region were previously reported to affect expression of *AGPAT6* gene and were involved in milk fat synthesis. The associated significant SNPs within this gene were not significant in our study; however, our result detected five significant SNPs on BTA27 at 41 Mb which is speculated to be a new region for lactation persistency. One significant SNP in this region is assigned to the gene thyroid hormone receptor, beta (*THRB*). This gene encodes a nuclear hormone receptor for triiodothyronine and mutations in this gene are known to be the

cause of generalized thyroid hormone resistance (GTHR). It has also been shown that there is a positive correlation between thyroid hormones in blood and energy metabolism (Reist et al., 2002). This proposed gene might be considered as a new potential candidate gene affecting production traits in the dairy cattle.

The GWAS analysis for herd life, indirect herd life and direct herd life (DHL) identified significant SNPs on several chromosomes. The highest number of significant SNPs for HL is detected on BTA6 and 18. Significant SNPs for IHL were mostly located on BTA5, 6 and 18. The overlapping region on BTA18 for HL, IHL and DHL was spanning at 42 to 65 Mb. This region was reported to be associated with one of the fertility-associated traits, calving ease, in a previous study. Fertility and postpartum performance of replacement heifers are important traits to minimize the non-productive period of the animal's life (Akanno et al., 2015). The SNPs identified in this study for BTA6 (for both herd life and indirect herd life) are located at 88 Mb. This region was reported associated with somatic cell score, milk yield and body condition score in Holstein cattle breed. Most of the significant SNPs that we found for longevity-associated traits were located within the QTL regions that have been previously reported to affect calving traits (calving ease, calf size, stillbirth), clinical mastitis, milk yield and conformation traits in Holstein and other cattle breeds. According to the fact that herd life evaluations are based on a combination of fertility, conformation and udder health traits, it is very possible that these regions and associated candidate genes, particularly on BTA6 and 18 influence this trait. Taken together with these previous results, our findings indicate that these regions and significant SNPs may potentially have an effect on longevity traits in dairy cattle and may be used for further genomic prediction analysis to identify causal candidate variants that affect herd life traits.

The objective of the third study, described in Chapter 5, was to investigate the possibility of increasing the accuracy of genomic predictions by using a higher density marker panel targeting candidate genes and potential causal variants. For this purpose a custom Affymetrix panel was developed, containing 135K (135,256) markers identified through GWAS, candidate gene and RNA-sequencing studies. Three subsets of SNPs were then evaluated in terms of genomic breeding value accuracy for a variety of traits: the new Affymetrix panel (termed the 80K after quality control), the Affymetrix panel combined with the 50K (the 124K panel), and the subset of SNPs from the combined set within transcribed regions (transcriptome SNP panel).

The results indicated that the accuracy of genomic predictions using the custom 80K panel dropped by 0.34% (averaged across all the traits) when compared to 50K panel whereas when using the custom genotyping panel in conjunction with the 50K (124K genotypes), prediction accuracy was slightly increased (0.57%) over 50K panel, averaged across all the traits. The most promising result in the current study was obtained from a set of SNP genotypes that are located in transcribed regions. This subset of markers from the 124K SNP panel, consisting of 74K markers, worked well for all the traits and on average led to an accuracy of 0.72% over 50K genotypes. In conclusion, these results indicated that the inclusion of candidate QTN did little to increase prediction accuracy but that further efforts aimed at using genotypes from gene regions are warranted.

6.1. Limitations and general recommendations

The genome-wide association study for milk production, longevity, fertility and lifetime profitability index in our study resulted in identifying many significant regions previously reported in other independent studies. Our study could also identify several novel regions

associated with calving to first service interval, days open (on chromosome 21) and lactation persistency (on chromosome 27). According to our findings, there was a little improvement in the accuracy of genomic predictions by including candidate variants to the custom genotyping panel. One of the limitations of this study was our challenge in identifying these causal variants due to incomplete knowledge of gene function and variants, for example regulatory sequences, which affect the traits under investigation. Another challenge was the limited number of sequenced animals in our data set, which made the list of variants considered for inclusion in the panel far from comprehensive. Notably, when a subset of SNP genotypes was selected consisting of variants near transcribed regions, a greater gain in accuracy was observed than when all the candidate variants were used. This result suggests that further efforts to identify biologically relevant variants are warranted. Continued studies of gene expression and gene regulatory regions in cattle will be helpful in this regard by allowing for more accurate predictions of variant function. In addition, imputation to whole-genome sequence for genomic prediction might result in greater accuracy provided that causal variants can be imputed accurately. Together these efforts stand to increase the rate of genetic gain achieved through the application of genomic selection in dairy cattle, and could contribute to a better understanding of the molecular basis of traits.

6.2. References

Akanno EC, Plastow G, Fitzsimmons C, et al (2015) Genome-wide association for heifer reproduction and calf performance traits in beef cattle. Genome 557:1–9. doi: 10.1139/gen-2015-0031

Bionaz M, Loor JJ (2008) Gene networks driving bovine milk fat synthesis during the lactation cycle. 21:1–21. doi: 10.1186/1471-2164-9-366

Chmuzynska (2006) The multigene family of fatty acid-binding proteins (FABPs): Function, structure and polymorphism. J Appl Genet 47:39–48.

Huang W, Peñagaricano F, Ahmad KR, et al (2012) Association between milk protein gene variants and protein composition traits in dairy cattle. J Dairy Sci 95:440–9. doi: 10.3168/jds.2011-4757

Reist M, Erdin D, von Euw D, et al (2002) Estimation of Energy Balance at the Individual and Herd Level Using Blood and Milk Traits in High-Yielding Dairy Cows,. J Dairy Sci 85:3314–3327. doi: 10.3168/jds.S0022-0302(02)74420-2

Snelling WM, Cushman R a., Keele JW, et al (2013) Networks and pathways to guide genomic selection. J Anim Sci 91:537–552. doi: 10.2527/jas2012-5784

Bibliography

Abdel-Shafy H, Bortfeldt RH, Reissmann M, Brockmann GA (2014) Short communication: Validation of somatic cell score-associated loci identified in a genome-wide association study in German Holstein cattle. J Dairy Sci 97:2481–6. doi: 10.3168/jds.2013-7149

Ahmadzadeh A, Barnes M, Pearson R. (1998) Effect of Naloxone on Serum Luteinizing Hormone Concentration in Anovulatory Holstein Cows During the Early Postpartum Period. Domest Anim Endocrinol 15:177–181. doi: 10.1016/S0739-7240(98)00005-8

Akanno EC, Plastow G, Fitzsimmons C, et al (2015) Genome-wide association for heifer reproduction and calf performance traits in beef cattle. Genome 557:1–9. doi: 10.1139/gen-2015-0031

Appuhamy JADRN, Cassell BG, Cole JB (2009) Phenotypic and genetic relationships of common health disorders with milk and fat yield persistencies from producer-recorded health data and test-day yields. J Dairy Sci 92:1785–1795. doi: 10.3168/jds.2008-1591

Arranz JJ, Coppieters W, Berzi P, et al (1998) A QTL affecting milk yield and composition maps to bovine chromosome 20: A confirmation. Anim Genet 29:107–115. doi: 10.1046/j.1365-2052.1998.00307.x

Ashwell MS, Heyen DW, Sonstegard TS, et al (2004a) Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. J Dairy Sci 87:468–75. doi: 10.3168/jds.S0022-0302(04)73186-0

Ashwell MS, Heyen DW, Sonstegard TS, et al (2004b) Detection of quantitative trait loci

affecting milk production, health, and reproductive traits in Holstein cattle. J Dairy Sci 87:468–75. doi: 10.3168/jds.S0022-0302(04)73186-0

Baik M, Etchebarne BE, Bong J, VandeHaar MJ (2009) Gene expression profiling of liver and mammary tissues of lactating dairy cows. Asian-Australasian J Anim Sci 22:871–884.

