



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Genetics of calving performance

A Thesis Submitted to the University of Dublin for the Degree of
Doctor of Philosophy

2015

Deirdre Christina Purfield B.A (Hons)



¹Molecular Population Genetics,
Smurfit Institute of Genetics,
Trinity College Dublin,
Dublin 2

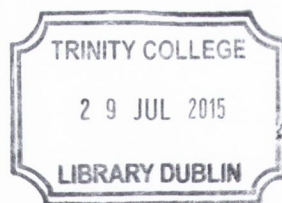


²Animal & Bioscience Research Dept.,
Animal & Grassland Research & Innovation Centre,
Teagasc, Moorepark,
Fermoy, Co. Cork

Research Supervisors

¹Prof. D.G Bradley

²Dr. D.P Berry



Thesis 10595

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to the Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed Deirdre Purfield Date 24/02/2015

Summary

Calving performance is a significant contributor to the profitability of cattle production systems whilst also becoming an increasingly important welfare issue in recent years. Calving performance includes both dystocia and perinatal mortality. Dystocia, commonly known as calving difficulty is defined as a prolonged or difficult parturition whereas perinatal mortality is defined as calf mortality shortly before, during or shortly after parturition. This thesis describes four separate bodies of work that were carried out to understand the genetic variation focusing on dystocia and perinatal mortality.

Dystocia and perinatal mortality traits were widely prevalent nationally (dystocia 4.29%, perinatal mortality 4.15%) with under half of mortalities occurring where no dystocia was observed (1.57%) from 2008 to 2012. Genetic parameters were estimated for dystocia and perinatal mortality using 100,664 calving events from the same time period. Heritability estimates differed substantially for calving difficulty between primiparous (direct 0.207, maternal 0.034) and multiparous dams (direct 0.104, maternal 0.008). Perinatal mortality was lowly heritable across all parities (primiparous direct 0.007 & maternal 0.017; multiparous direct 0.001 & maternal 0.001). The existence of genetic variation for both traits suggests the use of breeding programs to lower the prevalence will be fruitful.

The extent of homozygosity per animal was measured by characterising lengths of runs of homozygosity (ROH) in a variety of breeds across 891 HD genotyped sires and 1166 Bovine SNP50 genotyped sires. ROH were frequent across all breeds but variation existed between breeds due to a difference in breed origins and recent management. A strong correlation ($r=0.75$) existed between the pedigree inbreeding coefficient and a statistic based on the sum of ROH length $>0.5\text{Mb}$ suggesting this statistic is a sufficient predictor of inbreeding.

Regions of the bovine genome associated with dystocia and perinatal mortality in Holstein–Friesian dairy cows were detected using 54,001 SNPs on up to 1,970 Holstein-

Friesian sires. A genome-wide association study (GWAS) was carried out by fitting all SNPs simultaneously in a Bayesian framework and phenotypes were provided through the use of predicted transmitting abilities. Several associations were identified and one genomic region on chromosome 18 suggested strong evidence of an underlying causative mutation for direct calving difficulty. Genomic regions only associated with one calving trait were identified suggesting that direct selection for improvement in calving difficulty whilst overcoming the genetic antagonism that exists between the remaining calving traits is possible.

A frequentist and a Bayesian approach were used to identify genomic regions of association for calving performance in HD genotyped Holstein-Friesian (n=927), Charolais (n=770) and Limousin (n=963) populations. Genomic regions associated for each trait differed per breed emphasising the complex nature of calving performance. Genome sequence data was used to refine strong associations detected in the HD GWAS in the Holstein-Friesian population. Again a strong association for direct calving difficulty was detected on chromosome 18 and imputed sequence data refined this to a 1Mb region containing 16 gene/gene-products. Four highly significant missense variants were identified and the variant within CTU1 is suggested to impact protein structure (SIFT=0.06; PolyPhen=0.95). Genetic architecture surrounding genomic associations was further assessed by analysing copy number variation and ROH and double copy CNVs were commonly found for this QTL on chromosome 18.

In summary, clear genetic variation exists some of which is attributable to alleles of large effect but the vast majority is due to common alleles, each with a small effect in calving performance in Irish herds. Genetic improvement for calving performance is possible and an understanding of the genomic architecture will help facilitate a targeted breeding approach.

CONTENTS

INTRODUCTION.....	1
1.1. Irish Agriculture.....	2
1.1.1 Cattle Breeding in Ireland	2
1.1.2 Genetic trends in Irish dairy and beef performance	6
1.2. Quantitative genetics and statistical models	8
1.2.1 Terminology in animal quantitative genetics.....	9
1.2.2 Genetic Evaluations.....	12
1.2.2.1 Best Linear Unbiased Predicts (BLUP).....	13
1.2.2.2 Linear mixed model	13
1.2.2.3 Repeatability model	15
1.2.3 Deregression of EBVs	16
1.2.4 Inclusion of genomic data in genetic evaluations	17
1.3 Alternative structural variants.....	18
1.3.1 Runs of homozygosity	18
1.3.2 Copy number variation	18
1.4. Genome Wide Association Studies	19
1.4.1 Genotyping and Single Nucleotide Polymorphism quality control	22
1.4.2 Statistical approaches to genome-wide association studies	25
1.4.2.1 Regression approaches to GWAS	25
1.4.2.2 Bayesian hierarchical modelling	27
1.4.3 Advantages and disadvantages of the genome-wide association approach	31
1.5 Pathway analysis	33
1.6. Calving performance traits	34
1.6.1. Dystocia.....	34
1.6.2. Perinatal Mortality	35
1.6.3 Risk factors associated with calving performance traits	39
1.6.3.1. Effects of the calf	39
1.6.3.2 Effects of the dam	40
1.6.3.3. Genetic parameters of calving performance	42
1.7. Genome-wide association studies for calving performance	44
1.8. Gaps in Knowledge.....	47
 GENETIC PARAMETERS OF CALVING PERFORMANCE IN IRISH DAIRY AND BEEF COWS	 49
2.1 Preface	49
2.2 Abstract	50
2.3 Introduction.....	51

2.4 Methods 52

2.4.1 Data edits..... 52

2.3.2 Data Analysis 54

2.5 Results..... 55

2.5.1 Prevalence statistics..... 55

2.5.2 Genetic Parameters..... 56

2.5.3 Genetic correlations..... 57

2.6 Discussion 58

2.6.1 Dystocia..... 58

2.6.2 Perinatal mortality 60

2.6.3 Perinatal mortality not associated with dystocia 62

2.7 Conclusion 63

RUNS OF HOMOZYGOSITY AND POPULATION HISTORY IN CATTLE 65

3.1 Preface 65

3.2 Abstract 66

3.3 Introduction..... 67

3.4 Methods 68

3.4.1 Genotypes and Quality Control 68

3.4.2 Definition of a Run of Homozygosity 69

3.4.3 Inbreeding Coefficient vs. Runs of Homozygosity 71

3.5 Results..... 72

3.5.1 ROH in animals with HD panel genotypes..... 72

3.5.2 Genome locations of ROH 73

3.5.3 Correlation between ROH and inbreeding coefficient 75

3.5.4 Validation of Bovine SNP50 density..... 76

3.5.5 ROH in Bovine SNP50 genotypes across the 42 breeds 78

3.6 Discussion 83

3.7 Conclusion 88

**GENOME-WIDE ASSOCIATION STUDY FOR CALVING TRAITS IN
HOLSTEIN-FRIESIAN DAIRY CATTLE..... 89**

4.1 Preface 89

4.2 Abstract 90

4.3 Introduction..... 91

4.4 Methods 92

4.4.1 Phenotypic data 92

4.4.2 Genotypic data..... 94

4.4.3 Whole genome association	94
4.5 Results	98
4.5.1 Direct calving difficulty	98
4.5.2 Maternal calving difficulty	103
4.5.3 Perinatal Mortality	106
4.6 Discussion.....	109
4.6.1 Chromosome 18	109
4.6.2 Chromosome 17	110
4.6.3 Chromosomes 2, 16 & 12	111
4.6.4 Chromosome 23 & 28	112
4.6.5 Chromosome 10 & 14	113
4.6.6 Pathway analysis	114
4.7 Conclusion	115
 GENOME-WIDE ASSOCIATION STUDY FOR CALVING PERFORMANCE USING HIGH DENSITY GENOTYPES IN DAIRY AND BEEF CATTLE.	 117
5.1 Preface	117
5.2 Abstract	118
5.3 Introduction.....	119
5.4 Methods	120
5.4.1 Genotypic data.....	120
5.4.2 Phenotypic data	121
5.4.3 Whole genome association analysis	122
5.4.3.1 Single SNP regression.....	122
5.4.3.2 Bayesian analyses.....	123
5.4.4 Bioinformatics	125
5.4.5 Sequence analysis.....	125
5.4.6 Copy Number Variation	126
5.4.7 Runs of homozygosity	127
5.5 Results	128
5.5.1 Direct calving difficulty	128
5.5.2 Meta-analyses for direct calving difficulty	131
5.5.3 Maternal calving difficulty	135
5.5.4 Meta-analysis of maternal calving difficulty	137
5.5.5 Perinatal Mortality	139
5.5.6 Meta-analysis of perinatal mortality	140
5.5.7 Copy number variants.....	142
5.5.8 Runs of homozygosity and calving performance	143
5.6 Discussion	144
5.6.1 Direct calving difficulty	144
5.6.2 Maternal calving difficulty	146
5.6.3 Perinatal Mortality	148

5.7 Conclusion 149

DISCUSSION 151

6.1 Summary of thesis 152

6.1.1 Chapter 2: Genetic parameters of calving performance traits 154

6.1.2 Chapter 3: Runs of homozygosity and population history in cattle 155

6.1.3 Chapter 4: Genome-wide association study for calving traits in Holstein-Friesian dairy cattle..... 156

6.1.4 Chapter 5: Genome-wide association study for calving performance using high density genotypes in dairy and beef cattle. 157

6.2 Overall conclusions and implications 158

REFERENCES..... 165

PUBLICATIONS FROM THE THESIS 189

APPENDIX A 193

A. Runs of homozygosity and population history in cattle 193

APPENDIX B..... 201

B. Genome-wide association study for calving traits in Holstein-Friesian dairy cattle..... 201

List of figures

Figure 1.1 Schematic diagram showing the data inputs and outputs of the ICBF.....	3
Figure 1.2 Relative emphasis on various traits in the EBI over a fourteen year period.	4
Figure 1.3 Genetic trend of the Economic Breeding Index (EBI), Terminal Beef and Maternal Beef Indexes from 2000 to 2012.....	7
Figure 1.4 Stock chart showing the breed average, top 5 percentile and bottom 5 percentile of PTA for calving difficulty based on all pedigree animals in each breed born since 2007.	8
Figure 1.5 Graphic depiction of a positive, negative and no association between variable	10
Figure 1.6 A simple linear regression model	11
Figure 1. 7 Different types of copy number variation (CNV) found in the genome.	19
Figure 1.8 Graphical depiction of the benefits of discovering genes associated with diseases.	20
Figure 1.9 The principle behind Illumina genotyping and the GenTrain clustering method	23
Figure 1.10 Schematic overview of the Bayesian approach incorporating both prior and likelihood information.	27
Figure 1.11 Accuracy gained in predictions versus the number of animal records for four heritabilities.	33
Figure 3.1 The mean sum of runs of homozygosity (ROH) per HD genotyped animal measured in megabases (Mb) within each ROH length category.....	73
Figure 3.2 Incidence of each single nucleotide polymorphism (SNP) in a run of homozygosity (ROH) in the HD panel.....	74
Figure 3.3 Scatter plot of A) FROHon pedigree inbreeding coefficient B) FROH10000 on pedigree inbreeding coefficient ^a	76
Figure 3.4 The percentage of runs of homozygosity (ROH) in the HD panel that are recognised by the reduced density panel and the percentage of the reduced panel ROH that are correctly identified into the correct ROH length category.	78
Figure 3.5 Individual sum of total runs of homozygosity (ROH) per bovine SNP50 animal split into origin of breed.	80
Figure 3.6 Average sum of runs of homozygosity (ROH) per bovine SNP50 animal <20Mb in length within breed vs average sum of ROH>20Mb in length.....	82
Figure 4.1 The Bayesian posterior probability of model frequency from a Bayes B analysis for calving traits.....	100

Figure 4.2 Haplotype analysis of a 500kb region up and downstream surrounding ARS-BFGL-NGS-109285 (ss86324977) on chromosome 18. 101

Figure 4.3 Haplotype analysis of a 500kb region up and downstream surrounding BTA-63756-no-rs (rs41652463) on chromosome 28..... 106

Figure 4.4 Haplotype analysis of a 500kb region surrounding ARS-BFGL-NGS-16146 (ss86296129) on chromosome 14..... 108

Figure 5.1 Manhattan plots for single SNP regression analysis and Bayesian analysis for A) Holstein-Friesians B) Charolais C) Limousin and D) the meta-analysis for direct calving difficulty..... 132

Figure 5.2 Imputation to whole genome sequence single SNP regression results of 10Mb region surrounding ARS-BFGL-NGS-109285 on chromosome 18. 134

Figure 5.3 Imputed to sequence density single SNP regression (SSR) results for direct calving difficulty in the Holstein-Friesian population..... 132

Figure 5.4 Manhattan plots for single SNP regression analysis and Bayesian analysis for A) Holstein-Friesians B) Charolais, C) Limousin and D) the meta-analysis result for maternal calving difficulty..... 138

Figure 5.5 Manhattan plots for single SNP regression analysis and Bayesian analysis for A) Holstein-Friesians B) Charolais, C) Limousin and D) the meta-analysis manhattan for direct perinatal mortality..... 141

Figure 5.6 Size length distribution of copy number variants (CNV) and percentage of each copy number state identified in each breed. 142

List of tables

Table 1.1 Traits, economic values and emphasis that constitute the Economic Breeding Index.....	5
Table 1.2 Traits, economic values, and emphasis that constitute the Terminal Beef Index and the Maternal Beef Index.	6
Table 1.3 International prevalence rates for dystocia (defined as considerable or veterinary assistance) and perinatal mortality (defined as the death of a newborn calf within the first 24hrs of parturition) in primiparous (primi) and multiparous (multi) dams.....	38
Table 1.4 International direct and maternal heritability estimates for dystocia and perinatal mortality.	43
Table 2.1 Number of birth events, herds, sires, maternal grandsires (MGS) and contemporary groups (CG) in each dataset.....	54
Table 2.2 Prevalence of dystocia (scored 1 to 4), perinatal mortality and perinatal mortality not associated with dystocia.....	56
Table 2.3 Heritability estimates (direct and maternal) and genetic variation for dystocia, perinatal mortality and perinatal mortality not associated with dystocia (PMND) in primiparous, multiparous and a combination of both categories of animals, estimated using univariate analysis.	57
Table 3.1 Number of single nucleotide polymorphisms for each data edit implemented on each density panel.....	69
Table 4.1 Proportion of genetic variance explained by SNPs and the number of SNPs found in three different Bayesian posterior probability ranges for all traits.....	99
Table 4.2 Bayes Factors and posterior probability of associations (PPA) for all SNPs that showed 'substantial evidence' (Bayes factor >3.1) of an association with direct calving difficulty ordered by PPA.....	102
Table 4.3 Bayes Factors and posterior probability of associations (PPA) for all SNPs that showed 'substantial evidence' (Bayes factor >3.1) of an association with maternal calving difficulty ordered by PPA.....	105
Table 4.4 Bayes Factors and posterior probability of associations (PPA) for the ten most significant SNPs associated with direct perinatal mortality ordered by PPA.....	108
Table 5.1 Bayes Factors (BF) and single SNP regression p values for all SNPs with a BF >200 for direct and maternal calving difficulty and a BF >40 for perinatal mortality..	133

List of abbreviations

AI	Artificial insemination
BAF	B-allele frequency
BCS	Body condition score
BF	Bayes factor
BLAD	Bovine leukocyte adhesion deficiency
BLUP	Best linear unbiased prediction
CG	Contemporary group
CGE	Complete generation equivalents
CH	Charolais
CHR	Chromosome
CNV	Copy number variant
CVM	Complex vertebrate mutation
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DGV	Direct genomic value
DUMPs	Deficiency of uridine monophosphate synthase
EBI	Economic breeding index
EBV	Estimated breeding value
FROH	Statistic based on sum of runs of homozygosity
GCscore	GenCall score
GO	Gene ontology
GTscore	GenTrain score
GWAS	Genome wide association study
HD	High density
HOL-FR	Holstein-Friesian
HWE	Hardy weinberg equilibrium
ICBF	Irish cattle breeding federation
IgG	ImmunoglobulinG
KEGG	Kyoto encyclopedia of genes and genomes

KB	Kilobase
LD	Linkage disequilibrium
LDLA	Linkage disequilibrium linkage analysis
LIM	Limousin
Log	Logarithm
LRR	Log-R ratio
MAF	Minor allele frequency
MB	Megabase
MGS	Maternal grandsire
MME	Mixed model equation
PMND	Perinatal mortality not associated with dystocia
PO	Posterior odds
PPA	Posterior probability of association
PTA	Predicted transmitting ability
QTL	Quantitative trait loci
RBI	Relative breeding index
ROH	Runs of homozygosity
SNP	Single nucleotide polymorphism
SSR	Single SNP regression
Yr	Year

Acknowledgements

I would like to acknowledge the opportunity presented to me by Teagasc and Trinity College Dublin to carry out this work. I would also like to acknowledge Science Foundation Ireland for providing financial support for this work to be completed. In addition, I am grateful to Dr. Pat Dillon, Head of the Animal and Grassland Research and Innovation Programme for making facilities available to me throughout my PhD.

I would like to acknowledge the support, advice, and guidance of both my supervisors, Dr. Donagh Berry and Prof. Dan Bradley throughout my PhD. I would like to thank them for sharing their vast knowledge, skills, patience and expertise with me whilst providing me with a variety of opportunities to further my knowledge of animal genomics and breeding during the last four years. Their enthusiasm and dedication for their respective fields is contagious and enjoyable to be a part of. In particular, I would like to thank Dr. Donagh Berry for helping with the SAS tantrums, the crashed computers and all the little problems along the way. I would also like to thank my other co-authors who have helped contributed to this PhD, Dr. Ross Evans and Dr. Francis Kearney of the Irish Cattle Breeding Federation and Dr. Sinead McParland of Teagasc.

I would like to acknowledge the many friends that have provided support throughout the last four years. Firstly to the home crew; Flano, Nic, Moloney and Hussey, no matter how far the distance was over the last few years it never mattered. Secondly, to the Moorepark bunch, I seriously cannot thank all of you enough for all the laughs, the great nights out, the tea time chats, the support and general banter. I would specifically like to thank Noirin for taking me under her wing, Jess and Vinny for being the best office mates that could be asked for and to the Ardkeen crew; Ellen, Roberta, Amy, Aine and Mags for providing plenty of distractions. I would also like to say a huge thank you to Marion and Will for all the chats, support and nights out and finally to Stephen for his encouragement, laughter and friendship.

To my family, John, Julie and particularly Catriona, thank you for the encouragement and support. To Mam, to whom this thesis is dedicated, your support and encouragement over the years has been never ending, thank you so much for everything and lastly to Dad for providing me with one last challenge, thank you.

Chapter 1

Introduction

1.1. Irish Agriculture

Gross agricultural output in Ireland in 2012 was estimated to be €5.58 billion. The beef sector accounted for the largest proportion (i.e., 38%) followed by dairy production (29%) (BordBia, 2012). Of the 6.2 million cattle in Ireland, 1.06 million are dairy cows and 1.12 million are beef cows (CSO, 2013). Beef cattle production is the most prevalent enterprise on Irish farms, with 100,000 or 90% of farms having a beef cattle enterprise (BordBia, 2012). Herd size in Ireland is, on average, small for both dairy and beef enterprises with an average of 50 cows per dairy herd and 18 cows per beef herd (O'Mara, 2012). Both dairy and beef production systems predominately operate seasonal production systems (Berry et al., 2013, Berry and Evans, 2014); 69% of cows calve between January and March and a further 8% calve between October and December (CMMS, 2012). The seasonality of production is to maximise the utilisation of low-cost grazed grass in the diet.

1.1.1 Cattle Breeding in Ireland

The Irish Cattle Breeding Federation (ICBF) was founded in 1996 with the objective of establishing breeding schemes to deliver optimal rates of genetic improvement in cattle using the best available technologies. One of the key roles of the ICBF is to provide a centralised database to support the information requirements of cattle breeding in Ireland (Figure 1.1). The information stored in the database includes animal events recordings, movement notifications, milk recording for 50% of the national dairy, herd linear scoring for the dairy and beef traits, as well as dairy and beef genetic evaluations (ICBF, 2008, Wickham, 2004). The database now also includes genotypes of animals and genomic breeding values estimated using genomic selection algorithms (Wickham, 2004).

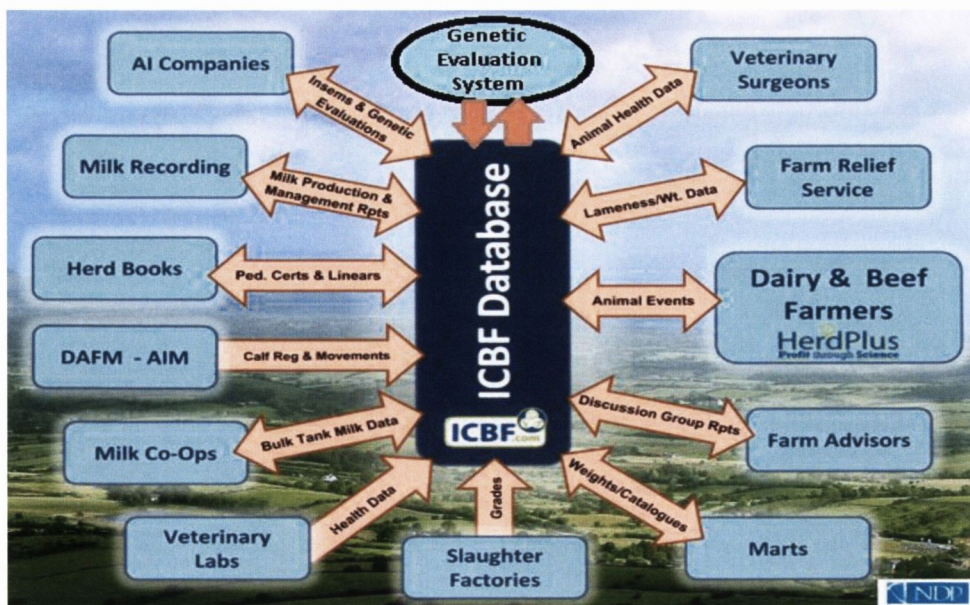


Figure 1.1 Schematic diagram showing the data inputs and outputs of the ICBF. (Source; <http://www.icbf.com>)

National beef and dairy genetic evaluations are computed by the ICBF on behalf of the Irish cattle breeding industry. The data are routinely extracted from the database (some 60 traits in total are evaluated) to undertake genetic evaluations which are then uploaded back into the database for widespread publication.