Barton NH (2000) Genetic hitchhiking. Philos Trans R Soc Lond B Biol Sci 355:1553–1562. doi: 10.1098/rstb.2000.0716

Bauman DE, Currie WB (1980) Partitioning of nutrients during pregnancy and lactation: a review of mechanisms involving homeostasis and homeorhesis. J Dairy Sci 63:1514–1529. doi: 10.3168/jds.S0022-0302(80)83111-0

Bauman DE, Griinari JM (2001) Regulation and nutritional manipulation of milk fat: Low-fat milk syndrome. Livest Prod Sci 70:15–29. doi: 10.1016/S0301-6226(01)00195-6

Bauman DE, Griinari JM (2003) Nutritional regulation of milk fat synthesis. Annu Rev Nutr 23:203–227. doi: 10.1146/annurev.nutr.23.011702.073408

Bauman DE, Mather IH, Wall RJ, Lock AL (2006) Major advances associated with the biosynthesis of milk. J Dairy Sci 89:1235–1243. doi: 10.3168/jds.S0022-0302(06)72192-0

Beigneux AP, Vergnes L, Qiao X, et al (2006) Agpat6—a Novel Lipid Biosynthetic Gene Required for Triacylglycerol Production in Mammary Epithelium. 47:734–744. doi: 10.1021/ja8019214.Optimization

Bell AW (1995) Regulation of Organic Nutrient Late Pregnancy Metabolism During Transition from to Early Lactation. J Anim Sci 73:2804–2819. doi: /1995.7392804x

Bell AW, Bauman DE (1997) Adaptations of glucose metabolism during pregnancy and lactation. J Mammary Gland Biol Neoplasia 2:265–278.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 57:289–300. doi: 10.2307/2346101

Berg J., Tymoczko J., Stryer L (2002) Biochemistry, 5th edition.

Bernard L, Leroux C, Chilliard Y (2008) Bioactive Components of Milk. Springer Science+BusinessMedia, Inc. New York, USA

Bionaz M, Loor JJ (2008a) Gene networks driving bovine milk fat synthesis during the lactation cycle. 21:1–21. doi: 10.1186/1471-2164-9-366

Bionaz M, Loor JJ (2008b) ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J Nutr 138:1019–1024. doi: 138/6/1019 [pii]

Blaschek M, Kaya a, Zwald N, et al (2011) A whole-genome association analysis of noncompensatory fertility in Holstein bulls. J Dairy Sci 94:4695–4699. doi: 10.3168/jds.2010-3728

Block SS, Butler WR, Ehrhardt RA, et al (2001) Decreased concentration of plasma leptin in periparturient dairy cows is caused by negative energy balance. J Endocrinol 171:339–348. doi: 10.1677/joe.0.1710339

Blott S, Kim J, Moisio S, et al (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth

hormone receptor is associated with a major effect on milk yield and composition. Genetics 163:253–66.

Blum JW, Kunz P, Leuenberger H, et al (1983) Thyroid hormones, blood plasma metabolites and haematological parameters in relationship to milk yield in dairy cows. Anim Prod 36:93–104. doi: 10.1017/S0003356100039982

Boichard D (1990) Estimation of the economic value of conception rate in dairy cattle. Livest Prod Sci 24:187–204. doi: 10.1016/0301-6226(90)90001-M

Boichard DB, Grohs C, Bourgeois F, et al (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. Genet Sel Evol 35:77–101. doi: 10.1051/gse

Bouwman AC, Bovenhuis H, Visker MH, van Arendonk JA (2011) Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet 12:43. doi: 10.1186/1471-2156-12-43

Bovenhuis H, Van Arendonk J a, Korver S (1992) Associations between milk protein polymorphisms and milk production traits. J Dairy Sci 75:2549–2559. doi: 10.3168/jds.S0022-0302(92)78017-5

Brand B, Baes C, Mayer M, et al (2009) Identification of a two-marker-haplotype on Bos taurus autosome 18 associated with somatic cell score in German Holstein cattle. BMC Genet 10:50. doi: 10.1186/1471-2156-10-50

Brinkman-Van Der linden ECM, Hurtado-Ziola N, Hayakawa T, et al (2007) Human-specific expression of Siglec-6 in the placenta. Glycobiology 17:922–931. doi: 10.1093/glycob/cwm065

Browning BL, Browning SR (2008) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210–223. doi: 10.1016/j.ajhg.2009.01.005

Buitenhuis B, Janss LL, Poulsen NA, et al (2014) Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15:1112. doi: 10.1186/1471-2164-15-1112

Burgos SA, Cant JP (2010) IGF-1 stimulates protein synthesis by enhanced signaling through mTORC1 in bovine mammary epithelial cells. Domest Anim Endocrinol 38:211–221. doi: 10.1016/j.domaniend.2009.10.005

Calvano SE, Xiao W, Richards DR, et al (2005) A network-based analysis of systemic inflammation in humans. Nature 437:1032–1037. doi: 10.1038/nature03985

Carlson DB, McFadden JW, D'Angelo A, et al (2007) Dietary l-Carnitine Affects Periparturient Nutrient Metabolism and Lactation in Multiparous Cows. J Dairy Sci 90:3422–3441. doi: 10.3168/jds.2006-811

Cerri RLA, Thompson IM, Kim IH, et al (2012) Effects of lactation and pregnancy on gene expression of endometrium of Holstein cows at day 17 of the estrous cycle or pregnancy. J Dairy Sci 95:5657–75. doi: 10.3168/jds.2011-5114

Chamberlain a J, Hayes BJ, Savin K, et al (2012) Validation of single nucleotide polymorphisms associated with milk production traits in dairy cattle. J Dairy Sci 95:864–75. doi: 10.3168/jds.2010-3786

Chatterjee S, Szustakowski JD, Nanguneri NR, et al (2009) Identification of Novel Genes and Pathways Regulating SREBP Transcriptional Activity. PLoS One 4:e5197. doi: 10.1371/journal.pone.0005197

Chmuzynska (2006) The multigene family of fatty acid-binding proteins (FABPs): Function, structure and polymorphism. J Appl Genet 47:39–48.

Cingolani P, Platts A, Wang L, et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118 ; iso-2; iso-3. Landes Biosci 4:74–77. doi: 10.4161/mabs.4.1.18566

Cochran SD, Cole JB, Null DJ, Hansen PJ (2013) Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genet 14:49. doi: 10.1186/1471-2156-14-49

Cohen-Zinder M, Seroussi E, Larkin DM, et al (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res 15:936–944. doi: 10.1101/gr.3806705

Cole JB, Null DJ (2009) Genetic evaluation of lactation persistency for five breeds of dairy cattle. J Dairy Sci 92:2248–58. doi: 10.3168/jds.2008-1825

Cole JB, VanRaden PM, O'Connell JR, et al (2009) Distribution and location of genetic effects for dairy traits. J Dairy Sci 92:2931–2946. doi: 10.3168/jds.2008-1762

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

Coleman RA, Lewin TM, Van Horn CG, Gonzalez-Baró MR (2002) Do long-chain acyl-CoA synthetases regulate fatty acid entry into synthetic versus degradative pathways? J Nutr 132:2123–2126.

Conte G, Mele M, Chessa S, et al (2010) Diacylglycerol acyltransferase 1, stearoyl-CoA desaturase 1, and sterol regulatory element binding protein 1 gene polymorphisms and milk fatty acid composition in Italian Brown cattle. J Dairy Sci 93:753–63. doi: 10.3168/jds.2009-2581

Curley M, Cairns MT, Friel AM, et al (2002) Expression of mRNA transcripts for ATP-sensitive potassium channels in human myometrium. Mol Hum Reprod 8:941–945.

Daetwyler HD, Capitan A, Pausch H, et al (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet 46:858–865. doi: 10.1038/ng.3034

Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JAB (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. J Dairy Sci 91:3225–36. doi: 10.3168/jds.2007-0333

Dann HM, Drackley JK (2005) Carnitine palmitoyltransferase I in liver of periparturient dairy cows: effects of prepartum intake, postpartum induction of ketosis, and periparturient

disorders. J Dairy Sci 88:3851–3859. doi: 10.3168/jds.S0022-0302(05)73070-8

Dekkers J (2012) Application of Genomics Tools to Animal Breeding. Curr Genomics 13:207–212. doi: 10.2174/138920212800543057

Dekkers JC (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci. doi: 82/13_suppl/E313 [pii]

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3:22–32. doi: 10.1038/nrg701

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004. doi: 10.1111/j.0006-341X.1999.00997.x

Dillon P, Berry DP, Evans RD, et al (2006) Consequences of genetic selection for increased milk production in European seasonal pasture based systems of milk production. Livest Sci 99:141–158. doi: 10.1016/j.livprodsci.2005.06.011

Djoković R, Cincović M, Kurćubić V, et al (2014) Endocrine and Metabolic Status of Dairy Cows during Transition Period. Thai J Vet Med Thai J Vet Med 44:59–66.

Djoković R, Šamanc H, Jovanović M, Nikolić Z (2007) Blood Concentrations of Thyroid Hormones and Lipids and Content of Lipids in the Liver of Dairy Cows in Transitional Period. Acta Vet Brno 76:525–532. doi: 10.2754/avb200776040525

Doran AG, Berry DP, Creevey CJ (2014) Whole genome association study identifies regions of the bovine genome and biological pathways involved in carcass trait performance in Holstein-Friesian cattle. BMC Genomics 15:837. doi: 10.1186/1471-2164-15-837

Drackley JK (1999) Biology of dairy cows during the transition period: the final frontier? J Dairy Sci 82:2259–2273. doi: 10.3168/jds.S0022-0302(99)75474-3

Duplus E, Glorian M, Forest C (2000) Fatty acid regulation of gene transcription. J Biol Chem 275:30749–30752. doi: 10.1074/jbc.R000015200

Egger-Danner C, Cole JB, Pryce JE, et al (2014) Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal 9:191–207. doi: 10.1017/S1751731114002614

Erbe M, Hayes BJ, Matukumalli LK, et al (2012) Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci 95:4114–29. doi: 10.3168/jds.2011-5019

Etherton TD, Bauman DE (1998) Biology of somatotropin in growth and lactation of domestic animals. Physiol Rev 78:745–761.

Fang M, Fu W, Jiang D, et al (2014) A Multiple-SNP Approach for Genome-Wide Association Study of Milk Production Traits in Chinese Holstein Cattle. PLoS One 9:e99544. doi: 10.1371/journal.pone.0099544

Fazioli F, Minichiello L, Matoska V, et al (1993) Eps8, a substrate for the epidermal growth factor receptor kinase, enhances EGF-dependent mitogenic signals. EMBO J 12:3799–808.

Feng Z, Wong WWL, Gao X, Schenkel F (2011) Generalized genetic association study with samples of related individuals. Ann Appl Stat 5:2109–2130. doi: 10.1214/11-AOAS465

Fenwick MA, Llewellyn S, Fitzpatrick R, et al (2008) Negative energy balance in dairy cows is

associated with specific changes in IGF-binding protein expression in the oviduct. Reproduction 135:63–75. doi: 10.1530/REP-07-0243

Fernando R. (1998) Genetic evaluation and selection using genotypic, phenotypic and pedigree information. In: Proceedings of the 6th World Congress on Genetics Applied on Livestock Production. Armidale, NSW, Australia, p Vol. 26, 329-336

Ficenec D, Osborne M, Pradines J, et al (2003) Computational knowledge integration in biopharmaceutical research. Brief Bioinform 4:260–278. doi: 10.1093/bib/4.3.260

Fielding BA, Frayn KN (1998) Lipoprotein lipase and the disposition of dietary fatty acids. Br J Nutr 80:495–502. doi: 10211047

Flori L, Fritz S, Jaffrézic F, et al (2009) The genome response to artificial selection: a case study in dairy cattle. PLoS One 4:e6595. doi: 10.1371/journal.pone.0006595

Fortes MRS, Deatley KL, Lehnert SA, et al (2013) Genomic regions associated with fertility traits in male and female cattle: advances from microsatellites to high-density chips and beyond. Anim Reprod Sci 141:1–19. doi: 10.1016/j.anireprosci.2013.07.002

Freyer G, Sørensen P, Kühn C, et al (2003) Search for pleiotropic QTL on chromosome BTA6 affecting yield traits of milk production. J Dairy Sci 86:999–1008. doi: 10.3168/jds.S0022-0302(03)73683-2

Frolov A, Cho TH, Murphy EJ, Schroeder F (1997) Isoforms of rat liver fatty acid binding protein differ in structure and affinity for fatty acids and fatty acyl CoAs. Biochemistry 36:6545–6555. doi: 10.1021/bi970205t

Frost JM, Moore GE (2010) The importance of imprinting in the human placenta. PLoS Genet 6:1–9. doi: 10.1371/journal.pgen.1001015

Gao H, Su G, Janss L, et al (2013) Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. J Dairy Sci 96:4678–87. doi: 10.3168/jds.2012-6406

Garcia MD, Michal JJ, Gaskins CT, et al (2006) Significant association of the calpastatin gene with fertility and longevity in dairy cattle. Anim Genet 37:304–305. doi: 10.1111/j.1365-2052.2006.01443.x

Gebert C, Wrenzycki C, Herrmann D, et al (2006) The bovine IGF2 gene is differentially methylated in oocyte and sperm DNA. Genomics 88:222–229. doi: 10.1016/j.ygeno.2006.03.011

Gengler N, Verkenne C, Unit AS (2007) Bayesian Approach to Integrate Molecular Data into Genetic Evaluations.

Georges M, Georges M, Nielsen D, et al (1995) Mapping Quantitative Trait Loci Controlling Milk Production in Dairy Cattle by Exploiting Progeny Testing. Genet Soc Am 139:907–920.

Gibbs R, Taylor JF, Van Tassell CP, et al (2012) NIH Public Access. Science (80-) 29:997–1003. doi: 10.1016/j.biotechadv.2011.08.021.Secreted

Gimeno RE, Cao J (2008) Thematic review series: glycerolipids. Mammalian glycerol-3-phosphate acyltransferases: new genes for an old activity. J Lipid Res 49:2079–2088. doi:

10.1194/jlr.R800013-JLR200

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science 298:2345–2349. doi: 10.1126/science.1076641

Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet 124:323–330.

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10:381–391. doi: 10.1038/nrg2575

Goddard ME, MacLeod L., Kemper K., et al (2014) A Research Plan for the Identification of QTL. In: Proceeding, 10th World Congress of Genetics Applied to Livestock Production.

Goff JP, Horst RL (1997) Physiological changes at parturition and their relationship to metabolic disorders. J Dairy Sci 80:1260–1268. doi: 10.3168/jds.S0022-0302(97)76055-7

Graber M, Kohler S, Müller A, et al (2012) Identification of plasma and hepatic parameters related to metabolic robustness in dairy cows. J Anim Physiol Anim Nutr (Berl) 96:75–84. doi: 10.1111/j.1439-0396.2010.01124.x

Grant JR, Arantes AS, Liao X, Stothard P (2011) In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. Bioinformatics 27:2300–2301. doi: 10.1093/bioinformatics/btr372

Grisart B, Coppieters W, Farnir F, et al (2002) Positional candidate cloning of a QTL in dairy cattle: identi cation of a missense mutation in the bovine DGAT gene with major effect on milk yield and composition. Genome Res 12:222–231. doi: 10.1101/gr.224202

Grisart B, Farnir F, Karim L, et al (2004) Genetic and functional confirmation of the causality of

the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc Natl Acad Sci U S A 101:2398–403.

Groen AF, Steine T, Colleau J-J, et al (1997) Economic values in dairy cattle breeding, with special reference to functional traits. Report of an EAAP-working group. Livest Prod Sci 49:1–21. doi: 10.1016/S0301-6226(97)00041-9

Gross JJ, Kessler EC, Albrecht C, Bruckmaier RM (2015) Response of the Cholesterol Metabolism to a Negative Energy Balance in Dairy Cows Depends on the Lactational Stage. PLoS One 10:e0121956. doi: 10.1371/journal.pone.0121956

Grummer RR (2007) Strategies to improve fertility of high yielding dairy farms: Management of the dry period. Theriogenology 68:281–288. doi: 10.1016/j.theriogenology.2007.04.031

Guo J, Jorjani H, Carlborg Ö (2012a) 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. doi: 10.1186/1471-2156-13-82

Guo J, Jorjani H, Carlborg Ö (2012b) 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. doi: 10.1186/1471-2156-13-82

Ha N-T, Gross JJ, van Dorland A, et al (2015) Gene-Based Mapping and Pathway Analysis of Metabolic Traits in Dairy Cows. PLoS One 10:e0122325. doi: 10.1371/journal.pone.0122325

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on

genome-assisted breeding values. Genetics 177:2389–97. doi: 10.1534/genetics.107.081190

Haile-Mariam M, Morton JM, Goddard ME (2003) Estimates of genetic parameters for fertility traits of Australian Holstein-Friesian cattle. Anim Sci 76:35–42.