A number of indexes have been developed by the ICBF, in collaboration with national and international collaborators (Berry et al., 2007a; Amer et al., 2001). The national breeding objectives in Ireland are:

- Dairy: Economic Breeding Index (EBI)
- Beef: Maternal index
- Beef: Terminal index

The ICBF officially launched the EBI to the dairy industry in February 2001. This index replaced the existing relative breeding index (RBI) which focused solely on the production traits of milk, fat and protein yield. The EBI is a single figure profit indexes aimed to aid producers identify the most profitable parents of the next generation. The EBI is currently

comprised of seven sub-indexes related to profit and the evolution of the EBI over time can be seen in Figure 1.2. As milk production and fertility greatly impact profit on Irish dairy farms, these two suites of traits represent the largest proportion of the emphasis within the EBI (33% and 35% of the weighing, respectively) (Table 1.1). Calving performance, such as calving difficulty and calf mortality also represent a substantial part of this index (9%). The inclusion of calving traits in the index has aided producers in identifying parents whose progeny are more likely to experience difficulty calving and more likely to die at birth. Regression analyses on a large dataset of herds with herd profit information clearly showed that differences in profit for different herd EBI were within the theoretical expectation (Ramsbottom et al., 2012).

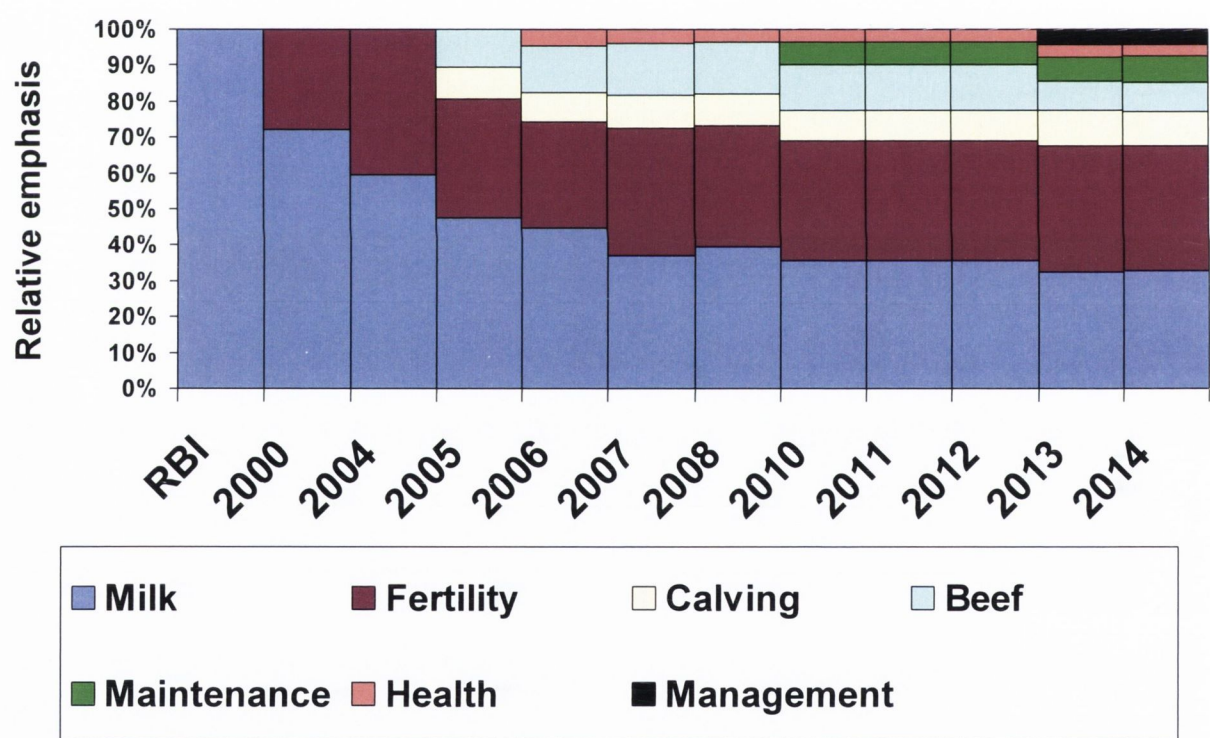


Figure 1.2 Relative emphasis on various traits in the EBI over a fourteen year period. (Source; <http://www.icbf.com>)

Table 1.1 Traits, economic values and relative emphasis that constitute the Economic Breeding Index

Sub-Index	Trait	Economic Weight	Emphasis	Overall Emphasis
Production	Milk	-€0.09	10.6%	33%
	Fat	€1.04	3.4%	
	Protein	€6.64	18.9%	
Fertility	Calving Interval	-€12.43	24.0%	35%
	Survival	€12.01	10.9%	
Calving	Direct Calving Difficulty	-€3.52	2.8%	9%
	Maternal Calving Difficulty	-€1.73	1.3%	
	Gestation Length	-€7.49	4.1%	
	Calf Mortality	-€2.58	1.0%	
Beef	Cull Cow Weight	€0.15	0.7%	9%
	Carcass Weight	€1.38	5.1%	
	Carcass Conformation	€10.32	1.7%	
	Carcass Fat	-€11.71	1.1%	
Maintenance	Cull cow weight	-€1.65	7.2%	7%
Management	Milking Time	-€0.25	2.1%	4%
	Milking Temperament	€33.69	1.9%	
Health	Lameness	-€54.26	0.6%	3%
	SCC	-€43.49	1.8%	
	Mastitis	-€77.10	0.8%	

The ICBF in 2012 revised the sucker beef breeding index to comprise of two separate indexes: Replacement Index and Terminal Index (Table 1.2). The development of these new indexes has provided producers with tools to more accurately identify appropriate animals to maximise potential for replacement or terminal traits in beef animals.

Table 1.2 Traits, economic values and relative emphasis that constitute the Terminal Beef Index and the Replacement Beef Index.

	Traits	Economic Weight	Maternal Index Emphasis	Terminal Index Emphasis
(A) Terminal Male and female progeny of a bull destined for slaughter	Calving difficulty	-€5.27	2.6%	22.1%
	Gestation	-€1.72	0.4%	3.7%
	Mortality	-€5.34	0.4%	3.6%
	Docility	€18.40	0.3%	2.4%
	Feed Intake	-€0.16	2.2%	18.4%
	Carcass Weight	€3.78	4.1%	35.1%
	Carcass Conformation	€18.93	1.1%	9.3%
	Carcass Fat	-€10.08	0.6%	5.5%
(B) Female progeny of bull who become cows	Age at 1st Calving	-€1.65	5.0%	
	Maternal Calving Difficulty	-€2.29	4.4%	
	Maternal Weaning weight	€1.81	9.1%	
	Calving Interval	-€2.20	7.0%	
	Survival	€4.00	5.0%	
	Mature Weight Heifer Intake	-€2.28	7.8%	
	Mature Weight Cow Intake	-€0.57	7.4%	
	Cow Docility	€34.40	2.9%	
	Cull Cow Carcass Weight	€3.04	5.5%	
(C) Grandprogeny of a bulls through his daughters	Calving difficulty	-€5.27	9.7%	
	Gestation	-€1.72	1.6%	
	Mortality	-€5.34	1.6%	
	Docility	€18.40	1.0%	
	Feed Intake	-€0.16	5.7%	
	Carcass Weight	€3.78	10.2%	
	Carcass Conformation	€18.93	2.7%	
	Carcass Fat	-€10.08	1.6%	

1.1.2 Genetic trends in Irish dairy and beef performance

Genetic trends for dairy performance indicate a steady increase in the average EBI of the national dairy herd with an average dairy herd EBI of €120 in 2012 (Figure 1.3). A substantial increase in EBI gain in the dairy herd has occurred in the last four years due to the improvement in fertility performance and emphasis on genomic selection. The recent introduction of the separate maternal and terminal indexes for beef herds may help improve the downward genetic trend that has observed for the maternal beef traits.

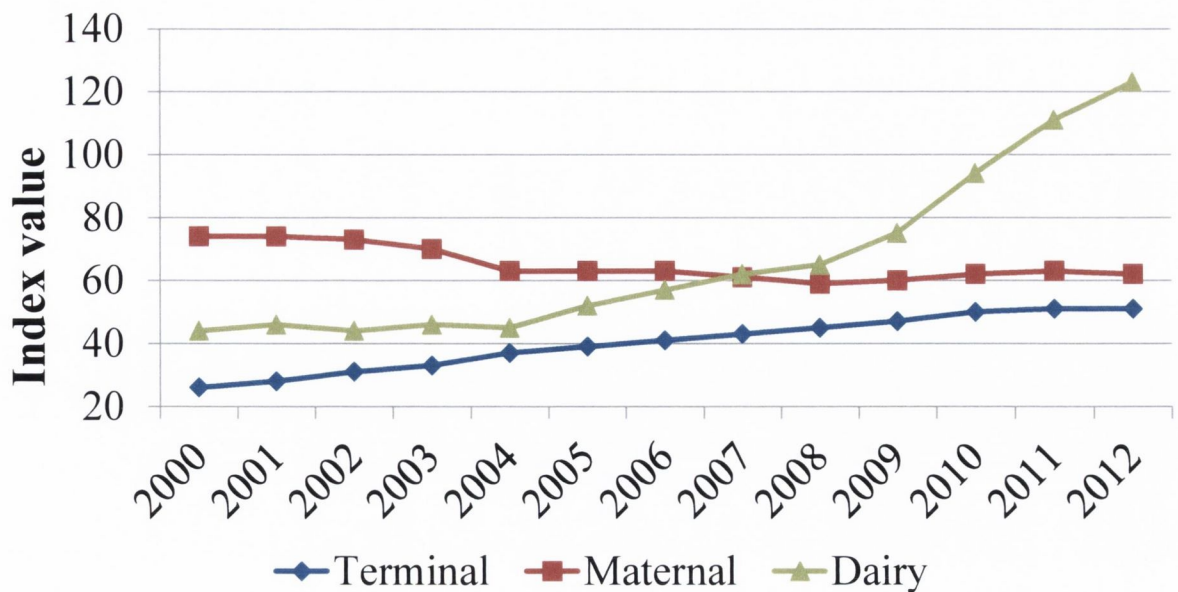


Figure 1.3 Genetic trend of the Economic Breeding Index (EBI), Terminal Beef and Maternal Beef Indexes from 2000 to 2012.(Source; <http://www.icbf.com>).

A predicted transmitting ability (PTA) is a measure of the difference in performance that can be expected in an animal’s progeny. The average breed PTAs for calving difficulty can be seen in Figure 1.4. The PTAs for calving difficulty and perinatal mortality are evaluated and published on an across breed basis allowing farmers to directly compare the PTA for different breeds. This is relevant due to the level of crossbreeding in both the dairy and commercial suckler herds (Evans and Pabiou, 2012).

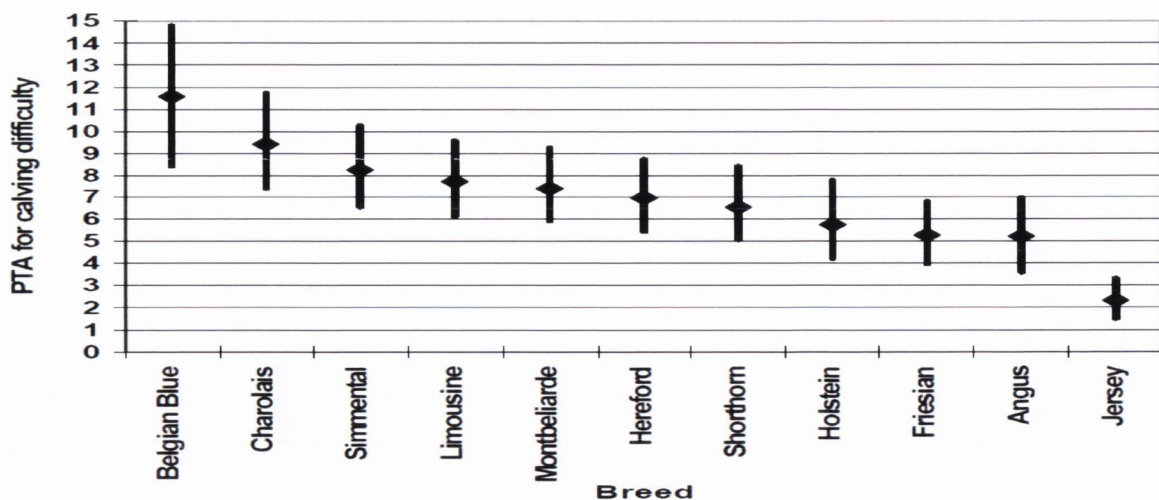


Figure 1.4 Stock chart showing the breed average, top 5 percentile and bottom 5 percentile of PTA for calving difficulty based on all pedigree animals in each breed born since 2007.

1.2. Quantitative genetics and statistical models

The phenotypic model that is the basis for animal breeding may be described as:

$$P = \mu + G + E$$

where P is the phenotypic value of an animal for a given trait, μ is the population mean or average phenotypic value for the trait of all animals in the population, G is the genotypic value of an animal for the trait and E is the effect of the environmental factors on the phenotype of the animal. A term describing how the genotype and environment interact is often ignored. Breeding values, also known as additive genetic values, reflect the genotypic value of the equation that is transmitted from parents to offspring. A breeding value is defined as sum of the additive effects of individual genes (Mrode and Thompson, 2005). Using animals with higher than average genetic merit as parents of the next generation will result in genetic gain since their progeny are, on average, expected to be genetically superior than their parental generation (Cameron, 1997). Substantial progress in recent years has been made in the selection of high genetic merit animals for traits using traditional quantitative genetics methods. This selection of superior animals has been based on a value known as the estimated breeding value (EBV) of an animal which is an estimate

of the true breeding value of an animal based on the information from the animal's phenotypic records, as well as phenotypic records on its relatives. As a PTA indicates the difference in performance that can be expected in an animal's progeny; a PTA is a half of the EBV.

1.2.1 Terminology in animal quantitative genetics

Variance

Variance is a measure of the variation within a sample population (i.e. the differences that exist among individuals). Phenotypic variation is composed of both genetic and environmental variation (Griffiths et al., 2000). Differences in genetic variation across populations can be due to the variation in alleles and their effects, can occur due to selection, inbreeding, mutations and genetic drift whereas environmental variation can arise from permanent (non-genetic) and temporary variation. Thus, phenotypic variance (σ_p^2) can be cumulatively written as:

$$\sigma_p^2 = \sigma_a^2 + \sigma_D^2 + \sigma_I^2 + \sigma_{EP}^2 + \sigma_{ET}^2$$

where σ_a^2 represents the additive genetic variation, σ_D^2 is the dominance genetic variation and σ_I^2 is the epistatic variation and σ_{EP}^2 and σ_{ET}^2 represent the permanent and temporary environmental variation, respectively.

Covariance

Covariance defines the relationship between variables. For two variables denoted X and Y , the covariance ($\hat{\sigma}_{XY}$) can be calculated as:

$$\hat{\sigma}_{XY} = \frac{\sum_i (X_i - \bar{x})(Y_i - \bar{y})}{N - 1}$$

where $\bar{x}$ and $\bar{y}$ are the respective means of the two variables, X and Y and N is the sample size. Covariance can be positive or negative and have units.

Pearson correlation coefficient

The Pearson correlation coefficient (r) is the measure of the strength of the linear relationship between two variables. This coefficient can range from -1 to +1, where 0 indicates no association between the two variables. Examples of Pearson correlation scatter plots are in Figure 1.5.

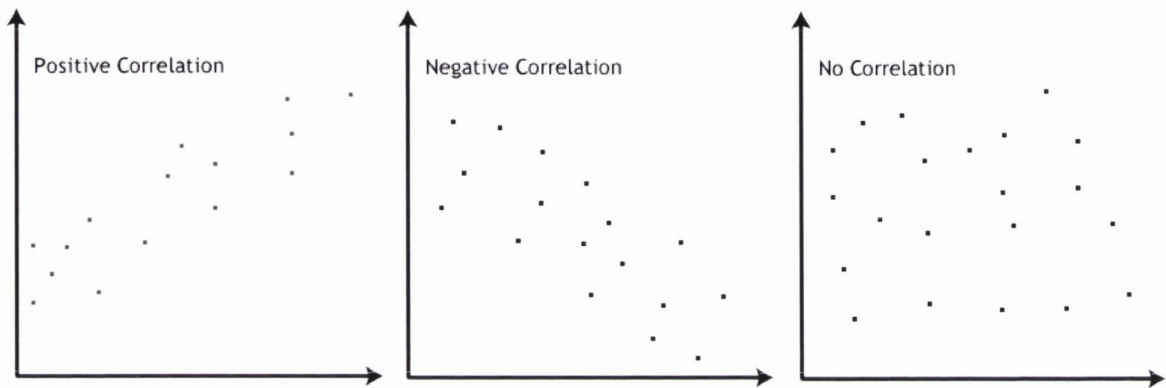


Figure 1.5 Graphic depiction of a positive, negative and no association between variables.

(Source; <http://www.statistics.laerd.com>).

The Pearson correlation coefficient can be calculated using the following formula:

$$r = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

where the covariance between the two variables, X and Y , is divided by the square root of the product of the variance of each variable. One advantage of the Pearson correlation coefficient is that it is unitless, allowing correlation coefficients between different traits to be compared without the necessity to take cognisance of the units of measurement of the traits used in the comparisons.

Regression

A regression quantifies the extent to which changes in one trait are associated with changes in another. Therefore, regression reveals the direction of the association between two traits

but it does not explain the strength of the relationship between the two traits. The linear relationship between two variables may be described as:

$$y = bX + a$$

where b is the slope of the line relating y to X and a is the intercept of that line. The regression coefficient is defined as the slope of the line (b), representing the expected change in y per unit change in X (Figure 1.6).

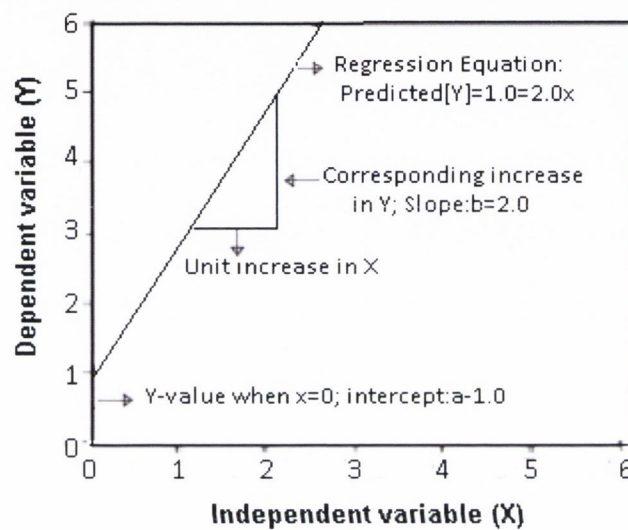


Figure 1.6 A simple linear regression model (Source; Zou et al., 2003).

The slope of the line can also be computed from the correlation coefficients using standard deviation of the two variables X and Y , as described below. If the two variables X and Y differ greatly in their standard deviations, it is possible to have a small slope and a high correlation (Zou et al., 2003).

$$b = \frac{\sigma_X}{\sigma_Y}.$$

Heritability

Heritability is defined as the ratio of genetic variance to phenotypic variance upon exclusion of the variance attributable to systematic environmental effects. Heritability is a proportion; therefore its numerical values range from 0.0 to 1.0. It is useful in quantifying the likely response to selection as well as being an indicator of the potential efficiency of

genome wide association studies (Berry et al., 2011b). When heritability of a trait is high, the measured phenotype is a good indicator of underlying true breeding value and phenotypic selection will be effective in changing the level of the trait. Broad-sense heritability (H^2) captures the proportion of phenotypic variation due to the total genetic values and includes effects due to dominance and epistasis. Broad sense heritability may be written as:

$$H^2 = \frac{V_G}{V_P} = \frac{\sigma_G^2}{\sigma_P^2}$$

The numerator of the narrow-sense heritability (h^2) is just the additive genetic variance and may be described as:

$$h^2 = \frac{\sigma_a^2}{\sigma_P^2}$$

The heritability (h^2) of milk, fat and protein yield in Holstein-Friesian cattle is moderately high ranging from 0.30 to 0.33 (More O'Ferrall, 1990, Hansen et al., 2002). Low heritability generally exists for health (Berry et al., 2011) and fertility (Berry et al., 2014) traits in cattle.

1.2.2 Genetic Evaluations

Linear models are commonly used in genetic evaluations. Two of the most common linear models are fixed effects models and mixed models. Fixed effects models contain only fixed effects while mixed models contain both fixed and random effects.

1. *Fixed effects.* Fixed effects correct for systematic differences in the population that has different mean levels. Fixed effects usually have a finite number of levels and all levels can potentially be accounted for within the model. They can be

continuous or class variables. Examples of fixed effects include gender, age and contemporary group management.

2. *Random effects*. Random effects usually have an infinite number of levels and the levels are drawn from a probability distribution. Inferences are then made to the greater population. Examples of random effects include experimental block and an individual.

1.2.2.1 Best Linear Unbiased Predicts (BLUP)

Estimates of genetic merit (i.e. EBVs) are generally estimated using Best Linear Unbiased Predictions (BLUP). Breeding values and fixed effects are simultaneously estimated in BLUP evaluations. The properties of the BLUP procedure include:

- *Best*: it maximises the correlation between the true breeding value and the predicted breeding value.
- *Linear*: predicted breeding values are linear functions of the observations.
- *Unbiased*: estimates of fixed effects are unbiased and the unknown true breeding values are distributed about the predicted breeding values.
- *Prediction*: the prediction of the true breeding value.

Over past decades, BLUP application has evolved from simple models such as sire models to more complex models such as the animal, multivariate, and random regression models (Mrode and Thompson, 2005).

1.2.2.2 Linear mixed model

A sire model is used when only EBVs of sires are of interest. The main advantage of sire models is that there is a considerable reduction in the number of equations to be

solved in comparison to an animal model; this therefore reduces the computational density. The major disadvantage of the sire model is that the genetic merit of the mate (dam) is not accounted for and all mates are therefore assumed to be genetically identical. This can result in bias in the predicted breeding values if assortative mating exists (Mrode and Thompson, 2005). A simple univariate sire model can be represented as:

$$y = Xb + Zs + e$$

where y is the vector of observations, X is the incidence matrix of the fixed effects, b is the vector of fixed effects, Z is the incidence matrix of the sire effects, s is the vector of random sire effects, and e is the vector of random residual effects.