Han J, Kraft P, Nan H, et al (2008) A Genome-Wide Association Study Identifies Novel Alleles Associated with Hair Color and Skin Pigmentation. PLoS Genet 4:e1000074. doi: 10.1371/journal.pgen.1000074

Hanson RW, Reshef L (2003) Glyceroneogenesis revisited. Biochimie 85:1199–1205. doi: 10.1016/j.biochi.2003.10.022

Harder B, Bennewitz J, Reinsch N, et al (2006) Mapping of quantitative trait loci for lactation persistency traits in German Holstein dairy cattle. J Anim Breed Genet 123:89–96. doi: 10.1111/j.1439-0388.2006.00577.x

Harris BL, Johnson DL (2010) The impact of high density SNP chips on genomic evaluation in dairy cattle. Interbull Bull 42:40–43.

Hawken RJ, Zhang YD, Fortes MRS, et al (2012) Genome - wide association studies of female reproduction in tropically adapted beef cattle. J Anim Sci 90:1398–1410. doi: 10.2527/jas2011-4410

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–43. doi: 10.3168/jds.2008-1646

Heid HW, Keenan TW (2005) Intracellular origin and secretion of milk fat globules. Eur J Cell Biol 84:245–258. doi: 10.1016/j.ejcb.2004.12.002

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108. doi: 10.1038/nrg1521

Höglund JK, Buitenhuis AJ, Guldbrandtsen B, et al (2009a) Overlapping chromosomal regions for fertility traits and production traits in the Danish Holstein population. J Dairy Sci 92:5712–9. doi: 10.3168/jds.2008-1964

Höglund JK, Buitenhuis B, Guldbrandtsen B, et al (2015) Genome-wide association study for female fertility in Nordic Red cattle. BMC Genet 16:110. doi: 10.1186/s12863-015-0269-x

Höglund JK, Guldbrandtsen B, Lund MS, Sahana G (2012) Analyzes of genome-wide association follow-up study for calving traits in dairy cattle. BMC Genet 13:71. doi: 10.1186/1471-2156-13-71

Höglund JK, Guldbrandtsen B, Su G, et al (2009b) Genome scan detects quantitative trait loci affecting female fertility traits in Danish and Swedish Holstein cattle. J Dairy Sci 92:2136–43. doi: 10.3168/jds.2008-1104

Höglund JK, Sahana G, Brøndum RF, et al (2014a) Fine mapping QTL for female fertility on BTA04 and BTA13 in dairy cattle using HD SNP and sequence data. BMC Genomics 15:790. doi: 10.1186/1471-2164-15-790

Höglund JK, Sahana G, Guldbrandtsen B, Lund MS (2014b) Validation of associations for female fertility traits in Nordic Holstein, Nordic Red and Jersey dairy cattle. BMC Genet 15:8. doi: 10.1186/1471-2156-15-8

Holmberg M, Andersson-Eklund L (2006) Quantitative trait loci affecting fertility and calving

traits in Swedish dairy cattle. J Dairy Sci 89:3664–71. doi: 10.3168/jds.S0022-0302(06)72406-7

Horton JD, Goldstein JL, Brown MS (2002) SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. Most 109:1125–1131. doi: 10.1172/JCI200215593.Lipid

Horton JD, Shimomura I, Brown MS, et al (1998) Activation of cholesterol synthesis in preference to fatty acid synthesis in liver and adipose tissue of transgenic mice overproducing sterol regulatory element-binding protein-2. J Clin Invest 101:2331–2339. doi: 10.1172/JCI2961

Houshdaran S, Zelenko Z, Irwin JC, Giudice LC (2014) Human Endometrial DNA Methylome Is Cycle-Dependent and Is Associated With Gene Expression Regulation. Mol Endocrinol 28:1118–1135. doi: 10.1210/me.2013-1340

Hsieh C-W, Huang C, Bederman I, et al (2011) Function of phosphoenolpyruvate carboxykinase in mammary gland epithelial cells. J Lipid Res 52:1352–62. doi: 10.1194/jlr.M012666

Hu J, Zhang Z, Shen W-J, Azhar S (2010) Cellular cholesterol delivery, intracellular processing and utilization for biosynthesis of steroid hormones. Nutr Metab (Lond) 7:47. doi: 10.1186/1743-7075-7-47

Huang W, Peñagaricano F, Ahmad KR, et al (2012) Association between milk protein gene variants and protein composition traits in dairy cattle. J Dairy Sci 95:440–9. doi: 10.3168/jds.2011-4757

Hunt CR, Ro JH, Dobson DE, et al (1986) Adipocyte P2 gene: developmental expression and homology of 5'-flanking sequences among fat cell-specific genes. Proc Natl Acad Sci U S A 83:3786–90. doi: 10.1073/pnas.83.11.3786

Iso-Touru T, Sahana G, Guldbrandtsen B, et al (2016) Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. BMC Genet 17:55. doi: 10.1186/s12863-016-0363-8

Jiang L, Liu J, Sun D, et al (2010) Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5:e13661. doi: 10.1371/journal.pone.0013661

Jiang L, Liu X, Yang J, et al (2014) Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. BMC Genomics 15:1105. doi: 10.1186/1471-2164-15-1105

Jorritsma R, Wensing T, A.M. K, et al (2003) Metabolic changes in early lactation and impaired reproductive performance in diary cows. Vet Res 34:11–26. doi: 10.1051/vetres

Jump DB, Botolin D, Wang Y, et al (2005) Fatty Acid Regulation of Hepatic. J Nutr 135:2503–2506.

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

Kasagic D, Radojicic B, Gvozdic D, et al (2011) Endocrine and metabolic profile in Holstein and red Holstein heifers during peripartal period. Acta Vet Brno 61:555–565. doi: 10.2298/AVB1106555K

Kawaguchi T, Osatomi K, Yamashita H, et al (2002) Mechanism for fatty acid "sparing" effect on glucose-induced transcription: Regulation of carbohydrate-responsive element-binding protein by AMP-activated protein kinase. J Biol Chem 277:3829–3835. doi: 10.1074/jbc.M107895200

Kay JK, Weber WJ, Moore CE, et al (2005) Effects of week of lactation and genetic selection for milk yield on milk fatty acid composition in Holstein cows. J Dairy Sci 88:3886–3893. doi: 10.3168/jds.S0022-0302(05)73074-5

Keenan TW, Mather IH (2009) Intracellular origin of milk fat globules and the nature of the milk fat globule membrane. Adv Dairy Chem 2:137–171. doi: 10.1007/0-387-28813-9_4

Kemper KE, Reich CM, Bowman PJ, et al (2015) Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. Genet Sel Evol 47:1–17. doi: 10.1186/s12711-014-0074-4

Kessler EC, Gross JJ, Bruckmaier RM, Albrecht C (2014) Cholesterol metabolism, transport, and hepatic regulation in dairy cows during transition and early lactation. J Dairy Sci 97:5481–5490. doi: 10.3168/jds.2014-7926

Khan MJ, Hosseini a, Burrell S, et al (2013) Change in subcutaneous adipose tissue metabolism and gene network expression during the transition period in dairy cows, including differences due to sire genetic merit. J Dairy Sci 96:2171–82. doi: 10.3168/jds.2012-5794

Khan RN, Matharoo-Ball B, Arulkumaran S, Ashford ML (2001) Potassium channels in the human myometrium. Exp Physiol 86:255–264.

Khatkar MS, Thomson PC, Tammen I, Raadsma HW (2004) Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genet Sel Evol 36:163–190. doi: 10.1186/1297-9686-36-2-163

Knight PG, Glister C (2006) TGF-beta superfamily members and ovarian follicle development. Reproduction 132:191–206. doi: 10.1530/rep.1.01074

Kolbehdari D, Wang Z, Grant JR, et al (2009) A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. J Anim Breed Genet 126:216–27. doi: 10.1111/j.1439-0388.2008.00793.x

Koufariotis L, Chen Y-PP, Bolormaa S, Hayes BJ (2014) Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. BMC Genomics 15:436. doi: 10.1186/1471-2164-15-436

Kuhn C, Bennewitz J, Reinsch N, et al (2003) Quantitative Trait Loci Mapping of Functional Traits in the German Holstein Cattle Population. J Dairy Sci 86:360–368. doi: 10.3168/jds.S0022-0302(03)73614-5

Kühn C, Freyer G, Weikard R, et al (1999) Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim Genet 30:333–340. doi: 10.1046/j.1365-2052.1999.00487.x

Kulak KK, Dekkers JCM, McAllister AJ, Lee AJ (1997) Relationships of early performance traits to lifetime profitability in Holstein cows. Can J Anim Sci 77:617–624. doi: 10.4141/A96-129

Laliotis GP, Bizelis I, Rogdakis E (2010) Comparative Approach of the de novo Fatty Acid Synthesis (Lipogenesis) between Ruminant and Non Ruminant Mammalian Species: From Biochemical Level to the Main Regulatory Lipogenic Genes. Curr Genomics 11:168–183. doi: 10.2174/138920210791110960

Lambeth JD (1986) Cytochrome P-450 $_{scc}$ A Review of the Specificity and Properties of the Cholesterol Binding Site. Endocr Res 12:371–392. doi: 10.3109/07435808609035446

Lande R, Thompson R (1998) Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. Heredity (Edinb) 80 (Pt 4):489–498. doi: 10.1046/j.1365-2540.1998.00308.x

Laven RA, Scaramuzzi RJ, Wathes DC, et al (2007) Recent research on the effects of excess dietary nitrogen on the fertility of dairy cows. Vet Rec 160:359–362. doi: 10.1136/vr.160.11.359

Lehner R, Kuksis A (1996) BIOSYNTHESIS OF TRIACYLGLYCEROLS. Prog Lipid Res 35:169–201. doi: O163-7827/96/$32.00

Lemay DG, Neville MC, Rudolph MC, et al (2007) Gene regulatory networks in lactation: identification of global principles using bioinformatics. BMC Syst Biol 1:56. doi: 10.1186/1752-0509-1-56

Leslie KE (1983) The events of normal and abnormal postpartum reproductive endocrinology and uterine involution in dairy cows: a review. Can Vet J 24:67–71.

Leury BJ, Baumgard LH, Block SS, et al (2003) Effect of insulin and growth hormone on plasma

leptin in periparturient dairy cows. Am J Physiol Regul Integr Comp Physiol 285:R1107–R1115. doi: 10.1152/ajpregu.00320.2003

Li G, Peñagaricano F, Weigel KA, et al (2012) Comparative genomics between fly, mouse, and cattle identifies genes associated with sire conception rate. J Dairy Sci 95:6122–9. doi: 10.3168/jds.2012-5591

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Li X, Li X, Chen H, et al (2013) Non-Esterified Fatty Acids Activate the AMP-Activated Protein Kinase Signaling Pathway to Regulate Lipid Metabolism in Bovine Hepatocytes. Cell Biochem Biophys 67:1157–1169. doi: 10.1007/s12013-013-9629-1

Lin X-Z, Luo J, Zhang L-P, et al (2013) MiR-27a suppresses triglyceride accumulation and affects gene mRNA expression associated with fat metabolism in dairy goat mammary gland epithelial cells. Gene 521:15–23. doi: 10.1016/j.gene.2013.03.050

Ling B, Alcorn J (2008) Acute administration of cefepime lowers L-carnitine concentrations in early lactation stage rat milk. J Nutr 138:1317–22.

Lipkin E, Mosig MO, Darvasi A, et al (1998) Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: Analysis of milk protein percentage. Genetics 149:1557–1567.

Littlejohn MD, Henty KM, Tiplady K, et al (2014a) Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. Nat Commun 5:5861. doi:

10.1038/ncomms6861

Littlejohn MD, Tiplady K, Fink TA, et al (2016) Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. Sci Rep 1–14. doi: 10.5061/dryad.457br

Littlejohn MD, Tiplady K, Lopdell T, et al (2014b) Expression variants of the lipogenic AGPAT6 gene affect diverse milk composition phenotypes in Bos taurus. PLoS One 9:e85757. doi: 10.1371/journal.pone.0085757

Liu F, Wollstein A, Hysi PG, et al (2010) Digital Quantification of Human Eye Color Highlights Genetic Association of Three New Loci. PLoS Genet 6:e1000934. doi: 10.1371/journal.pgen.1000934

Llewellyn S, Fitzpatrick R, Kenny D a., et al (2008) Endometrial expression of the insulin-like growth factor system during uterine involution in the postpartum dairy cow. Domest Anim Endocrinol 34:391–402. doi: 10.1016/j.domaniend.2007.11.003

Loor JJ, Everts RE, Bionaz M, et al (2007) Nutrition-induced ketosis alters metabolic and signaling gene networks in liver of periparturient dairy cows. Physiol Genomics 32:105–16. doi: 10.1152/physiolgenomics.00188.2007

Lucy MC (2001) Reproductive loss in high-producing dairy cattle: where will it end? J Dairy Sci 84:1277–93. doi: 10.3168/jds.S0022-0302(01)70158-0

Ma L, Corl B a (2012) Transcriptional regulation of lipid synthesis in bovine mammary epithelial cells by sterol regulatory element binding protein-1. J Dairy Sci 95:3743–55. doi:

10.3168/jds.2011-5083

Ma L, Wiggans GR, Wang S, et al (2012) Effect of sample stratification on dairy GWAS results. BMC Genomics 13:536. doi: 10.1186/1471-2164-13-536

Macdonald K a, Verkerk G a, Thorrold BS, et al (2008) A comparison of three strains of holstein-friesian grazed on pasture and managed under different feed allowances. J Dairy Sci 91:1693–1707. doi: 10.3168/jds.2007-0441

Mach N, Blum Y, Bannink a, et al (2012) Pleiotropic effects of polymorphism of the gene diacylglycerol-O-transferase 1 (DGAT1) in the mammary gland tissue of dairy cows. J Dairy Sci 95:4989–5000. doi: 10.3168/jds.2012-5348

Mandard S, Müller M, Kersten S (2004) Peroxisome proliferator-activated receptor a target genes. Cell Mol Life Sci 61:393–416. doi: 10.1007/s00018-003-3216-3

Mannaerts GP, Debeer LJ, Thomas J, De Schepper PJ (1979) Mitochondrial and peroxisomal fatty acid oxidation in liver homogenates and isolated hepatocytes from control and clofibrate-treated rats. J Biol Chem 254:4585–95.

Matukumalli LK, Lawley CT, Schnabel RD, et al (2009) Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS One. doi: 10.1371/journal.pone.0005350

Maxa J, Neuditschko M, Russ I, et al (2012) Genome-wide association mapping of milk production traits in Braunvieh cattle. J Dairy Sci 95:5357–64. doi: 10.3168/jds.2011-4673

Mcarthur MJ, Atshaves BP, Frolov A, et al (1999) Cellular uptake and intracellular trafficking of

long chain fatty acids. 40:1371–1383.

McGarry JD, Brown NF (1997) The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. Eur J Biochem 244:1–14. doi: 10.1111/j.1432-1033.1997.00001.x

McNamara JP (1988) Regulation of bovine adipose tissue metabolism during lactation. 4. Dose-responsiveness to epinephrine as altered by stage of lactation. J Dairy Sci 71:643–649. doi: 10.3168/jds.S0022-0302(88)79602-2

McNamara JP, Murray CE (2001) Sympathetic nervous system activity in adipose tissues during pregnancy and lactation of the rat. J Dairy Sci 84:1382–1389. doi: 10.3168/jds.S0022-0302(01)70169-5

Meredith BK, Kearney FJ, Finlay EK, et al (2012) Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet 13:21. doi: 10.1186/1471-2156-13-21

Meuwissen T, Goddard M (2010) Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185:623–31. doi: 10.1534/genetics.110.116590

Meuwissen THE, Goddard ME (2007) Multipoint identity-by-descent prediction using dense markers to map quantitative trait loci and estimate effective population size. Genetics 176:2551–2560. doi: 10.1534/genetics.107.070953

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829.