The mixed model equation used in a simple univariate sire model, accounting for both fixed and random effects is:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix} = \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix}$$

where X is the design matrix relating records to fixed effects (number of records by number of levels of fixed effects), Z is the design matrix relating records to sires (number of records by number of random effects), A is the numerator relationship as defined below, α is the ratio of residual variance to genetic variance (σ_e^2/σ_a^2), y is the vector of phenotypic records, $\hat{b}$ is the fixed effect solutions and $\hat{a}$ is the random effect solutions (i.e the sire predicted breeding values).

The matrix which defines the additive genetic relationship among individuals is called the numerator relationship matrix (A). The prediction of breeding value requires the inverse of the relationship matrix, A^{-1} and can be calculated as;

$$A^{-1} = (T^{-1})'D^{-1}T^{-1}$$

where; A^{-1} is the inverse of the relationship matrix

T^{-1} is a lower triangular matrix with ones along the diagonal. All elements in the upper triangle are zeros. The only non zero-elements to the left of the diagonal for animal i are -0.5 for columns corresponding to their parents.

D^{-1} is the variance and covariance matrix for Mendelian sampling and is the reciprocal of the diagonal elements D computed as;

1. If both parents, sire (s) and dam (d), of animal i are known, the diagonal element of D is $d_{ii} = 0.5 - 0.25 (F_s + F_d)$
2. If one parent is known, say sire, $d_{ii} = 0.75 - 0.25 F_s$
3. If no parents are known, this is the case in the base population, then $d_{ii} = 1$.

where F_i is the inbreeding coefficient of animal i .

In an animal model, the additive genetic value of each animal is included as random effects. A simple univariate animal model may be described as;

$$y = Xb + Za + e$$

where y is the vector of observations, X is the incidence matrix of the fixed effects, b is the vector of fixed effects, Z is the incidence matrix of animal effects, a is the vector of random animal effects and e is the vector of random residual effects. Animal linear mixed models are more suitable for estimating variance components because they account for all relationships among all animals and both assortative mating and selection (Mrode and Thompson, 2005). The computational requirements are greater though as equations for the animal, pedigree and progeny need to be solved (Mrode and Thompson, 2005).

1.2.2.3 Repeatability model

A repeatability model can be used where repeated records on the same animal exist. Repeatability represents the degree of association between measurements on the same

animal for traits which are measured more than once (Mrode and Thompson, 2005). An assumption made in the definition of repeatability is that the underlying genotypic effect is the same each time the trait is measured. The repeatability animal model is an extension of the simple animal model and can be written in the form:

$$y = Xb + Za + Wpe + e$$

where y is the vector of observations, X is the incidence matrix of the fixed effects, b is the vector of fixed effects, Z is the incidence matrix of animal effects, a is the vector of random animal effects, W is the incidence matrix of permanent environmental effects, pe is the vector of random permanent environmental effects, and e is the vector of random residual effects. The inclusion of multiple records on the same animal means that a proportion of the environmental effects an animal are exposed to is repeated. The inclusion of permanent environmental effects in the repeatability model allows for the repetition of environmental variation to be accounted for. The mixed model equation (MME) for the repeatability model is:

$$\begin{bmatrix} X'X & X'Z & X'Z \\ Z'X & Z'Z + A^{-1}\alpha_1 & Z'W \\ W'X & W'Z & W'W + I\alpha_2 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix} = \begin{bmatrix} \hat{b} \\ \hat{a} \\ \widehat{pe} \end{bmatrix}$$

where $\alpha_1 = \sigma_e^2 / \sigma_a^2$ and $\alpha_2 = \sigma_e^2 / \sigma_{pe}^2$ (Mrode and Thompson, 2005).

1.2.3 Deregression of EBVs

EBVs derived from BLUP evaluations are composed of the true breeding value (BV) plus a prediction error. However, there are two main problems with EBVs that can simply be solved through their deregression.

1. BLUP is a shrinkage estimator, meaning it shrinks observations towards the mean, resulting in an inconsistent regression of the EBV on the genotype. This means

that inferior animals tend to be over evaluated and superior animals are consequently under evaluated (Garrick et al., 2009).

2. If both a parent and its progeny are genotyped, the degree of double-counting decreases as the deregressed EBV excludes ancestral information (Ostersen et al., 2011).

Many different deregression approaches exist but the method below is commonly used in Ireland (Berry et al., 2009):

$$\tilde{y} = R(R^{-1} + A^{-1})\hat{a}$$

where $\tilde{y}$ is the deregressed EBV of the animal, $\hat{a}$ is a vector of traditional BLUP evaluations, R is a diagonal matrix containing one divided by the animal's reliability from his daughters less one and A is the numerator relationship of the matrix (as described previously).

1.2.4 Inclusion of genomic data in genetic evaluations

Genomic evaluation of animals in Ireland was launched in February 2009 (Berry et al., 2009). Direct genomic breeding values, like traditional breeding values are also estimated in Ireland using mixed model equations. In genomic evaluations, however, the traditional numerator relationship matrix (A) is replaced with a genomic relationship matrix computed from thousands of single nucleotide polymorphisms (SNPs). As all animals are not genotyped within the pedigree, a combination of direct genomic values (DGVs) and traditional EBVs is currently being used to produce genomically enhanced EBVs (Berry and Kearney, 2009).

1.3 Alternative structural variants

Copy number variation (CNV) and runs of homozygosity (ROH) represent two other important sources of structural genetic variation in the cattle genome. Both of these variants may provide additional genomic structural information complementary to SNP data. Genomic structural variation, including insertions, duplications, deletions, inversions and translocations of DNA are common variants in animal genomes but little is known about the extent to which these structural variants contribute to the genetic variation in cattle.

1.3.1 Runs of homozygosity

Runs of homozygosity (ROH) are contiguous lengths of homozygous genotypes that are present in an animal due to parents transmitting identical haplotypes to their offspring. The length and quantity of ROH in an animal may be influenced by consanguinity, population size reduction, and natural selection (Kirin et al., 2010). Long ROH can be indicative of recent inbreeding in a pedigree and short ROH may be indicative of ancestral relatedness. Long ROH may also persist in out bred individuals, perhaps due to unusual mutation, LD, and recombination rates at certain genomic locations (Gibson et al., 2006).

1.3.2 Copy number variation

Copy number variation (CNV) is a form of structural variation that may result in animals differing in the number of copies of certain chromosomal segments (Fadista et al., 2010). Certain genomic segments may have been deleted or duplicated resulting in an animal possessing more or less genomic regions (Figure 1.7). CNVs are common structural variants in the cattle genome with studies suggesting that between 2 to 4% of the genome are covered in these variants (Bickhart et al., 2012, Hou et al., 2011). Studies in humans

have demonstrated CNVs to contribute to phenotypic variability including disease susceptibility (Glessner et al., 2009). Limited studies have been undertaken in cattle, but a similar CNV disease susceptibility trend has been detected, as genes associated with pathogen and parasite-resistance have been found to be heavily duplicated in Nelore cattle whilst genes involved in lipid transport and metabolism were highly duplicated in beef breeds (Bickhart et al., 2012).

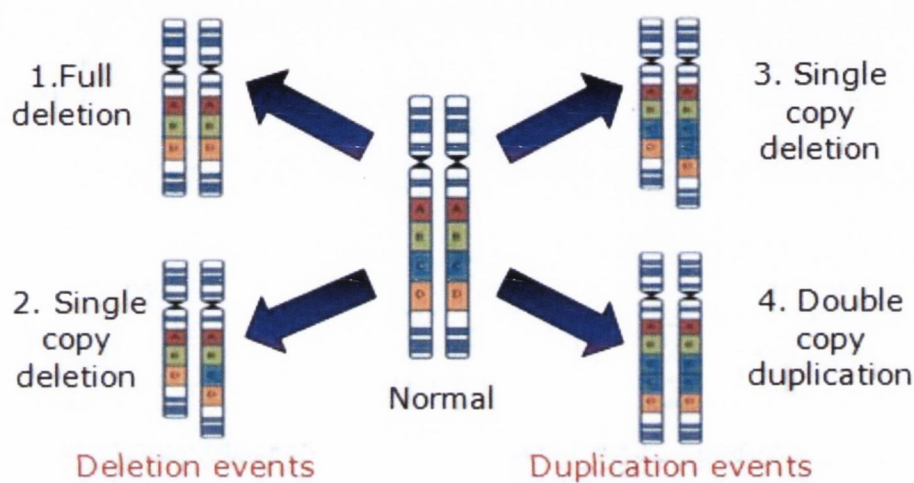


Figure 1. 7 Different types of copy number variation (CNV) found in the genome. (Source: http://readingroom.mindspec.org/?page_id=8221)

1.4. Genome Wide Association Studies

Genome wide association studies (GWAS) are used to identify regions of the genome associated with genetic differences in performance by using high-throughput genotyping technologies to assay (hundreds of) thousands single nucleotide polymorphisms (SNPs) (Pearson and Manolio, 2008). The ultimate goal of an association study is to identify candidate regions in the genome, within which genes and gene-products may contain structural variants that affect the phenotype of interest (Pearson and Manolio,

2008). Consequently, association studies help prioritise genes or genomic regions for further investigation and provide an overall description of the genetic architecture of a phenotype, such as the number of loci underlying the observed genetic (and therefore phenotypic) variation, and the distribution of effect sizes, as well as suggesting whether genetic interactions among loci (epistasis) or among traits (pleiotropy) exist (Stranger et al., 2011). GWAS are commonly used to identify genetic risk factors for diseases, which can be used for the development of new prevention and treatment strategies but also an understanding of the biology gained through these studies may help the development of new pharmacologic therapies (Bush and Moore, 2012) (Figure 1.8). Furthermore the identification of genomic regions associated with a phenotype can improve breeding as animals carrying the desirable alleles can be identified and these can be selected upon in a targeted breeding approach.

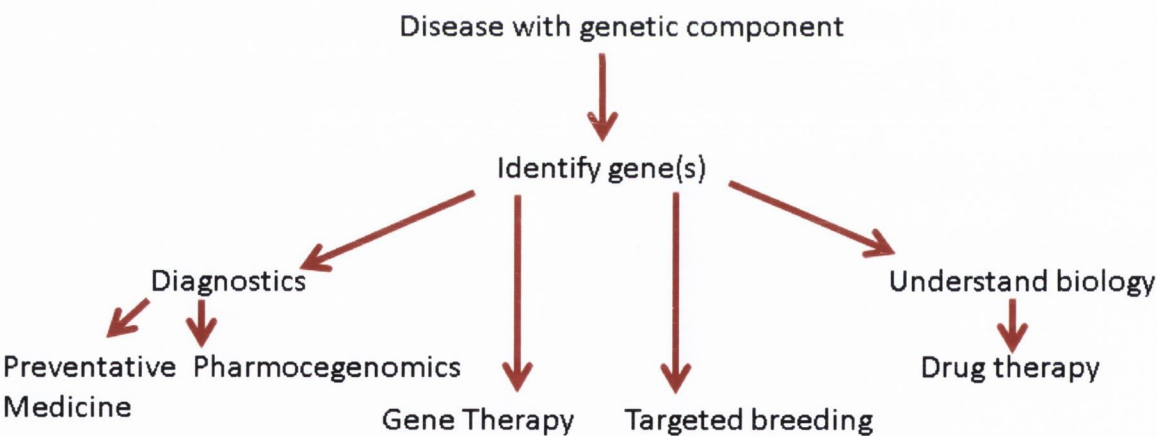


Figure 1.8 Graphical depiction of the benefits of discovering genes associated with diseases. (Source; Berry et al., 2014)

The most commonly studied genetic variants in DNA sequence are SNPs. A SNP is a single base pair change within DNA. Millions of SNPs exist within the cattle genome but only a small proportion of these SNPs are likely to be associated with phenotypes (Pearson and Manolio, 2008). GWAS is based on the premise that a causal variant is located in a haplotype, and therefore a marker allele (such as a SNP) in linkage disequilibrium (LD)

with the causal variant, should show an association with a trait of interest (Stranger et al., 2011). Linkage disequilibrium (LD) is a statistical measure of the co-occurrence of particular alleles at neighbouring SNPs and occurs due to genetic variation being passed on from generation to generation in large chunks (Bush and Moore, 2012). There are two forms of LD measure, D' and r^2 , where D' describes the normalised deviation of the haplotype frequency from linkage equilibrium and r^2 is a measure of the correlation between a pair of alleles at SNPs. Both measures can be calculated using the following formula:

$$D' = \frac{x_{11} - p_1 q_1}{\min(p_1 q_1, p_2 q_2)}$$

and

$$r^2 = \frac{(x_{11} - p_1 q_1)^2}{p_1 q_1 p_2 q_2}$$

where x_{ij} is the observed frequency of the haplotype $A_i B_j$, p_i is the frequency of A_i and q_i is the frequency of B_j . Through exploitation of LD, a reduced SNP marker set can be used to capture the majority of genomic variation that exists in the bovine genome. This has led to development of several types of SNPs arrays of varying densities that have been used to genotype sample populations.

In order to achieve sufficient statistical power in a GWAS, numerous factors must be accounted for while planning such an experiment (Allen et al., 2010);

1. *Well defined phenotypes.* Defining a phenotype is an important consideration because phenotypic heterogeneity can reduce statistical power. Accurate measures of genetic merit of sires based on their large progeny group sizes, such as EBVs (or PTAs) are commonly used in cattle association studies as they provide well-defined and highly accurate phenotypes. Also by using accurate measures of phenotypes such as EBVs, the typically large number of genotyped animals needed for lowly heritable traits can be partially overcome.

2. *The size of the study population.* This is a particularly important factor when studying lowly heritable and complex traits such as dystocia and perinatal mortality. Generally increasing the size of the study population enhances the chances of detecting low-frequency variants that have small effects on such phenotypes.
3. *The size of the effect of the variant on the phenotype.* Complex quantitative traits are likely influenced by the cumulative effect of many polymorphisms throughout the genome. An increase in the genotype density increases the statistical power to detect thousands of polymorphisms each of small effect.
4. *The extent of linkage disequilibrium (LD) between the marker allele and the causal allele.* The measure of LD between the causative allele and the genotyped allele can have an effect on study power. Power can be increased by using high marker density panels as it is thought the likelihood of the causative allele being in LD with a marker will increase. As LD in the bovine genome stretches far greater than in the human genome due to non-random mating or small effective population sizes that exist in cattle (Allen et al., 2010), the Bovine SNP50 density panel has proven to be effective in identifying genomic regions associated with genetic differences in performance particularly milk performance traits in the Holstein-Friesian population.

1.4.1 Genotyping and Single Nucleotide Polymorphism quality control

The principles of the commonly used Illumina beadchip genotyping platform is summarised in Figure 1.9. Upon the extraction of DNA from a sample, the DNA is added to the genotyping platform and following a single base extension base process carried out using fluorescently labelled nucleotides, fluorescent signals of green, red, and yellow are read from the platform by an array scanner. Using a clustering algorithm, the behaviour of

each locus is modelled and depending on the cluster shape and distance of each individual genotype to each other, a statistical score, GenTrain, is calculated for each SNP. Genotyping calls are then made by a calling algorithm, which relies on information provided by the GenTrain clustering algorithm and DNA intensity values. The GenCall score, the product of the GenTrain score and a data-to-model fit score, is a method to filter out failed individual genotypes. A GenCall score below 0.2 is an indication of a failed genotype and a score above 0.7 indicates a well behaved genotype but different thresholds can be set to exclude certain genotypes (http://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf).

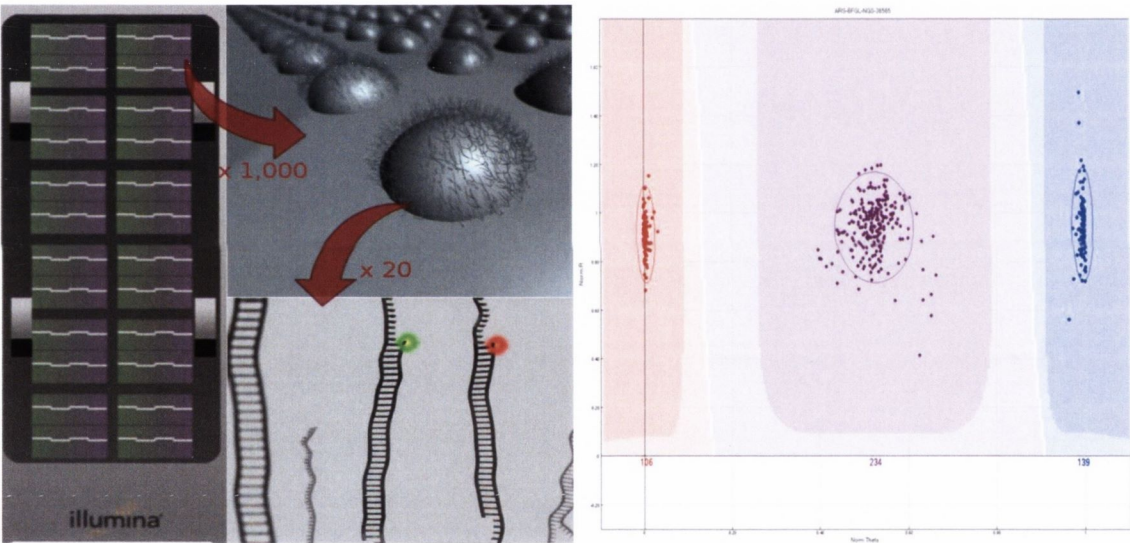


Figure 1.8 The principle behind Illumina genotyping and the GenTrain clustering method. When fragmented DNA is added to the genotyping platform, it hybridizes to the immobilised DNA on each bead and a single base extension is carried out with fluorescently labelled nucleotides. This way, the two alleles of the SNP of interest can be characterised by reading an overall fluorescence signal from the bead, where red or green means that the individual is only carrying allele A or B (interpreted as a homozygote AA or BB) and a composite yellow signal means that the individual is heterozygous, carrying both alleles A and B. The GenTrain clusters genotypes of their respective alleles and typically SNPs with a GenTrain <0.55 have failed to successfully cluster to their genotype and are removed from the analyses.

Strict quality control edits are applied to all genotype datasets prior to conducting a GWAS to limit the identification of false-positive associations. The following edits are typically applied before undertaking an association study;

1. *Removal of animals and SNPs that are poorly genotyped.* The initial step in all data analysis is to exclude samples with too much missing genotype data; samples with typically >5% missing genotypes are removed. Poorly genotyped SNPs are also removed. A typical analyses only retains SNPs with >95% genotyping rate (<5% missing) and SNPs with GenTrain scores >0.55 as below this threshold any lower there is doubts over the genotype called at this SNP. Usually, poor genotyping is indicative of poor quality DNA or a poor SNP (Pare, 2010).
2. *Hardy-Weinberg equilibrium.* The Hardy Weinberg Equilibrium states that both allele and genotype frequencies in a population remain constant from generation to generation unless the population is subjected to evolutionary forces. Standard thresholds for association studies range from $p < 10^{-4}$ to $p < 10^{-8}$ (Pare, 2010). Selection is common practise in animal breeding therefore typically low HWE cut of points ($p < 10^{-8}$) are implemented in animal association studies.
3. *Minor allele frequency.* This refers to the frequency of the less common allele at a SNP and a minimum allele frequency of between 1-5% is commonly applied to genome-wide association studies to ensure a genotyping error is not mistaken for an association.
4. *Mendelian inconsistencies.* Mendelian inconsistencies exist when an allele is present in an individual which could not have been received from its parents by Mendelian inheritance. Often in animal breeding there are many sire-son pairs genotyped allowing these opposing homozygotes to be identified and removed.

1.4.2 Statistical approaches to genome-wide association studies

Many different statistical approaches to undertake a GWAS have been developed in recent years and these can be partitioned into two main categories: a frequentist approach and a Bayesian approach. The frequentist approach focuses on probability of the data, given the hypothesis. This approach is called frequentist because it is concerned with the frequency which one expects to observe the data, given the hypothesis. Therefore, the data is treated as random and the hypothesis is fixed in a frequentist approach. Regression approaches are frequentist in nature. In contrast, the Bayesian approach focuses on the probability of the hypothesis, given the data; treating the data as fixed and the hypothesis as random (Fox, 2011).

1.4.2.1 Regression approaches to GWAS

There are two types of regression approaches used in genome-wide association studies, logistic and linear. Logistic regression is commonly used for association studies where the dependent variable is binary (e.g. an individual either has or has not succumbed to a disease), whereas linear regression is used when the dependent variable is a continuous trait.

To test for an association between a SNP and a continuous phenotype, a simple linear regression equation can be used;

$$y_i = b_m X_m + a$$

where y_i is the vector of the individuals phenotype values, a is the intercept term, X_m is a vector of the individuals genotype value for the m^{th} SNP and b_m is the regression coefficient corresponding to the m^{th} SNP (Chen and Witte, 2007). Use of the linear regression equation for every SNP will give the maximum-likelihood coefficient estimate for the association between SNP m and the phenotype. Using the t test statistic, the regression coefficient (β_m) divided by its standard error, the statistical significance can be

obtained for an association between the SNP and the phenotype. Ranking SNPs by their statistical significance identifies genomic regions of association which can be further investigated, to identify candidate genes and gene-products that may contain structural variants that affect the phenotype of interest.

An advantage of a regression approach is the ability to include interactions such as gene x environment and gene x gene interactions (Cantor et al., 2010). Also, the regression approach can be run independently so whole genome analysis is similar to a parallel process. The main disadvantage of the regression method is the high quantity of false-positive associations identified in this approach (Cantor et al., 2010). The use of large genotyping densities for association studies means that often a large proportion of SNPs are significant by chance in a regression approach. For example, assuming a 5% false discovery rate in a GWAS of 50,000 SNPs; 2,500 SNP association values are likely to be false-positive significant. As a large number of tests are performed in a GWAS, correcting for multiple hypothesis testing ensures the identification of true associations. The simplest approach to correct for multiple hypothesis testing is the Bonferroni correction, which multiplies the p-values by the number of statistical tests undertaken. Although SNPs are correlated due to linkage disequilibrium, the Bonferroni correction approach can lead to a conservative estimate of associations as it assumes each association test is independent of all other tests (Bush and Moore, 2012). Spectral decomposition of the square root of the pair-wise LD between SNP is an alternative method to correct for multiple testing and to determine the effective number of SNPs, whilst accounting for the covariance that exists (Nyholt et al., 2004).