Miglior F, Chesnais J, Van Doormaal BJ Genetic improvement: a major component of increased dairy farm profitability.

Minozzi G, Nicolazzi EL, Stella A, et al (2013) Genome wide analysis of fertility and production traits in Italian Holstein cattle. PLoS One 8:e80219. doi: 10.1371/journal.pone.0080219

Miyoshi H, Perfield JW, Souza SC, et al (2007) Control of adipose triglyceride lipase action by serine 517 of perilipin A globally regulates protein kinase a-stimulated lipolysis in adipocytes. J Biol Chem 282:996–1002. doi: 10.1074/jbc.M605770200

Mootha VK, Lindgren CM, Eriksson K-F, et al (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34:267–273. doi: 10.1038/ng1180

Mullen MP, Berry DP, Howard DJ, et al (2011) Single nucleotide polymorphisms in the insulin-like growth factor 1 (IGF-1) gene are associated with performance in Holstein-Friesian dairy cattle. Front Genet 2:1–9. doi: 10.3389/fgene.2011.00003

Nafikov R a., Schoonmaker JP, Korn KT, et al (2014) Polymorphisms in lipogenic genes and milk fatty acid composition in Holstein dairy cattle. Genomics 104:572–581. doi: 10.1016/j.ygeno.2014.10.001

Nafikov R a, Schoonmaker JP, Korn KT, et al (2013) Sterol regulatory element binding transcription factor 1 (SREBF1) polymorphism and milk fatty acid composition. J Dairy Sci 96:2605–16. doi: 10.3168/jds.2012-6075

Nayeri S, Sargolzaei M, Abo-Ismail MK, et al (2016) Genome-wide association for milk

production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet 17:75. doi: 10.1186/s12863-016-0386-1

Nebel R., Dransfield M., Jobst S., Bame J. (2000) Automated electronic systems for the detection of oestrus and timing of AI in cattle. Anim Reprod Sci 60–61:713–723. doi: 10.1016/S0378-4320(00)00090-7

Nejati-Javaremi a (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. J Anim Sci 75:1738–1745.

Nemcova E, Stipkova M, Zavadilova L, et al (2007) The relationship between somatic cell count, milk production and six linearly scored type traits in Holstein cows. Czech J Anim Sci 52:437–446.

Nguyen P, Leray V, Diez M, et al (2008) Liver lipid metabolism. J Anim Physiol Anim Nutr (Berl) 92:272–283. doi: 10.1111/j.1439-0396.2007.00752.x

Norman HD, Wright JR, Hubbard SM, et al (2009) Reproductive status of Holstein and Jersey cows in the United States. J Dairy Sci 92:3517–3528. doi: 10.3168/jds.2008-1768

Olsen HG, Hayes BJ, Kent MP, et al (2011) Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. Anim Genet 42:466–74. doi: 10.1111/j.1365-2052.2011.02179.x

Oltenacu PA, Algers B (2005) Selection for Increased Production and the Welfare of Dairy Cows : Are New Breeding Goals Needed ? Ambio A J Hum Environ 34:311–315.

Oltenacu PA, Broom DM (2010) The impact of genetic selection for increased milk yield on the

welfare of. 19:39–49.

Ong EC, Nesin V, Long CL, et al (2013) A TRPC1 protein-dependent pathway regulates osteoclast formation and function. J Biol Chem 288:22219–22232. doi: 10.1074/jbc.M113.459826

Pant SD, Verschoor CP, Schenkel FS, et al (2011) Bovine PGLYRP1 polymorphisms and their association with resistance to Mycobacterium avium ssp. paratuberculosis. Anim Genet 42:354–360. doi: 10.1111/j.1365-2052.2010.02153.x

Park S-H, Ramachandran S, Kwon S-H, et al (2008) Upregulation of ATP-sensitive potassium channels for estrogen-mediated cell proliferation in human uterine leiomyoma cells. Gynecol Endocrinol 24:250–256. doi: 10.1080/09513590801893315

Patti ME, Butte AJ, Crunkhorn S, et al (2003) Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. Proc Natl Acad Sci U S A 100:8466–71. doi: 10.1073/pnas.1032913100

Payne AH, Hales DB (2004) Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr Rev 25:947–970. doi: 10.1210/er.2003-0030

Peñagaricano F, Weigel KA, Khatib H (2012) Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. Anim Genet 43:65–71. doi: 10.1111/j.1365-2052.2012.02350.x

Pezzi C, Accorsi PA, Vigo D, et al (2003) 5′-Deiodinase Activity and Circulating Thyronines in Lactating Cows. J Dairy Sci 86:152–158. doi: 10.3168/jds.S0022-0302(03)73595-4

Pimentel ECG, Bauersachs S, Tietze M, et al (2011) Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. Anim Genet 42:251–262. doi: 10.1111/j.1365-2052.2010.02148.x

Pimentel ECG, Tietze M, Simianer H, et al (2009) Study of Relationships Between Production and Fertility Traits in Dairy Cattle Using Genomic Data.

Poland J, Rutoski J (2015) Advances and Challenges in Genomic Selection for Disease Resistance. Annu Rev Phytopathol. doi: 10.1146/annurev-phyto-080615-100056

Pryce JE, Bolormaa S, Chamberlain a J, et al (2010a) A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J Dairy Sci 93:3331–45. doi: 10.3168/jds.2009-2893

Pryce JE, Daetwyler HD (2012) Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim Prod Sci 52:107–114. doi: Doi 10.1071/An11098

Pryce JE, Haile-Mariam M, Verbyla K, et al (2010b) Genetic markers for lactation persistency in primiparous Australian dairy cows. J Dairy Sci 93:2202–14. doi: 10.3168/jds.2009-2666

Pryce JE, Royal MD, Garnsworthy PC, Mao IL (2004) Fertility in the high-producing dairy cow. Livest Prod Sci 86:125–135. doi: 10.1016/S0301-6226(03)00145-3

Puigserver P, Spiegelman BM (2003) Peroxisome Proliferator-Activated Receptor-gamma Coactivator 1 alpha (PGC-1alpha): Transcriptional Coactivator and Metabolic Regulator.

Endocr Rev 24:78–90. doi: 10.1210/er.2002-0012

Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. doi: 10.1086/519795

Purfield DC, Bradley DG, Evans RD, et al (2015) Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. Genet Sel Evol 47:47. doi: 10.1186/s12711-015-0126-4

Raven L-A, Cocks BG, Goddard ME, et al (2014a) Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genet Sel Evol 46:29. doi: 10.1186/1297-9686-46-29

Raven L-A, Cocks BG, Goddard ME, et al (2014b) Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genet Sel Evol 46:29. doi: 10.1186/1297-9686-46-29

Raven L-A, Cocks BG, Hayes BJ (2014c) Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:62. doi: 10.1186/1471-2164-15-62

Reist M, Erdin D, von Euw D, et al (2002) Estimation of Energy Balance at the Individual and Herd Level Using Blood and Milk Traits in High-Yielding Dairy Cows,. J Dairy Sci 85:3314–3327. doi: 10.3168/jds.S0022-0302(02)74420-2

Reshef L, Olswang Y, Cassuto H, et al (2003) Glyceroneogenesis and the triglyceride/fatty acid cycle. J Biol Chem 278:30413–30416. doi: 10.1074/jbc.R300017200

Reue K, Dwyer JR (2008) Lipin proteins and metabolic homeostasis. J Lipid Res 50:S109–S114. doi: 10.1194/jlr.R800052-JLR200

Reverter A, Fortes MRS (2012) Building single nucleotide polymorphism-derived gene regulatory networks: Towards functional genomewide association studies. J Anim Sci 91:530–536. doi: 10.2527/jas.2012-5780

Rhoads ML, Gilbert RO, Lucy MC, Butler WR (2004) Effects of urea infusion on the uterine luminal environment of dairy cows. J Dairy Sci 87:2896–2901. doi: 10.3168/jds.S0022-0302(04)73420-7

Roche JR, Friggens NC, Kay JK, et al (2009) Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. J Dairy Sci 92:5769–5801. doi: 10.3168/jds.2009-2431

Rolf MM, Taylor JF, Schnabel RD, et al (2012) Genome-wide association analysis for feed efficiency in Angus cattle. Anim Genet 43:367–374. doi: 10.1111/j.1365-2052.2011.02273.x