1.4.2.2 Bayesian hierarchical modelling

Recently Bayesian multiple-regression methods have been successfully implemented for genomic selection and genomic prediction (Fernando and Garrick, 2013). Bayesian models can simultaneously fit many more SNPs than the number of observations available for the analysis (Fernando and Garrick, 2013). Therefore the Bayesian approach differs from linear regression analysis where one SNP must be fitted at a time. As a result, Bayesian models are capable of handling high density data to estimate all SNP effects simultaneously, whilst also combining both prior information in the form of prior distributions and information from the data itself for inference. A Markov Monte Carlo chain is commonly used to sample and draw inferences from the dataset (Fernando and Garrick, 2013). Prior information may include results from previous studies such as identified SNPs of association or the expected number of SNPs in the dataset that are likely to be associated with the phenotype. The prior information that a SNP is associated with a phenotype can be denoted by the value π and it can be the same for all SNPs included in the analysis or can be allowed to vary depending on the Bayesian approach taken. Information such as the MAF of a SNP, the SNPs location and evidence for selection should be taken into account when choosing a value of π for each SNP (Stephens and Balding, 2009).

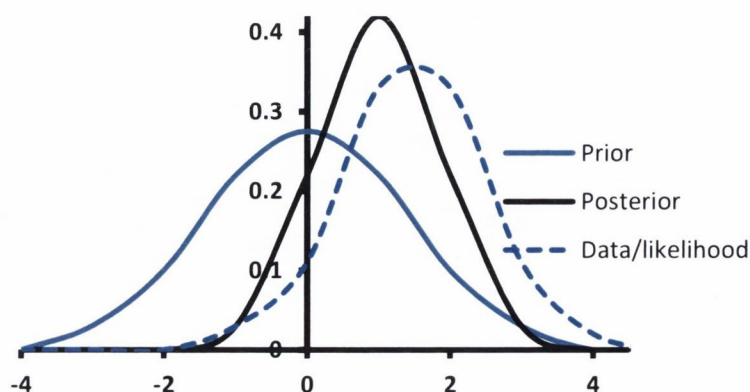


Figure 1.9 Schematic overview of the Bayesian approach incorporating both prior and likelihood information.

In comparison to regression approaches, Bayesian methods are less sensitive to false-positives due to the fact that SNPs are fitted simultaneously. Also, the ability to incorporate informative priors into a Bayesian model has resulted in the Bayesian approach becoming increasingly popular among biologists.

The Bayes Theorem, which is the basis of the Bayesian approach, gives the relationship between the probabilities of A , $P(A)$ and B , $P(B)$, and the conditional probabilities of A given B , $P(A|B)$ and B given A , $P(B|A)$ and can be written as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In terms of an association study, the posterior probability $P(A|B)$ is the conditional probability the marker is associated with the phenotype, after accounting for prior information, $P(A)$, and the likelihood from the dataset, $P(B|A)$:

$$P(A|B) \propto P(B|A) P(A)$$

In a GWAS Bayesian approach two hypotheses are compared;

1. H_0 Null model: There is no association between the SNP and the phenotype
2. H_1 Alternative model: There is an association between the SNP and the phenotype

The ratio between the probabilities of the data under the null hypothesis (H_0) and the alternative hypothesis (H_1) is known as the Bayes Factor (BF). If the BF is large then there is more support from the data for the alternative hypothesis and if the BF is equal to one then both hypothesis are deemed to be equally probable (Stephens and Balding, 2009). Bayes factors can indicate the strength of an association based on their value; a Bayes factor >3.1 is indicative of ‘substantial evidence’ that a SNP is associated with a phenotype, whereas Bayes factors >10.1 indicates ‘strong evidence’ and values greater

than >30.1 indicates ‘very strong evidence’ for association with a phenotype (Jeffreys, 1961). The BF for comparing opposing hypotheses can be calculated as:

$$BF = \frac{\frac{\Pr(H_1|y)}{\Pr(H_2|y)}}{\frac{\Pr(H_1)}{\Pr(H_2)}}$$

where y is the data, H_1 is the hypothesis that the marker is associated with a phenotype and H_2 is the hypothesis that the marker is not associated with the phenotype. The prior probability of the first hypothesis and alternative hypothesis are $\Pr(H_1)$ and $\Pr(H_2)$, respectively and $\Pr(H_1|y)$ and $\Pr(H_2|y)$ are the posterior probability of H_1 and H_2 . This can also be expressed in a simpler version as:

$$BF = \frac{\frac{\Pr(H_1|y)}{1 - \Pr(H_1|y)}}{\frac{\Pr(H_1)}{1 - \Pr(H_1)}}$$

The posterior probability of association (PPA) can be thought of as the Bayesian analogue of a frequentist p-value (Stephens and Balding, 2009). The PPA combines the evidence in the observed data (i.e., the BF) with the prior probability that the SNP is truly associated with the SNP. To calculate the PPA, the posterior odds (PO) must be calculated first as:

$$PO = BF \times \pi / (1 - \pi)$$

where BF is the bayes factor and π is the association of a SNP with the phenotype. This can be used to calculate the PPA as:

$$PPA = PO / (1 + PO)$$

A PPA value close to one means there is significant evidence for association but for this to be obtained the BF value has to be particularly large (10^4 - 10^6) (Stephens and Balding, 2009).

Several alternative Bayesian algorithms have been developed and a combination of these methods is generally used to identify genomic regions of association with the phenotype. A brief summary of some of these Bayesian alterations are detailed below:

BayesA

The BayesA method estimates the variance for each marker separately but all SNPs are assumed to contribute towards the variation in the phenotype, meaning there is a t-distribution of the SNP effects (Meuwissen et al, 2001). By allowing each SNP to have a distinct effect variance, the BayesA method will shrink SNPs of small effect towards zero and apply less shrinkage on SNPs of large effect (Meuwissen et al, 2001).

BayesB

The BayesB method assumes that many SNPs will have no association with the phenotype. This method is dependent on the prior value π given in the model and will fit a mixture distribution that assumes any SNP might come from a continuous distribution or distribution towards zero (Meuwissen et al, 2001).

BayesC

The BayesC is useful to fit as a first analysis to estimate the genotypic and residual priors as it is less sensitive to prior information than BayesB. The information from the data itself will overwhelm the prior genetic and residual parameters to give an accurate estimation of the genetic and residual priors for successive analysis.

BayesC π

The probability that a SNP is associated with a trait (π) can be estimated from the data using the BayesC π algorithm (Habier et al, 2011). This approach is commonly used as a first analysis to accurately estimate π .

BayesRS

This is a novel approach used to share information across populations for genomic predictions. It involves first mapping the variation explained by segments of the bovine genome for traits and subsequently using this information as genomic location specific priors to estimate EBVs (Brøndum et al., 2012).

1.4.3 Advantages and disadvantages of the genome-wide association approach

The main advantage of a GWAS approach is that this method scan the entire genome for genomic variation associated with a phenotype. This contrasts with the previously commonly used candidate-gene approach which focused on associations between genomic variation within pre-specified genes of interest and a phenotype. Candidate genes were most often selected based on prior biological knowledge meaning that it often failed to identify novel genes and pathways associated with the phenotype (Fan et al., 2010).

GWAS is now also the preferred method over traditional family-based linkage studies (linkage based QTL mapping). A key difference between linkage and association mapping is the precision with which they map the location of the quantitative trait loci (QTLs). A linkage analysis uses recombination events only within the recorded pedigree, therefore the confidence interval for the position of the QTL will be relatively large (~20cM) and further fine mapping is typically required to identify polymorphisms (Lynch

and Walsh, 1998). In contrast, association studies rely on LD between the QTLs and SNPs to detect polymorphisms, and as LD in cattle typically extends no further than 100kb ($r^2 > 0.15$) in the cattle genome (McKay et al., 2007), the confidence interval for the position of the QTL is generally smaller in association studies than that of a linkage analysis study (Kemper et al., 2012). Therefore, the high density of SNPs used in a GWAS approach allows smaller genomic regions of association to be identified, permits the identification of novel associations with the phenotype and no previous knowledge of the underlying biology is needed. Also there is a wide variety of computational tools readily available for analysis, allowing researchers to choose which method suits their requirements.

One major disadvantage of a GWAS approach is the ability to miss rare alleles that may be significantly associated with a phenotype. Minor allele frequency edits are often applied to genotypic data prior to analyses which may remove rare alleles of significance from the population. In addition, tag SNPs in a genotype density panel are often common SNPs that exist in the genome hence it is possible that significant genomic associations for a phenotype may have never been genotyped (Hirschhorn and Daly, 2005). One method to overcome this is to include more SNPs in the GWAS analyses to capture more of the genomic variation. Population stratification is another source of bias that may limit a GWAS approach particularly in a case-control study (Pritchard et al., 2000). If cases and controls are sampled from different subgroups of the population being studied, an association found may not be a true association with the disease but rather could be due to the underlying structure of the population (Devlin and Roeder, 1999). Also, large genotyped population sizes are often required to achieve statistical power, particularly for lowly heritable traits as seen in Figure 1.10, therefore a well-planned study can be costly with power studies suggesting that at least 2,000 to 5,000 samples are required when using general populations.

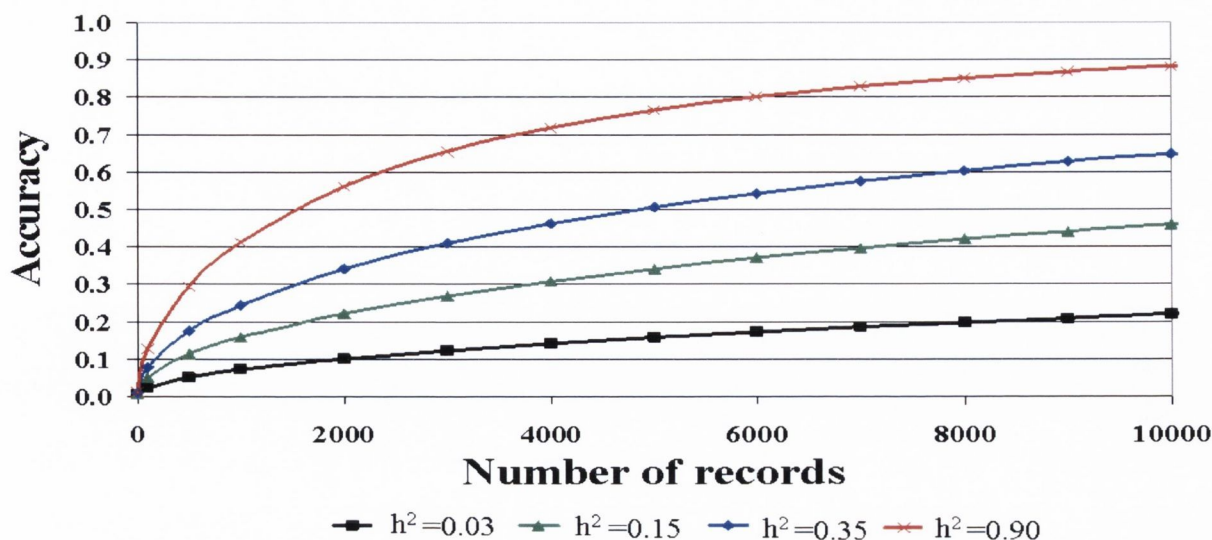


Figure 1.10 Accuracy gained in predictions versus the number of animal records for four heritabilities.

1.5 Pathway analysis

After identification of possible candidate genes from a GWAS, pathway analysis can provide another insight into the underlying biology of a complex trait. Pathways analysis has the ability to identify the genes and mechanisms that are involved in the expression of the phenotype (Khatri et al., 2012). By grouping long lists of individual genes identified in an association study into smaller sets of related genes or proteins, the complexity of an analysis can be reduced (Khatri et al., 2012). If a biological pathway is involved in the expression of the phenotype, then it is likely the associations detected in a GWAS represent candidate genes within that pathway (Szkiba et al., 2014). There are two main pathway repositories, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG).

The GO database is a collection of ontologies that describe and relate gene products in terms of their biological properties. These biological properties have been divided into three sections: cellular components, biological processes and molecular functions, with each containing a set of GO terms describing a function such as steroid metabolism (Moore

and Dunlap, 2010). Thus in pathway analysis the idea is to test whether the number of candidate genes linked to a GO term is greater than expected (enriched) (Szkiba et al., 2014).

The KEGG pathway database divides biological processes into five groups: metabolism, genetic information processing, environmental information processing (including signal transduction), cellular processes, and human diseases. By comparing to a background of all genes in the human genome over-represented KEGG pathways can be identified upon submission of a list of candidate genes.

1.6. Calving performance traits

1.6.1. Dystocia

Dystocia, commonly referred to as calving difficulty or calving ease, is broadly defined as a birth delivery requiring more assistance than desirable (Meijering, 1983). Calving difficulty may be partitioned into two components: direct calving difficulty and maternal calving difficulty. Direct calving difficulty refers to the characteristic of the calf itself (e.g., body size) while maternal calving difficulty describes the characteristics of the dam giving birth (e.g., pelvic dimensions). In Ireland calving difficulty is scored by producers on a scale of 1-4 as follows:

1. No assistance;
2. Slight assistance (assistance by one person, without needing to use a calf puller);
3. Considerable assistance (assistance by one person needing to use a calf puller or more than one person);
4. Veterinary assistance (including caesarean operations).

This scale approximates an increase in the duration and force required to deliver a calf (Mee et al., 2011). Dystocia has been documented to account for 50% of all cases of

perinatal mortality (Steinbock et al., 2003). This trait has also been reported to impact milk production (Bicalho et al., 2008), fertility (Olsen et al., 2010) and cow survival (Bicalho et al., 2007) as well as being a welfare concern; dystocia is considered one of the most painful conditions a cow can experience (Huxley and Whay, 2006). Consequently dystocia has huge economic implications for producers (Amer et al., 2001, Berry et al., 2007a, McGuirk et al., 2007).

The overall prevalence of calving assistance (i.e. calving difficulty score ≥ 2) documented in Irish Holstein-Friesian dams was 31.1% with 6.8% of calving events requiring considerable or veterinary assistance (Mee et al., 2011). Mee et al., (2011) documented that although the dystocia prevalence of 6.8% was quite high, it was considerably lower than reported twenty years previous (Crosse and Soede, 1988). A similar prevalence of dystocia was reported in beef cattle with 22.98% of calving events reporting some level of assistance and 5.93% requiring considerable or veterinary assistance (Mc Hugh et al., 2014). A detailed overview of the prevalence of dystocia in a variety of countries can be found in Table 1.3.

Many factors influence the degree of dystocia experienced during parturition including season of calving, parity of the dam, sex of the calf, twinning and cow phenotype.

1.6.2. Perinatal Mortality

Perinatal mortality in cattle is defined as calf mortality shortly before, during, or after parturition. Recent studies have highlighted that the prevalence of bovine perinatal mortality is increasing, with current prevalence statistics varying between 2% and 10% internationally (Table 1.3). In Ireland a perinatal mortality prevalence of 4.29% in grass based Holstein-Friesians was documented (Mee et al., 2008). A slightly higher prevalence of 5.48% perinatal mortality was reported in Irish beef cattle (Mc Hugh et al., 2014). The

majority of perinatal mortality (75%) occurs within 1 hour of calving with the remainder occurring either pre-partum (10%) or post-partum (15%) (Mee et al., 2008).

Perinatal mortality is a trait that is influenced by many factors, although dystocia is by far the largest contributor, accounting for 50% of mortality experienced within the first 24 hours of birth (Collery et al., 1996, Mee et al., 2008). Other significant risk factors include age at first calving, gestation length, season of birth and body condition score (Berry et al., 2007b, Johanson and Berger, 2003, Mee et al., 2008). An increase in the prevalence of perinatal mortality where no assistance was required during calving has been documented (Gustafsson et al., 2007, Mee et al., 2008). Genetic analyses suggest half the genetic variation remains in perinatal mortality after adjustment for a difficult calving (Berglund et al., 2003, Steinbock et al., 2003). Congenital defects, placental dysfunction, infection, pre-calving nutrition, or possibly other management issues such as age at first calving may be the cause of these deaths.

Lethal genetic recessive variants could also be another contributing factor to perinatal mortality where no dystocia has been observed. Currently in Ireland, there are four main deleterious mutations that breed societies test for;

1. *Complex vertebrate mutation (CVM)*. This recessive disease is characterised by misshapen and fused vertebrae around the cervico-thoracic junction due to a mutation in the SLC35A3 gene (Thomsen et al., 2006).
2. *Bovine leukocyte adhesion deficiency (BLAD)*. This disease is characterized by recurrent bacterial infections, delayed wound healing and stunted growth. The molecular basis of BLAD is a single point mutation (adenine to guanine) at position 383 of the CD18 gene (Nagahata, 2004).
3. *Deficiency of uridine monophosphate synthase (DUMPS)*. DUMPS is caused by a mutation of uridine monophosphate synthetase (UMPS) protein, resulting in pyrimidine deficiencies. The mutation leads to the loss of an *Ava*I site at codon 405,

resulting in a premature stop codon with a truncated C-terminal catalytic subunit of the proteinuridine monophosphatesynthetase (Schwenger et al., 1993).

4. *Brachyspinia*. Affected animals are characterised by severely reduced body weight, growth retardation, extensive vertebral malformations causing a significant shortening of the spine (brachyspina) and long and slender limbs. This defect is caused by a 3.3kb deletion in the FANCI gene which is essential for DNA interstrand crosslink repair (Charlier et al., 2012).

Table 1.3 International prevalence rates for dystocia (defined as considerable or veterinary assistance) and perinatal mortality (defined as the death of a newborn calf within the first 24hrs of parturition) in primiparous (primi) and multiparous (multi) dams.

Country	Breed	Dystocia prevalence		Mortality prevalence		Reference
		Prim	Multi	Prim	Multi	
America	Holstein X	16.4%	8.4%	15.1%	12.7%	Heins et al., (2006)
	Brown Swiss X	12.5%	4.9%	11.6%	5.6%	Heins et al., (2006)
	Normande cross	N/A	8.7%	N/A	7.3%	Heins et al., (2006)
	Montbeliarde X	11.6%	5.4%	12.7%	5.0%	Heins et al., (2006)
	Scandinavian Red X	5.5%	2.1%	7.7%	4.7%	Heins et al., (2006)
Australia	Holstein-Friesian	9.5%	4.1%	7.4%	7.1%	McClintock (2003), Pryce et al., (2006)
Germany	Holstein	N/A	2.66%	N/A	9.7%	Popescue et al.,(2013), Gundelach et al.,(2009)
Iran	Holstein	19.3%	7.0%	7.2%	3.3%	Hosseini-Zadeh (2014)
Ireland	Holstein-Friesian	9.3%	5.8%	7.7 %	3.5%	Mee et al., (2008,2011)
France	Holstein-Friesian	N/A	6.6%	6.9%	N/A	Fourichon et al.,(2001), Chassagne et al., (1999)
Norway	Norwegian red	N/A	2.14%	N/A	3.4%	Gulliksen et al., (2009)
Sweden	Charolais	6.6%	1.0%	5.9%	1.8%	Eriksson et al., (2004)
	Hereford	6.2%	1.2%	5.6%	1.8%	Eriksson et al., (2004)
	Holstein	8.3%	4.5%	7.1%	2.7%	Steinbock et al., (2003)
United Kingdom	Holstein-Friesian	4.0%	1.89%	11.6%	4.3%	Eaglen et al., (2012)

1.6.3 Risk factors associated with calving performance traits

Dystocia and perinatal mortality are both traits influenced by direct and maternal genetic components (Berger et al., 1992); that is the genetic makeup of both the calf and the dam will impact on the birth process and its outcome (Meijering, 1983).

1.6.3.1. Effects of the calf

A variety of animal level factors influence the direct component of dystocia and perinatal mortality mainly reflecting the size and shape of the calf.

Birthweight

Birthweight has long been known as a risk factor for dystocia and perinatal mortality; calves lighter and heavier than the average tend to have an increased risk of both phenomena (Berger et al., 1992). A study carried out by Johanson and Berger (2003) in Holstein cattle showed that if the birthweight of a calf exceeded the trial average of 40.3kg there was an exponential increase in the risk of mortality; a calf birthweight of 52kg had 9.6 percentage unit increase in the probability of perinatal mortality. Dam parity also influences birthweight as primiparous cows tend to have smaller calves (38.2 kg) than multiparous cows (41.7 kg) (Johanson and Berger, 2003). Heavier birthweights have also been reported in calves born from a longer than average gestation length (Johanson and Berger, 2003). Data on calf birthweight is currently lacking in Ireland. Birthweight is a moderately heritable trait for the direct component with estimates ranging from 0.44 to 0.60 in both dairy and beef cattle (Eriksson et al., 2004, Plum et al., 1965).

Gender

Calf gender is also associated with calving performance traits with male calves, on average, experiencing greater dystocia and mortality due to their bigger size and greater birth weight in comparison to female calves (Meijering, 1983). In Ireland, Mee et al (2008)

reported 1.12 greater odds of perinatal mortality in male calves than in females. Male calves have been documented to have longer gestation lengths which can also increase the risk of dystocia (Newman et al., 1993). Calf shape also differs between genders with broader and heavier skeletons observed in males calves which may increase the risk of dystocia and mortality due to incompatible feto-pelvic dimensions (Meijering, 1983). Secondary sex ratio, which is the sex ratio at birth has been shown to be lowly heritable in Irish cattle populations with a direct heritability estimate of 0.02 (Berry et al., 2011).

1.6.3.2 Effects of the dam

Cow pelvic area

Cow pelvic area defined as the height of the pelvic inlet multiplied by its width can impact dam dystocia (Meijering, 1983). A 1dm² (decimetre) greater pelvic area has been associated with a 11.0 percentage units decrease in the odds of experiencing dystocia in Holstein-Friesian dams (Johanson and Berger, 2003). Pelvic area differs by breed and one study focusing on Angus and Hereford breeds suggests that lower cow body weight was related to a more narrow pelvic opening (Bellows et al., 1971). Cow pelvic area is predominantly of concern in first parity animals, although the pelvic area itself explained only 10% of the variance associated with dystocia (Luo et al., 2002). Cow pelvic area is a moderately heritable trait with estimates ranging from 0.04 to 0.36 (Kriese et al., 1994).

Body weight and body condition score

Cow body weight and body condition score (BCS) have been shown to impact dystocia and perinatal mortality (Chassagne et al., 1999, Erb et al., 1985). Over conditioned cows (BCS >4) have an increased risk of dystocia and stillbirth at parturition due to a reduction in the dam pelvic area as a result of the accumulation of fat in the pelvic region (Chassagne et al., 1999). Both body weight and body condition score are moderately heritable with

estimates ranging from 0.39 to 0.5 for body weight and 0.27 to 0.37 for body condition score (Berry et al., 2002).

Twinning

Twinning may be considered a characteristic of the dam, as the dam may ovulate numerous eggs thus increasing the chances of multiple birth events. Twinning is often perceived as an unfavourable phenotype; greater incidence of abortion, abnormal presentation, retained placenta and metritis are associated with multiple births (Nielen et al., 1989). Twinning also impacts negatively on production by reducing fertility, milk production and survival (Nielen et al., 1989). Multiple ovulations were documented as lowly heritable (0.03) but moderately repeatable (0.33) in Irish dairy and beef herds with a heritability estimate of 0.02 for twinning (Fitzgerald et al., 2014),

Gestation length

Gestation length is non-linearly associated with perinatal mortality with a greater risk of mortality reported after long (>278 days) as well as short (<267 days) gestation periods (Philipsson et al., 1976, Hansen et al., 2004). Maternal heritability estimates for gestation length range from 0.07 to 0.13 (Azzam and Nielsen., 1987, Hansen et al., 2004, Jamrozik et al., 2005).