Ron M, Weller JI (2007) From QTL to QTN identification in livestock - Winning by points rather than knock-out: A review. Anim Genet 38:429–439. doi: 10.1111/j.1365-2052.2007.01640.x

Royal MD, Flint APF, Woolliams JA (2002) Genetic and phenotypic relationships among endocrine and traditional fertility traits and production traits in Holstein-Friesian dairy cows. J Dairy Sci 85:958–67. doi: 10.3168/jds.S0022-0302(02)74155-6

Rudolph MC, McManaman JL, Phang T, et al (2007) Metabolic regulation in the lactating mammary gland: a lipid synthesizing machine. Physiol Genomics 28:323–336. doi: 10.1152/physiolgenomics.00020.2006

Saatchi M, Garrick DJ, Tait RG, et al (2013) Genome-wide association and prediction of direct genomic breeding values for composition of fatty acids in Angus beef cattle. BMC Genomics 14:730. doi: 10.1186/1471-2164-14-730

Saatchi M, Schnabel RD, Taylor JF, Garrick DJ (2014) Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics 15:442. doi: 10.1186/1471-2164-15-442

Sahana G, Guldbrandtsen B, Bendixen C, Lund MS (2010) Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. Anim Genet 41:579–588. doi: 10.1111/j.1365-2052.2010.02064.x

Sahana G, Guldbrandtsen B, Lund MS (2011) Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. J Dairy Sci 94:479–486. doi: 10.3168/jds.2010-3381

Sahana G, Guldbrandtsen B, Thomsen B, Lund MS (2013) Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. Anim Genet 44:620–626. doi: 10.1111/age.12053

Sallam AH, Endelman JB, Jannink J-L, Smith KP (2015) Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. Plant Genome 8:0. doi: 10.3835/plantgenome2014.05.0020

Santos JEP, Bisinotto RS, Ribeiro ES, et al (2010) Applying nutrition and physiology to improve reproduction in dairy cattle. Soc Reprod Fertil Suppl 67:387–403. doi: 10.5661/RDR-VII-387

Sargolzaei M (2014) SNP1101 User ' s Guide. Version 1.0.

Sargolzaei M (2012) FImpute User ' s Guide.

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. doi: 10.1186/1471-2164-15-478

Sargolzaei M, Schenkel FS, Vanraden PM (2013) gebv : Genomic breeding value estimator for livestock. L'alliance boviteq, sainthyacinthe, quebec Cent Genet Improv livestock, Univ guelph, ontario 1–7.

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218–223. doi: 10.1111/j.1439-0388.2006.00595.x

Schennink A, Heck JML, Bovenhuis H, et al (2008) Milk Fatty Acid Unsaturation: Genetic Parameters and Effects of Stearoyl-CoA Desaturase (SCD1) and Acyl CoA: Diacylglycerol Acyltransferase 1 (DGAT1). J Dairy Sci 91:2135–2143. doi: 10.3168/jds.2007-0825

Schennink A, Stoop WM, Visker MHPW, et al (2007) DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Anim Genet 38:467–473. doi: 10.1111/j.1365-2052.2007.01635.x

Schillo KK, Halls JB, Hileman SM (1992) effects of nutrition and season in the onset of puberty

in the beef heifer. J Anim Sci 70:3994–4005.

Schlegel G, Keller J, Hirche F, et al (2012) Expression of genes involved in hepatic carnitine synthesis and uptake in dairy cows in the transition period and at different stages of lactation. BMC Vet Res 8:28. doi: 10.1186/1746-6148-8-28

Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS (2005) Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Anim Genet 36:408–16. doi: 10.1111/j.1365-2052.2005.01337.x

Schork AJ, Thompson WK, Pham P, et al (2013) All SNPs Are Not Created Equal: Genome-Wide Association Studies Reveal a Consistent Pattern of Enrichment among Functionally Annotated SNPs. PLoS Genet. doi: 10.1371/journal.pgen.1003449

Schulman NF, Sahana G, Iso-Touru T, et al (2011) Mapping of fertility traits in Finnish Ayrshire by genome-wide association analysis. Anim Genet 42:263–9. doi: 10.1111/j.1365-2052.2010.02149.x

Schulman NF, Sahana G, Lund MS, et al (2008) Quantitative trait loci for fertility traits in Finnish Ayrshire cattle. Genet Sel Evol 40:195–214. doi: 10.1186/1297-9686-40-2-195

Sciascia Q, Pacheco D, McCoard SA (2013) Increased milk protein synthesis in response to exogenous growth hormone is associated with changes in mechanistic (mammalian) target of rapamycin (mTOR)C1-dependent and independent cell signaling. J Dairy Sci 96:2327–38. doi: 10.3168/jds.2012-6267

Sewalem A, Miglior F, Kistemaker GJ, et al (2008) Relationship between reproduction traits and

functional longevity in canadian dairy cattle. J Dairy Sci 91:1660–1668. doi: 10.3168/jds.2007-0178

Shindou H, Hishikawa D, Harayama T, et al (2008) Recent progress on acyl CoA: lysophospholipid acyltransferase research. J Lipid Res 50:S46–S51. doi: 10.1194/jlr.R800035-JLR200

Shook GE (2006) Major advances in determining appropriate selection goals. J Dairy Sci 89:1349–1361. doi: 10.3168/jds.S0022-0302(06)72202-0

Short TH, Lawlor TJ (1992) Genetic parameters of conformation traits, milk yield, and herd life in Holsteins. J Dairy Sci 75:1987–98. doi: 10.3168/jds.S0022-0302(92)77958-2

Silva AA, Azevedo ALS, Verneque RS, et al (2011) Quantitative trait loci affecting milk production traits on bovine chromosome 6 in zebuine Gyr breed. J Dairy Sci 94:971–980. doi: 10.3168/jds.2009-2970

Sinclair KD, Sinclair LA, Robinson JJ (2000) Nitrogen metabolism and fertility in cattle: I. Adaptive changes in intake and metabolism to diets differing in their rate of energy and nitrogen release in the rumen. J Anim Sci 78:2659–2669.

Sivan E, Boden G (2003) Free fatty acids, insulin resistance, and pregnancy. Curr Diab Rep 3:319–322. doi: 10.1007/s11892-003-0024-y

Snelling WM, Cushman R a., Keele JW, et al (2013) Networks and pathways to guide genomic selection. J Anim Sci 91:537–552. doi: 10.2527/jas2012-5784

Sodeland M, Kent MP, Olsen HG, et al (2011) Quantitative trait loci for clinical mastitis on

chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. Anim Genet 42:457–465. doi: 10.1111/j.1365-2052.2010.02165.x

Solovieff N, Cotsapas C, Lee PH, et al (2013) Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 14:483–495. doi: 10.1126/scisignal.2001449.Engineering

Spicer LJ, Echternkamp SE (1995) The ovarian insulin and insulin-like growth factor system with an emphasis on domestic animals. Domest Anim Endocrinol 12:223–245. doi: 10.1016/0739-7240(95)00021-6

Stagg K, Spicer LJ, Sreenan JM, et al (1998) Effect of calf isolation on follicular wave dynamics, gonadotropin and metabolic hormone changes, and interval to first ovulation in beef cows fed either of two energy levels postpartum. Biol Reprod 59:777–783. doi: 10.1095/biolreprod59.4.777

Stevenson KR, Wathes DC (1996) Insulin-like growth factors and their binding proteins in the ovine oviduct during the oestrous cycle. J Reprod Fertil 108:31–40. doi: 10.1530/jrf.0.1080031

Stothard P, Choi J-W, Basu U, et al (2011) Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics 12:559. doi: 10.1186/1471-2164-12-559

Su G, Brøndum RF, Ma P, et al (2012a) Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. J Dairy Sci 95:4657–65. doi: 10.3168/jds.2012-5379

Su G, Madsen P, Nielsen US, et al (2012b) Genomic prediction for Nordic Red Cattle using one-step and selection index blending. J Dairy Sci 95:909–17. doi: 10.3168/jds.2011-4804

Sumner-Thomson JM, Vierck JL, McNamara JP (2011) Differential expression of genes in adipose tissue of first-lactation dairy cattle. J Dairy Sci 94:361–369. doi: 10.3168/jds.2010-3447