Age at first calving

Young calving heifers have an increased likelihood of requiring calving assistance; heifers calving at 22 months of age required more assistance than those calving at 24 months (Berry and Cromie, 2009).

Parity

Primiparous dams are increasingly likely to experience dystocia and perinatal mortality with a 4.7 greater risk of dystocia reported in primiparous dams than multiparous dams (Berry et al., 2007, Johanson and Berger., 2003).

Season of calving

Season of calving also impacts calving performance with increased levels of perinatal detected in winter months (Johanson et al., 2001, McGuirk et al., 1999, Mee et al., 2008).

1.6.3.3. Genetic parameters of calving performance

Heritability estimates for calving difficulty and perinatal mortality differ widely internationally, but estimates are generally low (Table 1.4). Heritability estimates for the direct calving difficulty component range from <0.01 to 0.17 and from <0.01 to 0.12 for the maternal component of calving difficulty (Eaglen et al., 2012, Hickey et al., 2007, Steinbock et al., 2003). Heritability estimates tend to slightly lower for perinatal mortality ranging from <0.01 to 0.12 for the direct component and from <0.01 to 0.08 for the maternal component (Eriksson et al., 2004, Steinbock et al., 2003). The difference in genetic parameter estimates of perinatal mortality and dystocia can be attributed to differences between breeds (Heins et al., 2006), management systems and different statistical models (Erkisson et al., 2004, Hickey et al., 2007, Johanson et al., 2011, Steinbock et al., 2003). A negative covariance between direct and maternal dystocia components has been previously reported (Eaglen et al., 2012, Eriksson et al., 2004, Steinbock et al., 2012). The negative covariance may be due to small easy calved females who are likely to experience dystocia at maturity due to their reduced size and pelvic opening dimensions (Philipsson et al, 1976).

Table 1.4 International direct and maternal heritability estimates for dystocia and perinatal mortality.

Country	Breed	Parity	Dystocia		Perinatal Mortality		Reference
			Direct	Maternal	Direct	Maternal	
America	Holstein	All parities	0.11	0.14	0.13	0.15	Johanson et al., (2011)
Australia	Holstein-Friesian,	Primiparous	0.11	N/A	0.013	N/A	McClintock et al., (2003), Pryce et al., (2006)
	Jersey and crosses	Multiparous	0.04	N/A	0.010	N/A	
Canada	Holstein	Primiparous	0.14	0.08	N/A	N/A	Luo et al., (2002)
		Multiparous	0.10	0.04	N/A	N/A	Luo et al., (2002)
Ireland	Holstein-Friesian	Primiparous	0.130	0.040	0.02	N/A	Hickey et al., (2007), ICBF., (2006)
France	Charolais	All parities	0.13	0.012	N/A	N/A	Phocas et al.,(2004)
	Limousin	All parities	0.10	0.08	N/A	N/A	Phocas et al., (2004)
	Blonde d'Aquitaine	All parities	0.09	0.09	N/A	N/A	Phocas et al., (2004)
	Maine-Anjou	All parities	0.09	0.09	N/A	N/A	Phocas et al., (2004)
Sweden	Charolais	Primiparous	0.127	0.074	0.002	0.005	Eriksson et al., (2004)
		Multiparous	0.048	0.023	0.010	0.005	Eriksson et al., (2004)
	Hereford	Primiparous	0.164	0.110	0.004	0.001	Eriksson et al., (2004)
		Multiparous	0.021	0.004	0.006	0.003	Eriksson et al., (2004)
	Holstein	Primiparous	0.062	0.048	0.038	0.028	Steinbock et al., (2003)
		Multiparous	0.004	0.002	0.007	0.003	Steinbock et al., (2003)
United Kingdom	Holstein-Friesian	Primiparous	0.120	0.050	0.020	0.030	Eaglen et al., (2012)
		Multiparous	0.030	0.020	0.020	0.020	Eaglen et al., (2012)

1.7. Genome-wide association studies for calving performance

Dystocia and perinatal mortality are genetically complex traits influenced by both direct and maternal effects. Several studies have documented QTL associated with calving performance in Holstein populations (Schroteen et al., 2000, Kühn et al., 2003, Schnabel et al 2005) although the majority of QTL locations differed by population. This may be due to a difference in calving trait definitions, recording methods and statistical approaches but also possibly due to different allele frequencies segregating in various subpopulations (Seidenspinner al., 2009).

Several studies have reported a QTL associated with calving performance on chromosome 18. Kühn et al., (2003) reported a QTL associated with maternal dystocia and maternal stillbirth on chromosome 18 in the German Holstein population. Holmberg and Andersson-Eklund (2006) reported a QTL associated with maternal calving ease in the Swedish Reds and Swedish Holsteins, and Thomasen et al., (2008) documented a QTL associated with dystocia and stillbirth in Danish Holsteins. Cole et al., (2009) was the first to propose that a single QTL on chromosome 18 may be the cause of these calving performance observations. In this single QTL, the SNP, ss86324977, on chromosome 18 had the strongest association with several traits including sire and daughter calving ease, rump width, stature and body depth in a US Holstein population (Cole et al., 2009). Two studies in Danish and Swedish Holstein populations also documented the SNP ss86324977 to be significantly associated with calving difficulty (Höglund et al., 2012, Sahana et al., 2011). This SNP, ss86324977, is located within an intron of the sialic acid binding Ig-like lectin 12-like gene (Siglec-5). In humans, a similar gene, Siglec-6, is expressed in the placenta and may play a role in the initiation of parturition (Brinkman-Van der Linden et al., 2007). Cole et al (2009) hypothesized that animals homozygous 'AA' for the SNP

ss86324977 may experience a delayed parturition due to a lectin deficiency caused by high levels of sialic acid binding Ig-like lectin.

Genomic associations for calving performance have also been documented on chromosome 10 (Kühn et al., 2003, Schnabel et al., 2005) suggesting there is a causal mutation between the 87.75Mb to 90.75Mb region affecting calving difficulty (Seidenspinner et al., 2011). This 3Mb region on chromosome 10 was further fine-mapped to reveal three strong candidate genes; LTBP2 (encoding latent transforming growth factor- β -binding protein 2), ESRRB (encoding oestrogen-related receptor β) and PGF (encoding placental growth factor 1). The bovine placental growth factor encoding gene (PGF) was chosen as a potential candidate gene for the maternal effect of dystocia and perinatal mortality by Seidenspinner et al. (2011). Bovine placental growth factor was chosen as it is involved in the modulation of angiogenesis, vessel remodelling and vascular permeability during implantation and placentation (Seidenspinner et al., 2011). Within the *PGF* gene, 37 SNPs and two insertions/deletions were identified of which 17 SNPs were significantly associated with maternal stillbirth and maternal calving difficulty (Seidenspinner et al., 2011). A total of 8% to 14% of the additive genetic variation of maternal stillbirth and 8% to 15% of maternal calving ease was explained by allele substitutions in the 17 significant SNPs in the *PGF* gene (Seidenspinner et al., 2011).

Chromosome 6 has also been implicated in multiple studies to be associated with calving performance (Schrooten et al., 2004, Olsen et al., 2010, Biogorni et al., 2012). In the GWAS study by Olsen et al., (2010) in the Norweigen Red population, chromosome 6 had the strongest associations with calving performance; five SNPs between the 30Mb to 47Mb region were associated with the direct effect of dystocia (Olsen et al., 2010). This QTL region on chromosome 6 was further refined using linkage disequilibrium linkage analysis (LDLA) to identify two QTL associated with the direct effect of dystocia and one QTL associated with the direct effect of perinatal mortality (Olsen et al., 2010). Several

candidate genes (*SPP1*, *IBSP* and *MEPE*) known to affect bone and cartilage formation were identified within this region (Olsen et al., 2010). Also a study by Bongiorno et al., (2012) in Piedmontese cattle reported a QTL associated with the direct effect of dystocia between the 37.8 and 38.7Mb region on chromosome 6. Three genes (*LAP3*, *NCAPG* and *LCORL*) were identified within this 900kb region and to confirm their effect on dystocia, additional SNPs within these three genes were analysed for association. The gene *LAP3* which is involved in oxytocin metabolism was the most probable candidate gene after the association analyses of the additional SNPs although it was found to be in very strong LD with *LCORL* which has been shown to control stature in cattle (Bongiorno et al., 2012). These studies suggest chromosome 6 is associated with the direct effect of dystocia via an impact on calf morphology and calf shape.

Association studies have also been carried out on many genetic factors that contribute to calving performance such as birthweight, body size, bone percentage and gestation length (Ashwell et al., 2005, Gutierrez-Gil et al., 2009, Maltecca et al., 2011, McClure et al., 2010). Overlap between calving performance associations and associations with these genetic factors is to be expected as they strongly influence calving performance.

1.8. Gaps in Knowledge

Gaps in knowledge that will be examined as part of this thesis include:

- Estimates of genetic parameters for dystocia and perinatal mortality in Ireland by more recent data. No previous study has attempted to quantify the genetic variation in perinatal mortality not associated with dystocia.
- The size and distribution of SNP effects on dystocia and perinatal mortality in both dairy and beef cattle.
- The identification of possible candidate genes associated with calving performance, as well as the identification of prominent underlying biological pathways.
- The analysis of whole genome sequence data to finemap significant associations identified through GWAS for dystocia and perinatal mortality.
- The assessment of amino acid substitutions on protein function for all significant missense variants associated with calving performance.
- The validation of runs of homozygosity as a predictor of inbreeding in cattle species.
- The analysis of alternative genetic variants such as runs of homozygosity and copy number variants in dairy and beef cattle breeds and their impact on calving performance.

Chapter 2

Genetic parameters of calving performance in Irish dairy and beef cows

2.1 Preface

Deirdre Purfield was the primary author who edited the data, calculated the prevalence and heritability estimates and drafted the manuscript. Donagh Berry and Daniel Bradley conceived the study, participated in the design and co-ordination of this study and helped draft the manuscript. Ross Evans provided the data which the analysis was completed on.

2.2 Abstract

The aim of this study was to estimate the direct and maternal genetic parameters for dystocia, perinatal mortality, and perinatal mortality not associated with dystocia in primiparous and multiparous dams for Irish dairy and beef cows. Univariate repeatability animal-dam models were used to estimate the variance components using 100,664 birth events. The overall prevalence of calving assistance and dystocia was 20.33% and 4.29% in primiparous dams and 14.19% and 2.61% in multiparous dams, respectively. The prevalence of perinatal mortality was 4.15% in primiparous dams and 1.79% in multiparous dams. The heritability of dystocia in primiparous dams was 0.21 and 0.03 for the direct and maternal effects, respectively and 0.10 and 0.01 for the direct and maternal effects in multiparous dams. Heritability estimates for perinatal mortality were 0.01 and 0.02 for the direct and maternal effects in primiparous dams and 0.001 for both the direct and maternal effect in multiparous dams. Perinatal mortality not associated dystocia had a prevalence of 1.57% in primiparous dams and 0.96% in multiparous dams. Only the direct component of perinatal mortality not associated with dystocia was significantly different from zero (direct h^2 of 0.019) in primiparous dams. The genetic correlation between the direct and maternal component of dystocia was negative across all parities (primiparous - 0.34, multiparous -0.68). Reasonable genetic variation existed for all traits suggesting that breeding programs should continue to exploit this variation to improve calving performance within the population.

2.3 Introduction

Calving is an important event on both dairy and beef farms as a successful calving event is essential for economic success. Complications during a calving event can lead to impaired performance of a dam, loss of the dam or calf, and can compromise animal welfare (Berry et al., 2007b, Bicalho et al., 2008, Olsen et al., 2010). In recent years, functional traits have become increasingly important as producer emphasis has shifted from traits that increase revenue, towards traits that reduce costs (Eaglen et al., 2012). Functional traits such as dystocia, commonly known as calving difficulty, and perinatal mortality are regularly included in breeding programs to improve calving performance through genetic selection methods. Nevertheless, recent studies have highlighted that the prevalence of both perinatal mortality and dystocia is increasing, with current population prevalence levels varying between 2% and 10% internationally for perinatal mortality and between 2% and 14% for serious cases of dystocia (Jamrozik et al., 2005, McGuirk et al., 1999, Meyer et al., 2001, Pryce et al., 2006, Steinbock et al., 2003). In Ireland, a prevalence of 6.8% for dystocia (Mee et al., 2011) and 4.29% for perinatal mortality (Mee et al., 2008) was documented in seasonal calving grazing Holstein-Friesian dams. A similar prevalence of dystocia was reported in beef cattle with 5.93% requiring considerable or veterinary assistance and a prevalence of 5.48% for perinatal mortality (Mc Hugh et al., 2014).

Perinatal mortality and dystocia are influenced by both direct and maternal genetic components (Meijering, 1983, Philipsson, 1979). Heritability estimates for the additive genetic component of dystocia range from <0.001 to 0.17, while the maternal heritability for dystocia ranges from <0.01 to 0.12 (Eriksson et al., 2004, Hickey et al., 2007., Steinbock et al., 2003). Heritability estimates for perinatal mortality for both direct and maternal are lower ranging from <0.01 to 0.12 (Eriksson et al., 2004, Johanson et al., 2011, Steinbock et al., 2003). The difference in genetic parameter estimates of perinatal mortality

and dystocia can be attributed to differences between breeds (Heins et al., 2006), management systems and different statistical models (Berry et al., 2011a, Eriksson et al., 2004, Johanson et al., 2011, Steinbock et al., 2003).

Dystocia is one of the main contributing factors to perinatal mortality with over 50% of stillborn calves experiencing some level of calving difficulty (Collery et al., 1996). Other significant risk factors for perinatal mortality include age at calving, parity of dam, gestation length, season of birth and body condition score (Berry et al., 2007b, Johanson and Berger, 2003, Mee et al., 2008, Meyer et al., 2000). An increase in the prevalence of cases of perinatal mortality where no assistance was required at birth has been reported (Gustafsson et al., 2007, Mee et al., 2008). Previous genetic analysis of perinatal mortality suggest half the genetic variation in perinatal mortality remains after phenotypic adjusting for calving difficulty in the model, suggesting a large proportion of the variation in perinatal mortality is to some extent phenotypically independent of dystocia (Berglund et al., 2003, Steinbock et al., 2003).

The objective of this study was therefore to estimate genetic parameter for dystocia and perinatal mortality with particular emphasis on a trait describing perinatal mortality which was not associated with a dystocia (PMND) event in Irish beef and dairy herds. The analysis was undertaken in both primiparous and multiparous animals separately.

2.4 Methods

2.4.1 Data edits

Data on calf, sire, dam, maternal grand sire and herd identification number, as well as calving date, age and parity of dam, gender of calf, degree of calving assistance, and occurrence of perinatal mortality (recorded as calf dead at birth or within 24hrs) were available from the Irish Cattle Breeding Federation database between the years 2008 to

2012, inclusive. In total 3,225,562 calving records were available where the sire and maternal grandsire of the calf were known. Of these records, 2,630,037 had information on the degree of calving assistance required at birth and 1,824,729 had information on perinatal mortality. Animals were from a multitude of different dairy and beef breeds and crossbreds. Dams of unknown parity and parities >10 were discarded ($n=27,417$ records discarded), as well as calvings with more than two calves (1,426 calving events removed). Cows that calved for the first time less than 608 days of age ($n=6,567$) as well as animals calving more than 2 yrs from the median, within parity ($n=144,371$), were discarded. Dam parity was recoded into five classes; 1, 2, 3, 4, and ≥ 5 . Parity number of the dam in this study was the parity of the cow the day after she calved.

In Ireland, calving difficulty is subjectively scored by producers on a linear scale of 1 to 4, where 1=no calving assistance /unobserved, 2=slight assistance; 3=severe assistance and 4=veterinary assistance (including caesarean). Severe dystocia in this study refers to scale 3 and 4. Perinatal mortality is recorded as a binary variable by producers, indicating whether the calf was dead within a 24 hour period. Information on a novel trait, perinatal mortality not associated with a dystocia event was created for the present study. This trait (PMND) was defined as 1 if perinatal mortality occurred where no calving assistance of any kind was required (i.e., calving dystocia score of 1) and was defined as 0 if no perinatal mortality occurred and set to missing if perinatal mortality was recorded with a dystocia recording of ≥ 2 .

An algorithm was used to generate contemporary groups based on grouping dams, within herd, that calved in close proximity (Crump et al., 1997). If this initial grouping of dams had less than 10 records, it was merged with an adjacent group from the same herd if they were within 182 days of each other. For each of the three calving traits investigated (i.e. dystocia, perinatal mortality, and PMND) contemporary group was defined separately

for primiparous and multiparous dams. Only contemporary groups with at least 5 records were retained. A random selection of contemporary groups from dairy and beef herds were retained for the estimation of variance components. The number of herds, dams, sires and maternal grandsires and other descriptive statistics of each dataset are detailed in Table 2.1.

Table 2.1 Number of birth events, herds, sires, maternal grandsires (MGS) and contemporary groups (CG) in each dataset

	First Parity	Later Parities
Number of records	33,526	67,138
Number of dams	33,526	40,045
Number of sires of the calf	3,163	3,674
Number of MGS	3,572	4,695
Number of Herds	1238	937
Number of dystocia CG records	2,714	5,456
Number of stillbirth CG records	2,514	5,465
Number of still birth no dystocia CG records	2,439	5,386

2.3.2 Data Analysis

Univariate repeatability animal-dam models in ASreml (Gilmour et al., 2008) were used to estimate variance components for dystocia, perinatal mortality, and PMND. Fixed effects included in the models were contemporary group, parity of the dam (for the multiparous cows), age of the dam in months nested within parity, calf gender, heterosis coefficient of both the calf and the dam, and recombination loss coefficient of both the calf and the dam. Random effects considered for inclusion in the models were the additive genetic effect of the calf, the maternal genetic effect and the dam permanent environmental effect; the latter was only considered in the multiparous models where repeated records per

dam existed. A covariance between the direct and maternal components was also estimated for all traits. The pedigree of all animals was traced back to the founder population where founder animals were allocated to breed groups based on breed. The pedigree consisted of 306,273 animals. All variance components were estimated separately for primiparous, multiparous and combination of both categories of animals.

2.5 Results

2.5.1 Prevalence statistics

Table 2.2 presents the prevalence of dystocia, perinatal mortality and PMND in primiparous and multiparous animals. The prevalence of all calving performance traits was greater in primiparous animals. The likelihood of severe dystocia (levels 3 and 4) was greatest for male calves; prevalence of dystocia in primiparous dams with male calves (5.7%) was nearly double than females calves (2.98%) with a similar trend in multiparous dams (3.37% and 1.81%, respectively). A similar gender effect also existed for perinatal mortality with a greater prevalence of mortality in male calves (primiparous 4.44%; multiparous 2.03%) in comparison to female calves (primiparous 3.87%; multiparous 1.54%). Perinatal mortality not associated with dystocia was slightly greater in female (1.68%) than male calves (1.45%) in primiparous dams. Some level of assistance (calving assistance score =2) was reported in 27.38% and 18.83% incidences of perinatal mortality in primiparous and multiparous dams, respectively whereas severe dystocia (calving assistance score ≥ 3) was reported in 24.21% and 21.85% of mortalities in primiparous and multiparous dams, respectively.

Table 2.2 Prevalence of dystocia (scored 1 to 4), perinatal mortality and perinatal mortality not associated with dystocia.

Trait	Dataset	Prevalence			
		1	2	3	4
Dystocia	Primiparous	75.38	20.33	3.16	1.13
	Multiparous	83.19	14.19	1.80	0.81
	Combined	80.79	16.08	2.22	0.91
Perinatal Mortality	Primiparous	4.15			
	Multiparous	1.79			
	Combined	2.49			
PMND	Primiparous	1.57			
	Multiparous	0.96			
	Combined	1.14			

2.5.2 Genetic Parameters

Genetic parameters from the univariate analyses for all three calving performance traits in primiparous, multiparous and the combination of all parities are detailed in Table 2.3. The heritability estimates of all traits were <0.02 with the exception of direct (primarous 0.21; multiparous 0.10) and maternal dystocia in primiparous dams (0.03). Direct and maternal heritability estimates for perinatal mortality were low in both primiparous and multiparous; only the maternal genetic effect in primiparous animals was different from zero (0.02 s.e 0.007). Perinatal mortality not associated with dystocia was only heritable for the direct effect in primiparous animals (0.02 s.e. 0.009). The repeatability of calving difficulty was 0.12 in mutliparous dams and 0.15 across all parities. The repeatability of perinatal

mortality and PMND was much lower in multiparous dams, 0.003 and 0.004, respectively and 0.015 and 0.006, respectively across all parities.

Table 2.3 Heritability estimates (direct and maternal) and genetic standard deviation for dystocia (σ_g), perinatal mortality and perinatal mortality not associated with dystocia (PMND) in primiparous, multiparous and a combination of both categories of animals, estimated using univariate analysis.

Trait		Direct		Maternal	
		Heritability	σ_g	Heritability	σ_g
Dystocia	Primiparous	0.207 (0.026)	0.226	0.034 (0.013)	0.092
	Multiparous	0.104 (0.012)	0.130	0.008 (0.004)	0.036
	Combined	0.131(0.011)	0.158	0.015 (0.004)	0.053
Perinatal	Primiparous	0.007 (0.008)	0.017	0.017 (0.007)	0.026
Mortality	Multiparous	0.001 (0.002)	0.005	0.001 (0.001)	0.005
	Combined	0.002 (0.002)	0.006	0.005 (0.002)	0.011
PMND	Primiparous	0.019 (0.009)	0.017	0.000 (0.000)	0.000
	Multiparous	0.000 (0.000)	0.000	0.000 (0.000)	0.002
	Combined	0.006 (0.001)	0.003	0.000 (0.000)	0.000

2.5.3 Genetic correlations

The correlation between the direct and maternal component for dystocia was negative with a correlation of -0.34 (s.e. 0.138) in primiparous dams, -0.68 (s.e 0.079) in multiparous dams and -0.53 (s.e. 0.074) across all parities. The correlation between the direct and maternal component for perinatal mortality was positive in primiparous dams (0.39 s.e. 1.208), negative in multiparous dams (-0.86 s.e. 0.311) and negative across all parities (-

0.27 s.e. 0.4907. The correlation between the direct and maternal component for PMND was negative in primiparous (-0.56 s.e 0.485) dams and was not estimable in multiparous dams and across all parities.

2.6 Discussion

Genetic parameters for calving performance traits currently included in genetic evaluations were estimated from a limited dataset in the year 2006. Recording of calving performance data is now routine in Irish dairy and beef herds. The objective therefore of this study was to re-estimate genetic parameters for dystocia and perinatal mortality. Moreover, the previous genetic parameters were estimated using data from all parities. In the present study we report lower heritability estimates for all traits but estimate genetic parameters separately in primiparous and multiparous dams as well as the genetic direct and maternal correlations.

2.6.1 Dystocia

The prevalence of dystocia in the present study (primiparous 4.29%, multiparous 2.61%) was lower than previously documented by Mee et al. (2011) (primiparous 9.3% multiparous 5.8%) although Mee et al. (2011) solely focused on the Holstein-Friesian population where only matings to AI sires were retained. The prevalence of dystocia in the Holstein-Friesian population in the present study (primiparous 4.95%, multiparous 2.51%) was also lower than reported by Mee et al., (2011). A decreasing trend in dystocia prevalence in recent years, has been reported by Mee et al. (2011) and is possibly due to changes in breeding schemes, animal genotype and nutritional management.

Dystocia prevalence was greater in primiparous dams in the present study corroborating the international literature (Berry et al., 2007b, Eriksson et al., 2004, Steinbock et al., 2003). Primiparous dams may have an greater risk of dystocia due to fetopelvic incompatibility (Phillipsson et al., 1976). Calf gender particularly influenced dystocia prevalence in primiparous dams in the present study, with male calves almost twice as likely to experience severe dystocia (5.7%) than females (2.9%). Morphological differences that exist between genders may be one cause; male calves are more likely to have a broader skeleton but also tend to be heavier at birth than their female counterparts (Eriksson et al., 2004, Phillipsson, 1976). Longer gestation lengths, which are associated with an increased risk of dystocia, have also been reported in males could also be a contributing factor; Casas et al (2012) reported a one-day longer gestation length in male calves compared to female calves.

The heritability of dystocia in the present study, although low was similar to estimates in other cattle populations (Eaglen et al., 2012, Eriksson et al., 2004, Hickey et al., 2007, Steinbock et al., 2003). Steinbock et al. (2003) reported a heritability of 0.17 for the direct effect of dystocia in primiparous dams, although the maternal heritability in the Steinbock study (0.12) was four times the maternal heritability estimated in the present study. The negative correlation between the direct and maternal components for dystocia in both primiparous and multiparous dams is corroborated by studies for other cattle populations (Eaglen et al., 2012, Eriksson et al., 2004, Steinbock et al., 2003). This correlation between the direct and maternal components suggests that animals with a lower genetic risk of being born without calving difficulty are genetically prone themselves to have a more difficulty calving. Female calves born without dystocia are likely to be small framed calves, hence once these calves become cows the small frame (e.g pelvic width) may make them more prone to calving difficulty. A stronger-negative direct-maternal correlation was reported in multiparous compared to primiparous dams in the present

study. Similar values of -0.43 in primiparous and -0.55 in multiparous dams were reported by Carnier et al., (2000), with a similar trend also reported by Eriksson et al., (2004). This is contrary to previous reports by Eaglen et al., (2012) and Steinbock et al., (2003) who reported a stronger-negative direct-maternal correlation in primiparous dams. Despite the contrasting results, all studies confirm the antagonistic relationship between direct and maternal calving difficulty, suggesting easily born female calves are more likely to exhibit greater difficulties when giving birth as dams (Carnier et al., 2000).

2.6.2 Perinatal mortality

The prevalence of perinatal mortality reported in the present study (primiparous 4.15%, multiparous 1.79%) is similar to prevalence statistics reported internationally in Norweigen Red (Heringstad et al., 2007) and Swedish Holsteins (Steinbock et al., 2003). Although, a previous Irish study (Mee et al., 2008) involving the Holstein-Friesian population, reported a greater prevalence of perinatal mortality (primiparous 7.7%, multiparous 3.5%) than found here. A greater prevalence of perinatal mortality has been documented in Holstein-Friesians in comparison to other breeds (Steinbock et al., 2003). An increase in perinatal mortality with increasing proportion of Holstein-Friesian genes in both the calf and the dam in comparison to the original Danish black and white population was also documented by Hansen et al. (2004). Inbreeding may be a causative factor increasing the risk of perinatal mortality within this breed possibly due to the accumulation of harmful deleterious effects; McParland et al (2007a) documented a 1% greater incidence of perinatal mortality in a 12.5% inbred Holstein-Friesian animal. Therefore, the moderate level of relatedness reported in the Holstein-Friesian population (Hammami et al., 2007, McParland et al., 2007) may explain the high level of mortality in this breed.

The prevalence of perinatal mortality has been reported to be greater in primiparous dams in several different cattle populations (Berry et al., 2007, Eriksson et al., 2004, Johanson and Berger, 2003, Mee et al., 2008, Steinbock et al. 2003) and a similar phenomenon was observed in the present study. Dystocia has been documented to be one of the main contributory factors to perinatal mortality (Philpsson et al., 1976) and the greater prevalence of dystocia in primiparous dams in the present study and elsewhere may explain the high level of mortality. The results from the present study and from other studies suggest that even slight assistance at calving is associated with an increased risk of perinatal mortality (McGuirk et al., 1999, Mee et al., 2008, Meyer et al., 2001). Calving assistance has been shown to have detrimental effects on adaptation of calves to extrauterine life in comparison to calves delivered without assistance, with anoxic and traumatic lesions commonly reported in perinatal mortality events that required assistance during delivery (Hosseini-Zadeh, 2014, Mee, 2013).

Heritability estimates of perinatal mortality were extremely low in the present study; which is not uncommon as international heritability estimates for perinatal mortality range from 0.001 (Eriksson et al., 2004; Swedish Hereford primiparous dams) to 0.12 (Steinbock et al., 2003; primiparous Swedish Holsteins). The positive direct-maternal correlation in primiparous dams for perinatal mortality was similar to that reported by Eaglen et al., (2012) across all parities (0.37, s.e 0.57), although a high standard error was reported in the present study. Currently in Ireland, no maternal EBV for perinatal mortality are generated in the national genetic evaluations. Sufficient additive genetic standard deviation for selection was however detected in the present study for the maternal estimate in primiparous dams and across all parities for perinatal mortality (Table 2.3) suggesting that this component should be included in the national genetic evaluations. The inclusion of a maternal heritability in the genetic evaluation system may permit producers to identify which dam's will be more likely to produce viable calves. The maternal estimate of

perinatal mortality could also be a reflection of the dams ability to produce quality colostrum of a high immunoglobulin G (IgG) concentration within the first 24hrs of parturition. Transfer of sufficient IgG to the neonatal calf via colostrum is vital to provide the calf with immunological protection and resistance against disease and a recent study reported a heritability of 0.10 for IgG concentration in Irish dairy cows with a genetic standard deviation of 16.0 g/l (Conneely et al., 2013).

2.6.3 Perinatal mortality not associated with dystocia

Recent studies have suggested that the prevalence of perinatal mortality which was not associated with calving difficulty is increasing (Gustafsson et al., 2007, Muskens, 2008). The prevalence of PMND was low in the present study (primiparous 1.57%, multiparous 0.96%) but nonetheless is a common occurrence on farms. This suggests that other factors such as congenital defects, placental dysfunction, infections, precalving nutrition or possibly management decisions such as age at first calving may be the cause of these mortality events. Smaller and less efficient placentas have been reported in primiparous dams (Echternkamp, 1993) with a similar trend documented in sheep (Dwyer et al., 2005) which may contribute to the higher prevalence of PMND reported in primiparous dams in the present study. Gestation length has also been shown to be shorter in primiparous dams and shorter gestation lengths are associated with an increased risk of mortality (Johanson and Berger, 2003). Calves prematurely born are often small and weak (Niskanen. and Juga., 1998, Philipsson, 1976) and a gestation length less than 273 days is increasingly associated with perinatal mortality as organs such as lungs are not yet fully developed (McClintock et al., 2005).

Perinatal mortality not associated with dystocia was only lowly heritable for the direct component in primiparous dams (0.019). The lack of genetic variation in later

parities may be because the genetic variation has been purged out through selection, especially in beef herds since a primiparous cow with no calf may be slaughtered immediately. Further research is needed to fully comprehend the genetic variation of this trait, although currently the available genetic variation could be exploited through the use of breeding programs to help reduce the prevalence.

2.7 Conclusion

The prevalence of calving difficulty and perinatal mortality was lower than previously reported in Irish populations, although the heritability estimates of these traits were similar to previously reported international estimates. The existence of PMND suggests that a greater understanding of the possible causes and risk factors of this trait should be examined before it becomes widely prevalent. All three calving performance traits exhibited reasonable genetic variation suggesting that exploitation through breeding programs to lower the prevalence is possible. The inclusion of the maternal component of perinatal mortality in the national genetic evaluations should be considered to permit for further genetic improvement in this trait.

Chapter 3

Runs of Homozygosity and population history in cattle

3.1 Preface

At the time of thesis submission this chapter was published in BMC Genetics (Accepted August 3, 2012; doi: 10.1186/1471-2156-13-70). The full reference is:

Purfield DC, Berry DP, McParland S, Bradley DG: Runs of homozygosity and population history in cattle. *BMC Genetics* 2012, 13:70.

Deirdre Purfield was primary author and performed the SNP edits, ROH analysis, and statistics and drafted the manuscript. Donagh Berry and Daniel Bradley conceived the study, participated in the design and co-ordination of this study and contributed to the manuscript. Sinead McParland calculated the pedigree inbreeding coefficients.

Formatting and referencing style has been edited for consistency throughout the thesis.

Figure and table captions have been assigned with a chapter prefix. Competing interests and acknowledgements have been removed. All other aspects are consistent with the published manuscript.

3.2 Abstract

Runs of homozygosity (ROH) are contiguous lengths of homozygous genotypes that are present in an individual due to parents transmitting identical haplotypes to their offspring. The extent and frequency of ROHs may inform on the ancestry of an individual and its population. Here we use high density ($n=777,962$) bi-allelic SNPs in a range of cattle breed samples to correlate ROH with the pedigree-based inbreeding coefficients and to validate subsequent analyses using 54,001 SNP genotypes. This study provides a first testing of the inference drawn from ROH through comparison with estimates of inbreeding from calculations based on the detailed pedigree data available for several breeds.

All animals genotyped on the HD panel displayed at least one ROH that was between 1-5 Mb in length with certain regions of the genome more likely to be involved in a ROH than others. A strong correlation ($r=0.75$, $p<0.0001$) existed between the pedigree-based inbreeding coefficient and a statistic based on sum of ROH of length $> 0.5\text{Mb}$ and suggests that in the absence of an animal's pedigree data, the extent of a genome under ROH may be used to infer aspects of recent population history even from relatively few samples. Our findings suggest that ROH are frequent across all breeds but differing patterns of ROH length and extent illustrate variations in breed origins and recent management.

Keywords; Runs of homozygosity, inbreeding, cattle population history

3.3 Introduction

Runs of homozygosity (ROH) are contiguous lengths of homozygous genotypes that are present in an animal due to parents transmitting identical haplotypes to their offspring. The extent and frequency of these may inform on the ancestry of an individual and its population. Particularly, consanguinity may be indicated from the presence of long ROH; the longer such segments are, the more likely that recent inbreeding occurred within a pedigree (Kirin et al., 2010). However, unusually long runs of homozygosity may also persist in outbred individuals, perhaps due to unusual mutation, linkage disequilibrium (LD), and recombination rates at certain genomic locations (Gibson et al., 2006).

The distribution of shorter ROH may also inform on the presence of more ancient relatedness which is unaccounted for in an individual's recorded pedigree due to the limitations in the recording process (McQuillan et al., 2008). In the context of domestic animals these may result from breed or population founder effects or other restrictions.

The domestication of cattle, which occurred ~10,000 years ago, was a complex process, with evidence suggesting that it occurred in a minimum of two domestication events (Loftus et al., 1994). However, both natural and artificial selection of cattle, as well as regional variations due to drift has resulted in breeds that differ extensively in phenotypes. These processes and the extent of breeding control have differed greatly among populations and ROH may provide useful information on these disparate histories. Particularly, in recent times the practices of intense selection of sires, artificial insemination, and embryo transfer have featured heavily in some breeds, reducing effective population sizes, genetic diversity and affecting levels of homozygosity.

Runs of homozygosity have been extensively studied in human populations and are an established method of distinguishing a population history of consanguinity, and with

homozygosity mapping analysis showing a relationship with susceptibility to recessive diseases (Gibson et al., 2006, Hildebrandt et al., 2009, Kirin et al., 2010, Lencz et al., 2007, McQuillan et al., 2008). Here we use high density (n=777,962) bi-allelic SNP data in a range of cattle breeds to correlate ROH with the pedigree-based inbreeding coefficient and to validate further analysis using 54,001 SNP genotypes. This allows examination and interpretation of the level of ROH that exist in a wide range of cattle breeds samples.

3.4 Methods

3.4.1 Genotypes and Quality Control

Single nucleotide polymorphisms (SNPs) genotypes consisting of 777,972 biallelic SNPs from the Bovine HD BeadChip (Illumina Inc., San Diego, CA) were generated for 891 artificial insemination sires of multiple breeds. Breeds represented included Angus (n=39), Belgian Blue (n=38), Charolais (n=117), Friesian (n=98), Hereford (n=40), Holstein (n=262), Holstein-Friesian crosses (n=111), Limousin (n=128) and Simmental (n=58). Additionally, the 48,734 SNPs common to both the HD and the Illumina BovineSNP50 Beadchip were retained in a reduced HD panel which was used to evaluate the lesser *circa* 50k SNP density for identifying ROH; here on in referred to as the reduced HD panel. In addition, three published genotype datasets at this density comprising of a total of 1166 animals from 42 different breeds (detailed in appendix A. 1) generated from the Illumina Bovine50 Beadchip, hereon in known as the SNP50 panel, were also available (Gautier et al., 2009, Gautier et al., 2010, Matukumalli et al., 2009).

The impact of the following SNP edits on the number of SNPs remaining for each of the three density panels is summarised in Table 3.1. Only biallelic SNPs on the 29 autosomes were retained and both animals and SNPs with call rates <90% were discarded.

Also, monomorphic SNPs and SNPs that deviated from the Hardy Weinberg equilibrium ($p < 0.0001$) within breed were discarded.

Table 3.1 Number of single nucleotide polymorphisms for each data edit implemented on each density panel. The SNP edits that were applied to each dataset detailing the amount of SNPs removed and the number of animals in each dataset is listed in parenthesis.

Genotype edit	HD Panel	Reduced	Hapmap	Gautier
		HD Panel	SNP50	SNP50
Initial data set	777,962	48,734	54008	58336
	(891)	(891)	(533)	(633)
Autosomes only	735,293	47,270	51587	55200
	(891)	(891)	(533)	(633)
SNP and animals with >90% Call Rate	727,559	46,761	50824	44687
	(867)	(867)	(530)	(633)
Hardy Weinburg Equilibrium in each breed ($p > 0.0001$)	723,502	46,379	50783	44602
	(867)	(867)	(530)	(633)
Monomorphic SNPs removed	665,058	43,028	48128	44160
	(867)	(867)	(530)	(633)

3.4.2 Definition of a Run of Homozygosity

Runs of homozygosity (ROH) were defined in the populations of animals for each of the three SNP density panels using PLINK v1.07, which slides a window of 50 SNPs, in one SNP intervals, across the genome estimating homozygosity. No more than two SNPs with missing genotypes were allowed per window and up to one possible heterozygous genotype was permitted. To ensure low SNP density did not affect ROH length, the minimum required density of SNPs per kb differed between the two genotyping densities,

as well as the maximum distance length between two consecutive homozygous SNPs in a run. To minimize the number of ROH that occur by chance in the HD panel, the minimum number of SNPs that constituted a ROH (l) was calculated by a method similar to that proposed by Lencz et al., (2007).

$$l = \frac{\log_e \frac{\alpha}{n_s \cdot n_i}}{\log_e (1 - \overline{het})}$$

where n_s is the number of SNPs per individual, n_i is the number of individuals, α is the percentage of false positive ROH (set to 0.05 in the present study), $\overline{het}$ is the mean SNP heterozygosity across all SNPs. For the exclusion of very short and common ROH that occur prevalently throughout the genome due to LD, a minimum ROH length of 500 kb was set.

For analysis of the HD panel genotypes, the minimum SNP density was 1 SNP every 50kb to ensure low SNP density did not falsify ROH length, a minimum run length of 58 SNPs was needed to produce <5% randomly generated ROH and the maximum gap between two consecutive homozygous SNPs in a run was set at a 100kb. In the analysis of the reduced HD panel and the Bovine SNP50 panel, the minimum SNP density was altered to 1 SNP every 120kb and no restriction was placed on the minimum number of SNPs in a ROH and the maximum gap length between two consecutive homozygous SNPs in a run was kept at the default value of 1000kb to account for the lesser genotype density.

Runs of homozygosity were identified for each animal separately on the HD and reduced panel and for comparative purposes only, the maximum gap length in the HD panel for ROH identification was altered to 1000kb, SNP density to 1 SNP every 120kb and there was no restriction on the minimum number of SNPs that constituted a ROH, in order to limit bias. To establish that the reduced density also predicts the correct ROH

length category for ROH, the extent that reduced panel ROH were correctly assigned to the HD ROH length category was plotted.

Animals with overlapping ROH, and those ROH that were an allelic match, were also identified in the HD panel. The identification of overlapping regions was done by using the sliding window approach as mentioned above, and then for each SNP by calculating the proportion of homozygous windows in the population dataset that overlap that same position. The percentage of animals that had the region with the most overlapping ROH on each chromosome was plotted and the percentage of these overlapping ROH that were an allelic match $\geq 95\%$ were identified.

The percentage population of each breed with ROH present at different ROH length categories was calculated, as well as the mean overall ROH sum per animal for each breed. The mean sum of ROH within each ROH length category was also calculated by summing all ROH per animal in each ROH length category and averaging this per breed population. The percentage of SNP involvement in ROH was also calculated by counting the amount of times a SNP appeared in a ROH in the population dataset.

3.4.3 Inbreeding Coefficient vs. Runs of Homozygosity

The measure of homozygosity per animal was calculated using the HD panel and reduced HD panel, by a method similar to that as proposed by McQuillan et al., (2008) except in the present study the centromeric region was also included in the calculation,

$$F_{ROH} = \sum \frac{L_{ROH}}{L_{AUTO}}$$

in which L_{ROH} is the sum of ROH per animal above a certain criterion length and L_{AUTO} is the total length of autosome covered by SNPs. L_{ROH} was calculated separately as the sum

of all ROH >500kb and the sum of all ROH >10,000kb where SNP autosomal genome coverage was 2,510,611 kb in the HD panel and 2,500,265 kb in the reduced HD panel. Pedigree based inbreeding coefficients for all animals were calculated using the Meuwissen and Luo (1992) algorithm. Depth of pedigree known was measured in complete generation equivalents (CGE) for all animals as described in (Mc Parland et al., 2007b) and correlations between all measures of inbreeding were calculated only on animals (n=230) with a CGE value ≥ 6 .

3.5 Results

3.5.1 ROH in animals with HD panel genotypes

We used a definition of ROH as tracts of homozygous genotypes that were >500kb in length identified in a genome sliding window of 50 SNPs. No more than two missing genotypes and one possible heterozygous genotype were allowed in a window, and within our HD SNP panel data, all 867 animals tested, displayed at least one ROH that was between 1-5 Mb in length with almost all (i.e., 98%) of the population dataset also having at least one ROH between 5-10Mb in length. Differences among breeds existed in their frequency in different ROH length categories (Figure 3.1). The two British breeds, the Angus and the Hereford, had a larger mean portion of their genome, 198.6 Mb and 198.7 Mb, respectively, covered in shorter length (i.e. 1-5 Mb) ROH; coverage ranged from 80.58 to 93.48 Mb in the remaining breeds. For all breeds, most ROH segment coverage was in the shorter length categories with the Holstein, Holstein-Friesian and Friesian samples showing the greatest frequency in the longer ROH categories which are more indicative of recent inbreeding. A smaller effective population size and its associated unfavourable effects due to inbreeding depression in these dairy cattle is a current concern in the industry (Mc Parland et al., 2007a). The mean sum of the ROH per animal for all

ROH length categories > 15-20Mb was less than 20Mb and no animal had more than five ROH that were >30 Mb in length. The three most homozygous animals present in our dataset were from the dairy breeds and had on average 700.3 Mb classified as ROH; this is equivalent to almost a quarter of their genome.

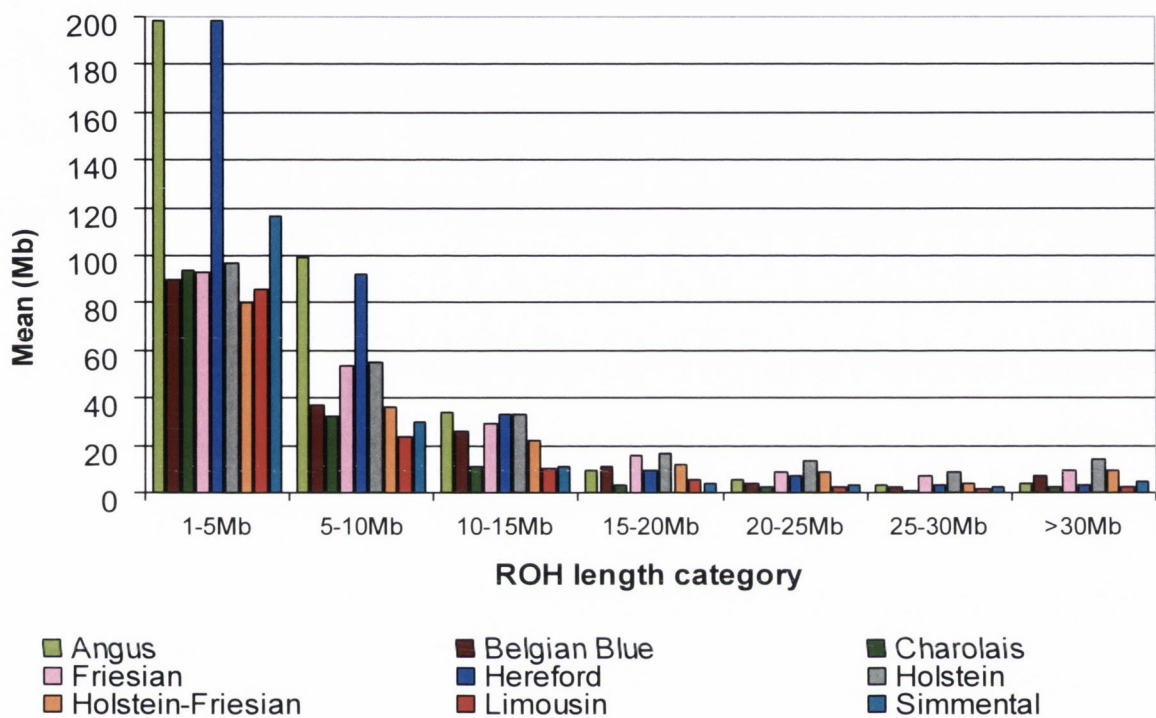


Figure 3.1 The mean sum of runs of homozygosity (ROH) per HD genotyped animal measured in megabases (Mb) within each ROH length category. Breeds from left to right are Angus, Belgian Blue, Charolais, Friesian, Hereford, Holstein, Holstein-Friesian, Limousin and Simmental. Angus and Hereford had high amount of mean short ROH (<5 Mb) possibly due to ancestral relatedness whereas the Holsteins had the greater mean long ROH (>20 Mb) due to recent consanguineous matings.