Sumner JM, McNamara JP (2007) Expression of lipolytic genes in the adipose tissue of pregnant and lactating Holstein dairy cattle. J Dairy Sci 90:5237–46. doi: 10.3168/jds.2007-0307

Sutherland KD, Lindeman GJ, Visvader JE (2007) The molecular culprits underlying precocious mammary gland involution. J Mammary Gland Biol Neoplasia 12:15–23. doi: 10.1007/s10911-007-9034-8

Szkiba D, Kapun M, von Haeseler A, Gallach M (2014) SNP2GO: functional analysis of genome-wide association studies. Genetics 197:285–9. doi: 10.1534/genetics.113.160341

Takeuchi K, Reue K (2009) Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis. Am J Physiol Endocrinol Metab 296:E1195–E1209. doi: 10.1152/ajpendo.90958.2008

Taylor VJ, Cheng Z, Pushpakumara PGA, et al (2004) Relationships between the plasma concentrations of insulin-like growth factor-I in dairy cows and their fertility and milk yield. Vet Rec 155:583–588. doi: 10.1136/vr.155.19.583

The Bovine Hapmap Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528–532. doi: 10.1126/science.1167936

Thomasen JR, Guldbrandtsen B, Sørensen P, et al (2008) Quantitative trait loci affecting calving traits in Danish Holstein cattle. J Dairy Sci 91:2098–2105. doi: 10.3168/jds.2007-0602

Van den Top AM, Wensing T, Geelen MJ, et al (1995) Time trends of plasma lipids and enzymes synthesizing hepatic triacylglycerol during postpartum development of fatty liver in dairy cows. J Dairy Sci 78:2208–2220. doi: 10.3168/jds.S0022-0302(95)76848-5

Van Dorland HA, Richter S, Morel I, et al (2009) Variation in hepatic regulation of metabolism during the dry period and in early lactation in dairy cows. J Dairy Sci 92:1924–1940. doi: 10.3168/jds.2008-1454

VanRaden PM (2004) Invited Review: Selection on Net Merit to Improve Lifetime Profit. J Dairy Sci 87:3125–3131. doi: 10.3168/jds.S0022-0302(04)73447-5

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423. doi: 10.3168/jds.2007-0980

VanRaden PM, O'Connell JR, Wiggans GR, Weigel K a (2011) Genomic evaluations with many more genotypes. Genet Sel Evol 43:10. doi: 10.1186/1297-9686-43-10

VanRaden PM, Van Tassell CP, Wiggans GR, et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24. doi: 10.3168/jds.2008-1514

VanRaden PM, Wiggans GR (1995) Productive Life Evaluations: Calculation, Accuracy, and Economic Value. J Dairy Sci 78:631–638. doi: 10.3168/jds.S0022-0302(95)76674-7

Veerkamp RF, Beerda B (2007) Genetics and genomics to improve fertility in high producing

dairy cows. Theriogenology 68:266–273. doi: 10.1016/j.theriogenology.2007.04.034

Veerkamp RF, Coffey MP, Berry DP, et al (2012) Genome-wide associations for feed utilisation complex in primiparous Holstein–Friesian dairy cows from experimental research herds in four European countries. Animal 6:1738–1749. doi: 10.1017/S1751731112001152

Vernon RG, Denis RGP, Sørensen a. (2001) Signals of adiposity. Domest Anim Endocrinol 21:197–214. doi: 10.1016/S0739-7240(01)00121-7

Viitala S, Szyda J, Blott S, et al (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. doi: 10.1534/genetics.105.046730

Viollet B, Foretz M, Guigas B, et al (2006) Activation of AMP-activated protein kinase in the liver: a new strategy for the management of metabolic hepatic disorders. J Physiol 574:41–53. doi: 10.1113/jphysiol.2006.108506

Viturro E, Koenning M, Kroemer A, et al (2009) Cholesterol synthesis in the lactating cow: Induced expression of candidate genes. J Steroid Biochem Mol Biol 115:62–7. doi: 10.1016/j.jsbmb.2009.02.011

Walsh SW, Williams EJ, Evans ACO (2011) A review of the causes of poor fertility in high milk producing dairy cows. Anim Reprod Sci 123:127–138. doi: 10.1016/j.anireprosci.2010.12.001

Wang M, Moisá S, Khan MJ, et al (2012a) MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation. J Dairy Sci 95:6529–35. doi:

10.3168/jds.2012-5748

Wang X, Ma P, Liu J, et al (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 16:111. doi: 10.1186/s12863-015-0263-3

Wang X, Wurmser C, Pausch H, et al (2012b) Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. PLoS One 7:e40711. doi: 10.1371/journal.pone.0040711

Wathes DC, Cheng Z, Chowdhury W, et al (2009) Negative energy balance alters global gene expression and immune responses in the uterus of postpartum dairy cows. 1–13. doi: 10.1152/physiolgenomics.00064.2009.

Wathes DC, Cheng Z, Fenwick MA, et al (2011) Influence of energy balance on the somatotrophic axis and matrix metalloproteinase expression in the endometrium of the postpartum dairy cow. Reproduction 141:269–281. doi: 10.1530/REP-10-0177

Wathes DC, Fenwick M, Cheng Z, et al (2007) Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. Theriogenology 68:232–241. doi: 10.1016/j.theriogenology.2007.04.006

Wathes DC, Taylor VJ, Cheng Z, Mann GE (2003) Follicle growth, corpus luteum function and their effects on embryo development in postpartum dairy cows. Reprod Suppl 61:219–237.

Weikard R, Goldammer T, Brunner RM, Kuehn C (2012) Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk

production in cows. Physiol Genomics 44:728–739. doi: 10.1152/physiolgenomics.00007.2012

Weng L, Macciardi F, Subramanian A, et al (2011) SNP-based pathway enrichment analysis for genome-wide association studies. BMC Bioinformatics 12:99. doi: 10.1186/1471-2105-12-99

Wickramasinghe S, Hua S, Rincon G, et al (2011) Transcriptome profiling of bovine milk oligosaccharide metabolism genes using RNA-sequencing. PLoS One 6:e18895. doi: 10.1371/journal.pone.0018895

Wiggans GR, Sonstegard TS, VanRaden PM, et al (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. doi: 10.3168/jds.2008-1758

Wu X, Lund MS, Sahana G, et al (2015) Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. Genet Sel Evol 47:50. doi: 10.1186/s12711-015-0129-1

Xie Y, Yang S, Cui X, et al (2014) Identification and expression pattern of two novel alternative splicing variants of EEF1D gene of dairy cattle. Gene 534:189–96. doi: 10.1016/j.gene.2013.10.061

Yamashita A, Nakanishi H, Suzuki H, et al (2007) Topology of acyltransferase motifs and substrate specificity and accessibility in 1-acyl-sn-glycero-3-phosphate acyltransferase 1. Biochim Biophys Acta - Mol Cell Biol Lipids 1771:1202–1215. doi: 10.1016/j.bbalip.2007.07.002

Yang J, Goddard ME, Visscher PM (2010) A commentary on "common SNPs explain a large proportion of the heritability for human height" by Yang et al. (2010). Twin Res Hum Genet 13:517–524. doi: 10.1038/ng.608

Zhang Q, Boichard D, Hoeschele I, et al (1998) Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics 149:1959–1973.

Zhang Q, Koser SL, Bequette BJ, Donkin SS (2015) Effect of propionate on mRNA expression of key genes for gluconeogenesis in liver of dairy cattle. J Dairy Sci 98:8698–709. doi: 10.3168/jds.2015-9590

Zhou G, Myers R, Li Y, et al (2001) Role of AMP-activated protein kinase in mechanism of metformin action. J Clin Invest 108:1167–1174. doi: 10.1172/JCI200113505

Zhu LH, Armentano LE, Bremmer DR, et al (2000) Plasma Concentration of Urea, Ammonia, Glutamine Around Calving, and the Relation of Hepatic Triglyceride, to Plasma Ammonia Removal and Blood Acid-Base Balance. J Dairy Sci 83:734–740. doi: 10.3168/jds.S0022-0302(00)74935-6

Zhu M, Zhao S (2007) Candidate gene identification approach: Progress and challenges. Int J Biol Sci 3:420–427.