3.5.2 Genome locations of ROH

The number of ROH per chromosome was greatest for chromosome 1 (11,513 runs across all 867 animals with the HD panel) with on average, 12.39% of the chromosome consisting of a ROH. The number of ROH per chromosome tended to decrease with chromosome

length (see Appendix A.2). The fraction of chromosome residing in ROH was greatest on chromosomes 14 and 16, with 13.71% and 14.16%, respectively; these had the largest extent of overlapping ROH positions among individuals (see Appendix A.3). In particular, 756 animals (87.2% of the dataset) had a ROH segment on chromosome 14 involving the same genomic region centred around the 25Mb position that consisted of 28 SNPs and was 127.3kb in length (Figure 3.2; Appendix A.4). When interrogated using Haploview (Barrett et al., 2005) the majority of SNPs in this region were in high LD with each other (Appendix A.5). Also, regions on each of chromosomes 7, 8, 14, 16, 18 featured in a ROH in over 55% of the sampled animals (Figure 3.2). Chromosome 12 displayed the longest run of contiguous SNPs (197) that were uninvolved in any ROH within the sample population and interrogation in Haploview (Barrett et al., 2005) revealed that this genomic region had corresponding low levels of LD between SNPs (Appendix A.6).

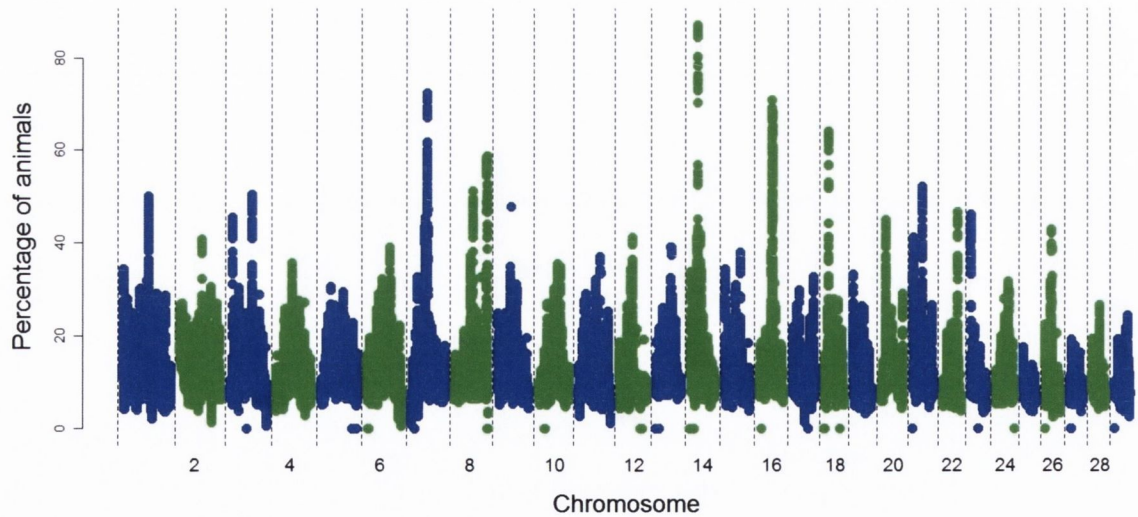
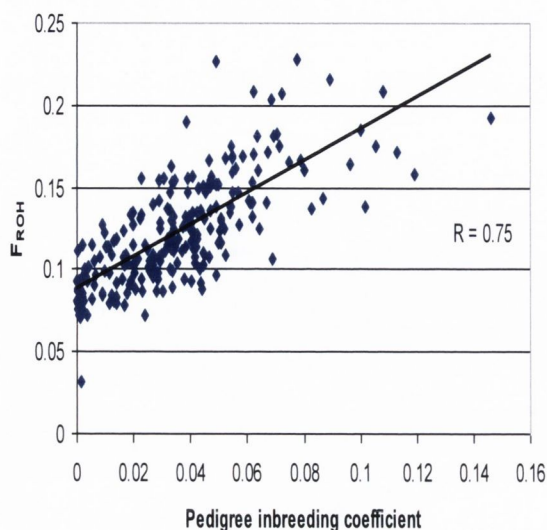


Figure 3.2 Incidence of each single nucleotide polymorphism (SNP) in a run of homozygosity (ROH) in the HD panel. Particular genomic positions can be seen, on each of chromosomes 7, 8, 14, 16, 18 featured in ROH which were shared in over 55% of the sampled animals

3.5.3 Correlation between ROH and inbreeding coefficient

Cattle offer an opportunity to assess ROH as an indicator of inbreeding by comparing pedigree-based estimates of inbreeding using the same individuals. Figure 3.3 illustrates a scatter plot of either individual animal F_{ROH} or $F_{ROH10000}$ (the sums of all ROH per animal that are >0.5Mb and >10Mb respectively) against the inbreeding coefficient based on pedigree recorded (McParland et al., 2007). Conservatively, we restricted our sample to the 230 animals that had complete generation equivalent values ≥ 6 of recorded pedigree ancestry. A clear linear relationship exists between the ROH and pedigree based estimates of inbreeding. The correlations between the pedigree inbreeding coefficient and individual sums of ROH were different from zero with values of $r=0.75$ ($P<0.0001$) for F_{ROH} and $r=0.71$ ($P<0.0001$) for $F_{ROH10000}$ using the HD panel and $r=0.73$ ($P<0.0001$) for F_{ROH} and $r=0.70$ ($P<0.0001$) for $F_{ROH10000}$ using the reduced HD panel. The intercept (0.088, s.e=0.002) of the regression of F_{ROH} on the pedigree inbreeding coefficient was greater than zero suggesting that the pedigree-based inbreeding coefficient may underestimate the levels of ancestral genomic relatedness that may exist. The intercept (0.006; s.e=0.002) of the regression of $F_{ROH10000}$ on the pedigree inbreeding coefficient was also greater than zero, but to a lesser extent when the 10 Mb ROH length limit is applied; consistent with longer ROH resulting from more recent inbreeding, such as that recorded in these pedigrees.

A)



B)

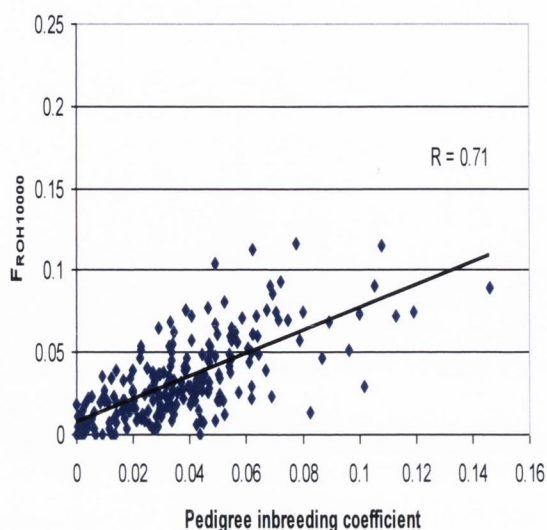


Figure 3.3 Scatter plot of A) FROH on pedigree inbreeding coefficient B) FROH10000 on pedigree inbreeding coefficient^a

^a F_{ROH} is the sum of all runs of homozygosity (ROH) per animal that are $>0.5\text{Mb}$, $F_{ROH10000}$ is the sum of all ROH per animal that are $>10\text{ Mb}$ and the pedigree inbreeding coefficient is based on pedigree recorded data calculated by the Meuwissen and Lou algorithm (1992).

3.5.4 Validation of Bovine SNP50 density

The most common SNP density platform for cattle is that from 54,000 SNP genotyping arrays (Illumina BovineSNP50beadchip); an important question is whether this panel can reliably infer ROH that have been detected with the greater accuracy provided by the HD array. To facilitate this, a reduced genotype dataset comprising 48,734 SNPs that were common to both the HD and Bovine50 Beadchip genotype panel on the 867 genotyped sires was used for comparison purposes. The same ROH criteria were applied to each dataset, where no more than two missing genotypes and one possible heterozygote were permitted per window, no restriction was placed on the number of SNPs in a ROH in either

density and the maximum gap between two consecutive SNPs was kept at the default value of 1000kb to account for the lesser genotype density in the reduced HD panel and to limit bias for comparison of ROH. There were in total 157,600 ROH identified using the HD panel, whereas only 19,078 ROH were identified in the same individuals using the reduced HD panel. However, the vast majority of ROH that the reduced panel failed to recognise were between 0.5-1 Mb in length (Figure 3.4). Also, only 27.7% of the 1-5 Mb ROH length category in the HD panel were identified by the reduced panel but almost all ROH in length categories of 5Mb and above were identified (Figure 3.4). In the larger ROH length categories, ROH from the HD panel that were recognised by the reduced panel were correctly assigned >70% of the time to the correct length category where the majority of incorrectly assigned ROH were due to inflation of length in the reduced HD panel. When the reduced panel density ROH estimates were compared to pedigree inbreeding values using the same individuals the correlations obtained were very similar (0.73; $P < 0.0001$) using F_{ROH} and 0.70 ($P < 0.0001$) using $F_{ROH10000}$ to those observed for the HD genotype panel.

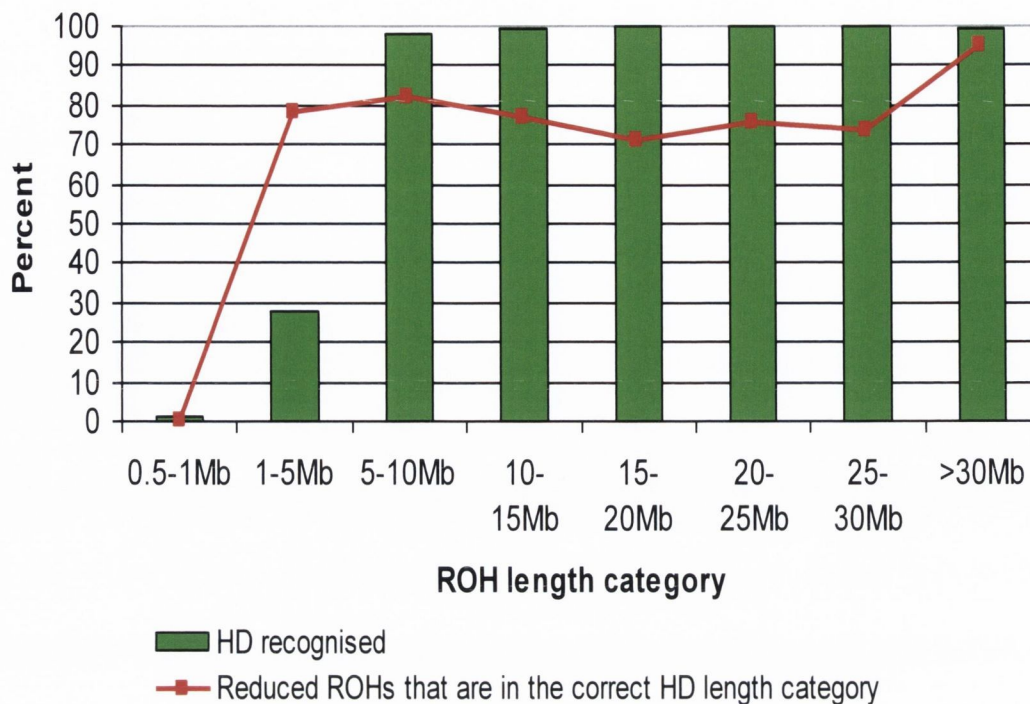


Figure 3.4 The percentage of runs of homozygosity (ROH) in the HD panel that are recognised by the reduced density panel and the percentage of the reduced panel ROH that are correctly identified into the correct ROH length category. Greater than 70% of the ROH identified using the reduced panel were placed in the correct length category. The reduced density failed to identify ROHs between 0.5-1 Mb in length due to the low SNP density.

3.5.5 ROH in Bovine SNP50 genotypes across the 42 breeds

The success of SNP50 validation exercise allows the testing of a large previously published data collection of wide provenance (Gautier et al., 2009, Gautier et al., 2010, Matukumalli et al., 2009) for ROH distributions using 1Mb as a ROH minimum. Figure 3.5 describes the total (Mb) per individual genome estimated to be under ROH in different breed groupings. There were clear differences among breeds in both the levels and variation of ROH frequency. Also, there were discernible patterns with respect to breed origin. The British Isles breed samples (Angus, Guernsey, Hereford and Jersey) clearly

displayed the highest individual sum of ROH per animal. Mainland European breed samples showed varying ranges of ROH levels and zebu breeds had intermediate levels. African breeds had a tendency towards low ROH but the West African taurine breeds (Baoule, Lagune, N'Dama and Somba) showed high variability in individual sums of ROH within breed samples (Figure 3.5).

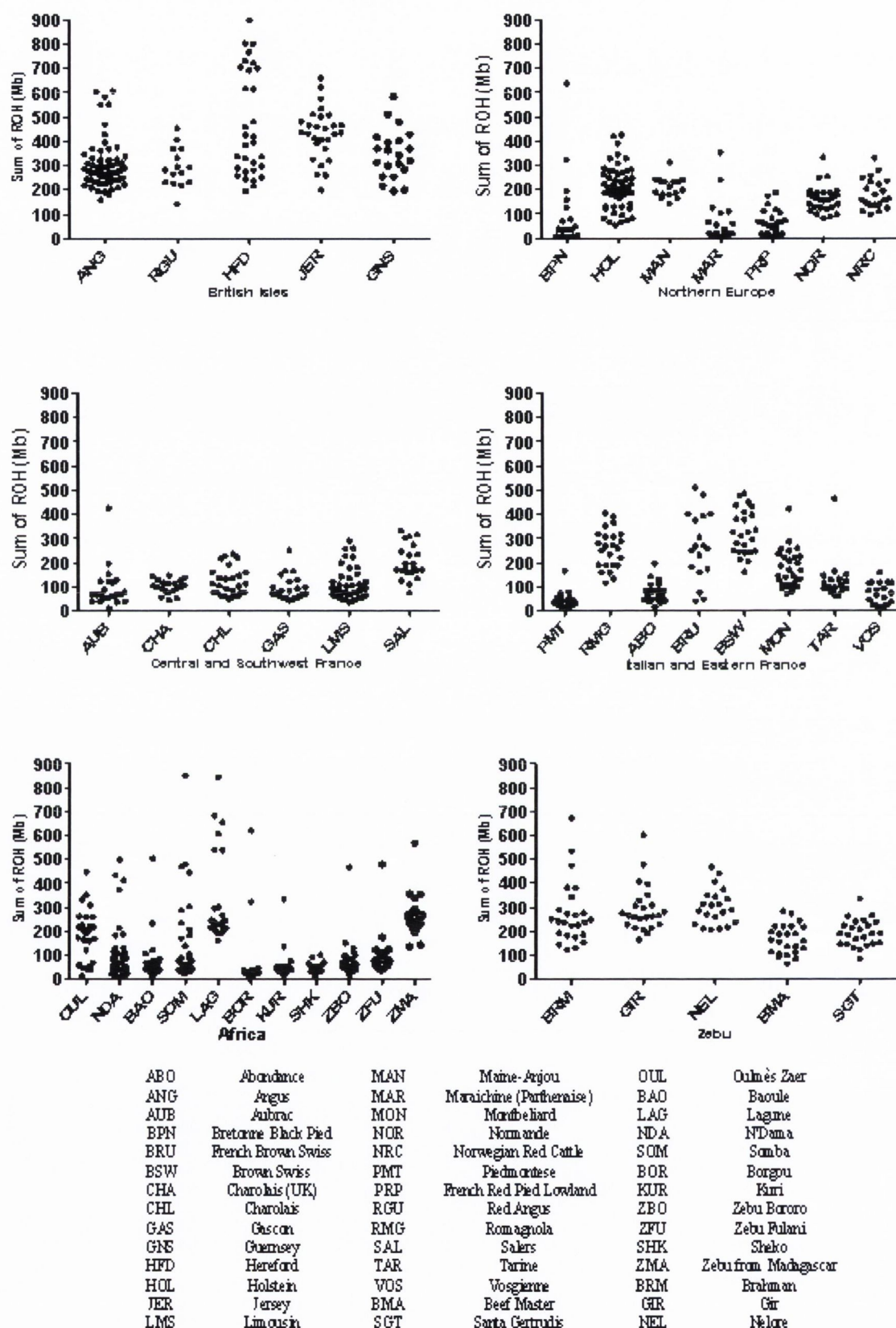


Figure 3.5 Individual sum of total runs of homozygosity (ROH) per bovine SNP50 animal split into origin of breed. The British Isles breeds had higher sums of ROH per animal in comparison to other geographic origins. Breeds originating in Africa have the most disparate range of sums a reflection of the current breeding practices.

Long ROH arise as a result of recent inbreeding, while shorter ROH can indicate more distant ancestral effects such as breed founder effects. In order to discriminate between these effects the average sum of the genome under ROH of size >20Mb against the average sum <20Mb for each breed were compared (Figure 3.6). A correlation of $r=0.68$ ($P<0.0001$) existed between the two measures but the relative contributions of long and short ROH differed sharply among breeds and breed types. The African *Bos taurus* breeds showed low levels of ROH for both measures, except for the Somba, Oulmès Zaer and Lagune breeds where elevated levels were clearly highly influenced by long ROH and hence recent consanguinity. The two Channel Islands breeds also showed this trend. In contrast, the zebu breeds showed intermediate values of ROH but these were more strongly influenced by shorter ROH, perhaps indicating a founder effect rather than recent inbreeding.

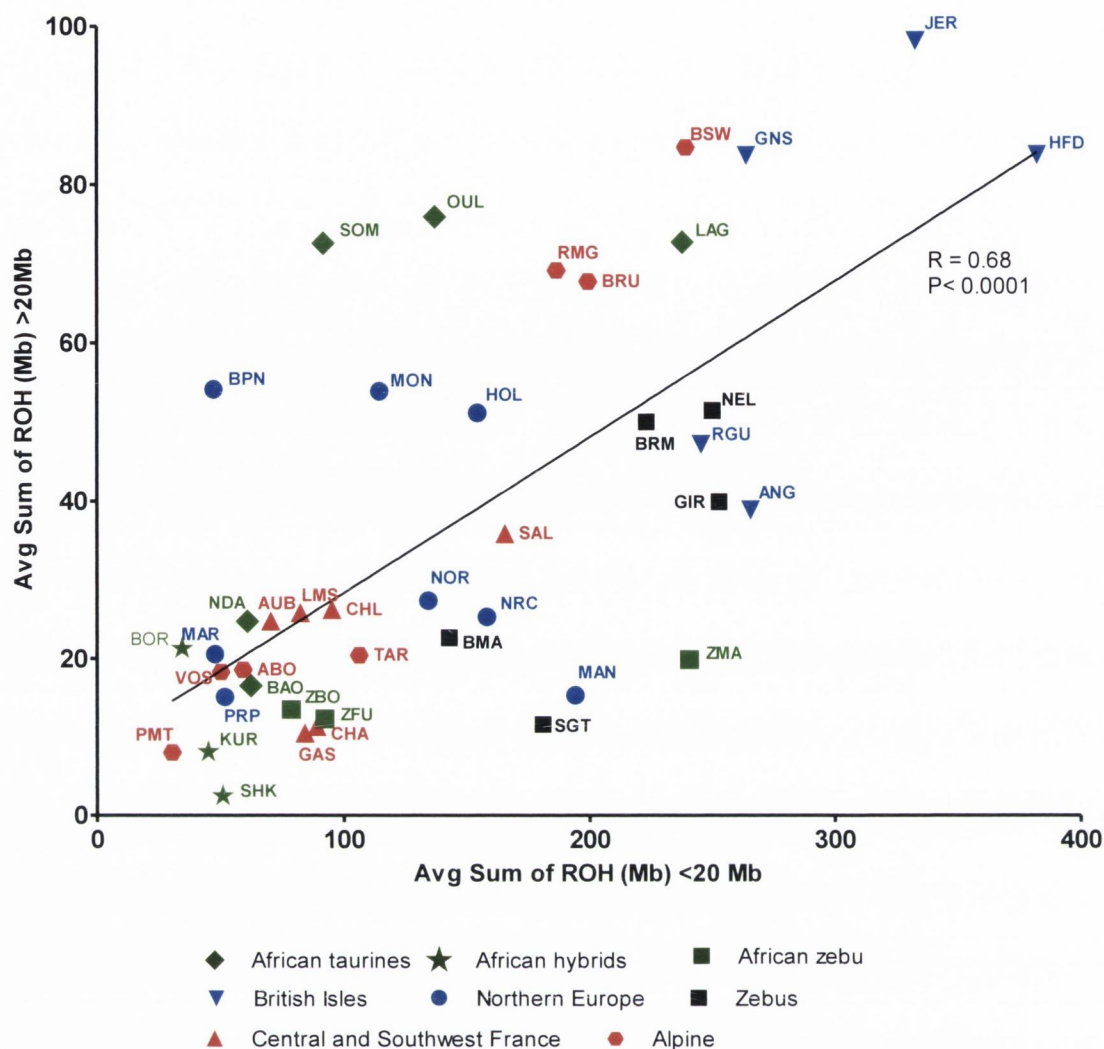


Figure 3.6 Average sum of runs of homozygosity (ROH) per bovine SNP50 animal <20Mb in length within breed vs average sum of ROH>20Mb in length. Breeds that originate from the same geographical area tend to cluster together with African *Bos taurus* breeds showing low levels for both measures, except for the Somba, Oulmès Zaer and Lagune breeds where elevated levels were clearly highly influenced by long ROH whereas the zebu breeds showed intermediate values of ROH but these were more strongly influenced by shorter ROH.

3.6 Discussion

Our findings show that ROH are frequent across all breeds and that certain ROH length categories can be used as an indication of consanguinity. They can also inform on breed population history as the effects of population bottlenecks, selection pressure and breeding management on the bovine genome may potentially leave an imprint on ROH length.

The bovine HD SNP assay allows an analysis of ROH at similar density to that employed to generate genomic signatures of endogamy that differ markedly among human populations (Gibson et al., 2006, Kirin et al., 2010, McQuillan et al., 2008). Moreover, cattle allow a first comprehensive testing of the inference drawn from ROH through comparison with estimates of inbreeding from calculations based on the detailed pedigree data available for many breeds. The strong correlation between the pedigree inbreeding coefficient and sum of ROH of length $> 0.5\text{Mb}$ suggests that, in the absence of an animal's pedigree data, the extent of a genome under ROH may be used to infer aspects of recent population history even from relatively few samples, as previously suggested by McQuillan et al (2008). However, 44% of the variance in ROH distribution remains unexplained by pedigree inbreeding and may partly reflect the limitations of ancestry recording in cattle where founder animals are generally, and often inaccurately, assumed to be unrelated (McParland et al., 2007b). Additionally, the propensity for multiple megabase scale ancestral haplotypes in certain genome regions to persist even in outbred animals, perhaps due to localised low levels of recombination and high levels of LD may contribute (McKay et al., 2007). Lastly, we note that pedigree relatedness gives an expected, not actual, proportion of genomic identity by descent among individuals and it might be anticipated that genotype-based estimates provide greater accuracy on relatedness (Visscher et al., 2006).

Whereas HD SNP panels facilitate more accurate detection of ROH, the vast majority of cattle SNP genotype data, and emerging data in other livestock, is available at ~50,000 SNP density. It is therefore of interest as to whether these sparser genotypic data can reliably inform on ROH and inbreeding history. We found that HD ROH were not accurately identified in such a reduced panel if between 0.5-1Mb in length. However, the SNP50 density genotypes were sufficient to recognise almost all ROH >5 Mb but they also has the potential to inflate ROH length. Importantly, ROH levels at the lower SNP density correlate equally well with the pedigree estimates of inbreeding. We conclude that this prevalent marker density is appropriate in identifying ROH.

We used three published (Gautier et al., 2009, Gautier et al., 2010, Matukumalli et al., 2009) SNP50 genotype collections to examine patterns in ROH distribution and compare aspects of population history among a range of cattle breeds. The domestication process itself featured a limited sampling from the wild with a more recent bottleneck detectable 50-100 generations ago, presumably corresponding to breed formation (Gautier et al., 2007). However, this traditional breed formation is largely a European phenomenon and its absence is most apparent in the data from African cattle. These samples, including *B.taurus* breeds, humped *B.Indicus* breeds and indicine/taurine hybrids, tended toward low levels of ROH per genome, reflecting traditional management practices in Africa, characterised by less controlled mating (Ibeagha-Awemu and Erhardt, 2005).

An open village breeding system may also predispose to random consanguineous matings and many African breeds show outlying highly inbred individuals (Figure 3.5). The length distribution of ROH can help to distinguish different types of parental relatedness. Samples from human populations where cousin marriage is common show an excess of long ROH; whereas for example, Papuan and Melanesian human populations show an excess of shorter ROH, consistent with effects of reduced population and isolation

rather than first degree relative unions (Kirin et al., 2010). Figure 3.6 compares the contributions of long (>20Mb) and short ROH to breed homozygosity in order to differentiate the effects of ancient and more recent relatedness among ancestors. Here, the three African taurine breeds (Oulmes Zaer, Somba and Lagune) with higher homozygosity clearly show a strong influence of ROH of length greater than 20 Mb and hence of recent inbreeding. We note that Gautier et al (2009) reported a high F_{IS} value, as well as extensive linkage disequilibrium within the Lagune breed.

Zebu-taurus hybridisation is also a dynamic and contemporary process within Africa (Hanotte et al., 2000, Kolbe et al., 2008, Lavergne and Molofsky, 2007). This acts to increase genetic diversity and contributes to the interruption of stretches of homozygous genotypes within individuals. The effects of this process are evident in three hybrid breeds which show the lowest extent of ROH in African breeds. These include the Kuri; where previous work has shown a near 50:50 genetic admixture between surrounding zebu and themselves (Meghen et al., 2000), and the Sheko breed, where the original taurine African *Y* chromosome is in danger of disappearing from this breed due to the use of zebu bulls (Hanotte et al., 2000).

Within European breeds, British breeds tended toward higher quantities of ROH, reflecting results of previous microsatellite research, where British Isles breeds had lower levels of observed heterozygosities and gene diversities in comparison to other Mediterranean and Northern European breeds analysed (Cymbron et al., 2005). The Channel Island breeds showed strong influence from long ROH reflecting their unusually closed population histories due to strict importation restrictions on both the Jersey and Guernsey Islands implemented during the 1800s (Sonstegard et al., 2011, Wiener et al., 2004).

The zebu breeds represented in this study have contrasting histories. The mainland African zebu breeds (Bororo and Fulani) which are products of ancient introductions from South Asia and are all hybrids to some extent, had much lower quantities of average ROH in comparison to the American zebu breeds analysed (Figure 3.6). Within the American and Madagascan zebu populations a stronger homozygosity signal, with a weighting toward smaller length ROH, suggests that these breeds were initially established by small founding populations but were not particularly affected by recent inbreeding (Figure 3.6). The initial introduction of the now prolific zebu animals in the Americas featured very limited numbers during the 19th and 20th century and Madagascan zebu were founded by ancient importations from Asia and East Africa which were probably limited in scope due to the isolation of the island (Hanotte et al., 2000, MacHugh et al., 1997).

The ascertainment bias (Van Tassell et al., 2009) towards European *Bos Taurus* breeds that is associated with the use of Bovine SNP50 genotyping chip, does not seem to invalidate the trends in ROH levels observed here, as ROH levels were in fact higher in those breeds with a higher number of polymorphic SNPs, as validated by Illumina (Appendix A.7). Also, the existence of long ROH (>20Mb) for example in many of the less polymorphic African village breeds (Oulmes Zaer, Somba and Lagune) are unlikely to be artefactual due to the vanishingly small probability of long continuous homozygous SNPs occurring by chance. However, some bias may exist in the *Bos indicus* ROH levels, as an over estimation of ROH amounts is possible due to the low amount of polymorphic markers found in these breeds due to the design of the genotyping chip (Matukamalli et al., 2009), as a result some caution must be taken when inferring ROH levels within these breeds. The bovine HD genotyping chip was designed from a more comprehensive range of breeds comprising several temperate and topically adapted *Bos taurus*, *Bos indicus* and hybrid breeds and thus does not exhibit the same level of ascertainment bias (Illumina, 2010).

The Hapmap population data also allow comparison with an alternative inference of past population size. Linkage disequilibrium may be used to infer past population size where higher r^2 indicates lower effective population size with LD at longer genetic distances corresponding to younger time depths (Hayes et al., 2003, Tenesa et al., 2007). Interestingly, the Hapmap breed samples analysed here show a strikingly similar ranking in LD at distances of >200 kb to that which they show in average ROH (Gibbs et al., 2009).

Analyses of human ROH have previously established a correlation between extensive LD, locally low rates of recombination and high incidence of homozygous runs (Gibson et al., 2006). Intensive selection intensity in cattle has possibly acted to maintain long lengths of homozygous tracts. Previous work carried out in over 500 animals from 8 breeds noted that high levels of LD, particularly in the Holstein breed, existed on chromosomes 14 and 16, the two chromosomes with highest proportions of ROH in our study (McKay et al., 2007). Conversely, chromosome 12 was found to have higher than average recombination rates and lower levels of LD ($r^2 < 0.2$) than the majority of chromosomes (McKay et al., 2007) and, interestingly, showed the highest proportion of SNPs uninvolved in a ROH within our sample population. There is a possibility of selection events at particular loci underlying the ROH incidence peaks on several chromosomes (Figure 3.2) but this requires further investigation.

The existence of QTL in ROH have been well documented in human studies (Gautier et al., 2010, Hildebrandt et al., 2009, Nalls et al., 2009). Here several of the highly involved genomic regions located on chromosomes 7, 14, 16 and 18 (Figure 3.2) all potentially contain genes of importance in cattle with associations ranging from immunity through to carcass and dystocia related traits (Casas et al., 2003, Kühn et al., 2003, Yun et al., 2011) when explored using three QTL databases available online (<http://genomes.sapac.edu.au/bovineqtl/index.html>,

<http://www.animalgenome.org/QTLdb/cattle.html>, <http://www.ncbi.nlm.nih.gov/>). In particular chromosomes 9 and 5 which had the greatest amount of long ROHs (>20Mb), are well documented to contain QTL pertaining to milk fat yield and weight related traits (Li et al., 2002, Lund et al., 2008, McClure et al., 2010, Wiener et al., 2000).

3.7 Conclusion

ROH analyses quantifies a feature of genomic variation that may be used in inference of population history and to associate with important production and disease traits and perhaps signatures of selection. We show that ROH analysis in cattle provides a sufficient predictor of the pedigree inbreeding coefficient and the prevalent SNP50 genotyping array may be a sufficient tool to predict ROH in the genome. Patterns of ROH may be decomposed to highlight the effects of recent and more ancient ancestral relatedness and match known aspects of breed history in a sample set of wide provenance.

Chapter 4

Genome-wide association study for calving traits in Holstein-Friesian dairy cattle

4.1 Preface

At the time of thesis submission this chapter was published in *Animal* (Published online 20 November 2013; DOI: 10.1017/S175173111300195X). The full reference is:

Purfield DC, Bradley DG, Kearney JF, Berry DP: Genome-wide association study for calving traits in Holstein-Friesian dairy cattle. *Animal* 2014;8(2)224-235

Deirdre Purfield was primary author and performed the SNP edits, GWAS and bioinformatics analysis and drafted the manuscript. Donagh Berry and Daniel Bradley conceived the study, participated in the design and co-ordination of this study and helped draft the manuscript. Francis Kearney estimated the PTAs for analysis in this study.

Formatting and referencing style has been edited for consistency throughout the thesis. Figure and table captions have been assigned with a chapter prefix. Competing interests and acknowledgements have been removed. All other aspects are consistent with the published manuscript.

4.2 Abstract

Dystocia and perinatal mortality are quantitative traits that significantly impact animal productivity and welfare. Their economic importance is reflected by their inclusion in the national breeding goals of many cattle populations. The genetic architecture that influences these traits, however, has still yet to be thoroughly defined. Regions of the bovine genome associated with calving difficulty (direct and maternal) and perinatal mortality were detected in this study using a Bayesian approach with 43,204 single nucleotide polymorphisms (SNPs) on up to 1970 Holstein-Friesian bulls. Several SNPs on chromosomes 5, 6, 11, 12, 17, 18 and 28 were detected to be strongly associated with these calving performance traits. Novel genomic regions with previously reported associations with growth, stature, birthweight and bone morphology were identified in the present study as being associated with the three calving performance traits. Morphological abnormalities are a known contributor to perinatal mortality and the most significantly associated SNP for perinatal mortality in the present study was located in a region in linkage disequilibrium with the gene SLC26A7. This gene, SLC26A7 has similarities and colocalizes with SLC4A2, which has previously been associated with osteoporosis and mortality in cattle populations. The HHIP gene which is known to be associated with stature in humans was strongly associated with direct calving difficulty in the present study; large calves are known to, on average, have a greater likelihood of dystocia. A stemloop microRNA, bta-mir-1256, on chromosome 12, involved in post-transcriptional regulation of gene expression was associated with maternal calving difficulty. Previously reported QTLs associated with calving performance traits in other populations were again identified in this study; with one genomic region on chromosome 18 supporting very strong evidence of an underlying causative mutation and accounting for 2.1% of the genetic variation in direct calving difficulty. Overlapping genomic regions associated with one or more of the calving traits were also detected substantiating the known genetic

covariances existing between these traits. Moreover, some genomic regions were only associated with one of the calving traits implying the selective genomic breeding programs exploiting these regions could help resolve genetic antagonisms.

Keywords: Dystocia, mortality, calving difficulty, genomic, quantitative trait loci

4.3 Introduction

Dystocia and perinatal mortality are quantitative traits that significantly impact animal productivity and welfare (Bicalho et al., 2007, Bicalho et al., 2008, Olsen et al., 2010, Seidenspinner et al., 2011) and are common phenomena in both dairy and beef cattle (Hickey et al., 2007, Mee et al., 2011). The contribution of dystocia and perinatal mortality to profit in cattle production systems is reflected by their inclusion in many national breeding programs. However, the genomic architecture influencing these traits has yet to be thoroughly elucidated.

Dystocia and perinatal mortality are influenced by both environmental factors such as age at first calving and season of calving (Berry and Cromie, 2009, Johanson et al., 2001, McGuirk et al., 1999) and also by genetic factors (Berger et al., 1992, Mee et al., 2011, Meijering, 1983). Heritability estimates range from <0.01 to 0.17 for direct calving difficulty and <0.01 to 0.12 for the maternal components of calving difficulty (Hickey et al., 2007, Johanson et al., 2011, Steinbock et al., 2003). Heritability estimates for direct and maternal components of perinatal mortality are generally lower and range from <0.01 to 0.12 (Steinbock et al., 2003, Johanson et al., 2011). Identification of genomic regions influencing lowly heritable traits typically requires a large, sometimes prohibitive, number

of genotyped animals. Using however accurate measures of genetic merit of sires and their large progeny group sizes can partially overcome this requirement.

Previous quantitative trait loci (QTL) mapping and genome wide association studies have identified several large genomic regions associated with calving difficulty and perinatal mortality. QTL regions on chromosomes 4, 6, 8, 12 and 18 have been associated with calving performance in several studies (Kühn et al., 2003, Olsen et al., 2010, Sahana et al., 2011, Schrooten et al., 2000). However, no sole causative QTL region seems to influence these traits, rather it is the cumulative effect of many of these regions and the allelic variation that lies within these that contributes to the observed phenotype.

The objective of this study was to identify regions of the bovine genome that are associated with dystocia and perinatal mortality in Holstein-Friesian dairy cows. Of particular interest is regions that are associated with only one trait thereby facilitating more targeted breeding that will help resolve genetic antagonisms which are known to exist; for example between direct and maternal calving difficulty (Steinbock et al., 2003).

4.4 Methods

4.4.1 Phenotypic data

Predicted transmitting abilities (PTAs) and their associated reliabilities for 4,683 Holstein-Friesian artificial insemination (AI) sires were obtained from the Irish national genetic evaluation in December 2012 for the direct and maternal component of calving difficulty and the direct component of perinatal mortality. In Ireland, calving difficulty is reported by the producers as a subjectively scored linear scale of 1 to 4, where 1=no calving assistance; 2=slight assistance (assistance by one person, without needing to use a calf puller); 3=considerable assistance (assistance by one person using a calf puller or more than one

person); 4=veterinary assistance (including caesarean). Perinatal mortality is recorded as a binary variable by the producers, indicating whether the calf was dead within a 24hr period.

Parental contribution to the reliability of each PTA was removed using the method described by Harris and Johnson (1998);

$$\tilde{R} = \frac{R_{TRAD} - R_{PA}}{(R_{TRAD} \times R_{PA}) + 1 - 2R_{PA}}$$

where $\tilde{R}$ is the reliability less the parental contribution, R_{TRAD} is the reliability from the traditional genetic evaluation from all relatives and R_{PA} is the parental average reliability. Only sires with an adjusted reliability of >40% for each trait were included in the analysis. Deregressed PTAs ($\tilde{y}$) were obtained as:

$$\tilde{y} = \dot{R}(R^{-1} + A^{-1})\hat{a}$$

where $\hat{a}$ is a vector of PTAs from the domestic Irish national genetic evaluation, A is the numerator relationship of the matrix and $\dot{R}$ is a diagonal matrix where the diagonal elements is

$$\dot{R} = \frac{1 - \tilde{R}}{\tilde{R}}$$

where $\tilde{R}$ is the animals reliability less parental contribution. The relationship matrix was generated by tracing the pedigree of each animal back at least five generations where available. After edits 1,970 animals remained for direct calving dystocia, 1,555 for maternal calving difficulty and 740 animals for direct perinatal mortality. The heritability of direct calving difficulty was 0.11 (s.e 0.0137), maternal calving difficulty was 0.03 (s.e 0.0126) and perinatal mortality was 0.02 (s.e 0.0156) (Ross Evans, ICBF, personal communication, April, 2013).

4.4.2 Genotypic data

Single nucleotide polymorphisms (SNPs) genotypes consisting of 54,001 SNPs from the BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) were available on the Holstein-Friesian AI bulls. Animals with genotype call rates <90% were removed. SNPs with >0.5% Mendelian inconsistencies between parents-offspring, with call rates <95%, monomorphic SNPs and SNPs that deviated from Hardy Weinberg equilibrium ($p < 1 \times 10^{-9}$) were all discarded. SNPs also had to have a GCscore>0.15 and a GTscore>0.55 where GCscore and GTscore are quality measures of the genotype calls from the genotyping assay. All SNPs with unknown position and those located on the X chromosome were discarded. After edits a total of 43,204 SNPs remained across 1,970 of animals. Missing genotypes were imputed using Beagle Version 3.1.0 (Browning and Browning, 2007). Default parameters were used which included 10 iterations to impute missing data.

4.4.3 Whole genome association

Association analyses were undertaken by fitting all SNP simultaneously in a Bayesian framework (Habier et al., 2011). The Bayesian approach was a two-step procedure: 1) Bayes-C π (Habier et al., 2011) followed by 2) Bayes B (Meuwissen et al., 2001). Bayes B is sensitive to the prior of the genetic variance and an inaccurate estimate could impact the results. However, Bayes C π is more tolerant to the prior genetic variance and can provide an accurate estimation of this prior for use in the Bayes B model. The general statistical model can be written as;

$$y = X\beta + u + \sum_{k=1}^K z_k a_k + e$$

where y is a vector of phenotypes, X is the incidence matrix of the fixed effects, β is the vector of fixed effects, K is the number of SNPs, z_k is the column vector representing the genotypes at SNP k, a_k is the additive effect of that SNP and e is a vector of residual effects

(Habier et al., 2011). Both algorithms were undertaken using GenSel, a web based program (<http://bigsc.ansci.iastate.edu/bigsgui/login.html>) developed by Fernando and Garrick (2011). The chain length for both Bayesian analyses was 50,000 iterations, with the first 10,000 excluded as burn in. Deregressed PTA phenotypes were weighted by the respective reliability of the PTA.

Bayes C π was initially used to estimate π , the probability of a SNP having no association with the phenotype under investigation. As proposed by Habier et al (2011), the priors of all SNP effects were assumed to have a common variance and the effect of a SNP fitted with probability $(1 - \pi)$ followed a mixture of multivariate Students t-distributions. Following the estimation of the posterior mean of π from the Bayes C π algorithm, the Bayes B algorithm was then invoked using the posterior mean of π value. This Bayes B method fits a different variance for every covariate (SNP) in the model and is dependent on the prior (π).

Summary statistics of SNPs described hereafter are from the Bayes B analyses. The proportion of post-burnin iterations that included the SNP in the model was calculated as the model frequency and this was used as evidence for an association. Bayes factors were used to quantify the strength of the posterior QTL probabilities for all loci (Kass and Raftery, 1995). Bayes factors can indicate the strength of an association based on their range of values; a Bayes factor >3.1 is indicative of ‘substantial evidence’ that the SNP is associated with a QTL, whereas Bayes factors >10.1 indicate ‘strong evidence’ and values greater than >30.1 indicate ‘very strong evidence’ for association with a QTL (Jeffreys, 1961). The Bayes factor, BF, for comparing opposing hypotheses, was calculated as;

$$BF = \frac{\frac{\Pr(H_1|y)}{\Pr(H_2|y)}}{\frac{\Pr(H_1)}{\Pr(H_2)}}$$

where y is the data, H_1 is the hypothesis that the marker is linked to a QTL and H_2 is the hypothesis that the marker is not linked to the QTL. The prior probability of the first hypothesis and alternative hypothesis was $\Pr(H_1)$ and $\Pr(H_2)$, respectively and $\Pr(H_1|y)$ and $\Pr(H_2|y)$ were the posterior probability of H_1 and H_2 . This can also be expressed in the simpler version of;

$$BF = \frac{\frac{\Pr(H_1|y)}{1 - \Pr(H_1|y)}}{\frac{\Pr(H_1)}{1 - \Pr(H_1)}}$$

The $\Pr(H_1)$ was the posterior mean of $(1-\pi)$ from the Bayes $C\pi$ analysis which was used as the prior for the Bayes B analysis.

The posterior probability of association (PPA) can be thought of as the Bayesian analogue of a p-value and this was calculated for all SNPs (Stephens and Balding, 2009). The PPA combines the evidence in the observed data (i.e., the BF) with the prior probability that the SNP is truly associated with the SNP. To calculate the PPA, the posterior odds (PO) was calculated first, where π was the value calculated from the Bayes $C\pi$ analyses and therefore $1 - \pi$ is the proportion of SNPs that are believed to be associated with the trait;

$$PO = (BF \times (1-\pi)) / (1-(1-\pi))$$

and the PPA was subsequently calculated as;

$$PPA = PO / (1 + PO)$$

The closer the PPA value is to one the more convincing the evidence for an association.

Because of the existence of genomic regions with strong linkage disequilibrium (LD) in the bovine genome (McKay et al., 2007), the effect of any particular QTL and the SNP model frequency may be distributed across numerous SNPs that are in LD with QTL.

This may hinder the identification of true QTL (Fan et al., 2011, Onteru et al., 2011). To accurately identify the presence of QTL, a sliding genomic window of five consecutive SNPs based on genomic location was used to predict the genomic breeding value of the region. This was achieved by summing the contribution of the SNPs to the genetic variance per window. This approach also helps control against false-positives arising from single putative SNPs (Fan et al., 2011, Onteru et al., 2011).

LD blocks were constructed on a 500kb region up and downstream surrounding each of the ten most associated SNPs per trait. LD was estimated using the solid spine of LD option in Haploview, which is an internally developed method that searches for a "spine" of strong LD running from one marker to another (Barrett et al., 2005). Gene search using ensemble (<http://ensembl.org>) on the Btau 4.0 genome build was completed for those regions that were located in LD blocks and also neighbouring LD blocks if they were side by side with the LD blocks containing the SNP of interest and strong LD was seen between SNPs. QTL regions were compared to previously reported QTL in CattleQTLdb for calving difficulty or perinatal mortality as well as likely underlying correlated traits such as birth weight and stature.

Genes within 500kb of SNPs that showed 'substantial evidence' of an association (i.e., $BF > 3.1$) were also identified using Ensembl Build BTA4.0 for all three calving performance traits. Biomart was then used to map the genes in each of these three datasets to their human orthologs using the mapping available from version hg19 of the human genome and GO terms were also identified. For each dataset the R package Goseq, without the correction for gene length bias, was used to identify the KEGG pathways which were significantly over-represented by the set of genes ($p < 0.05$) compared against a background of all genes in the human genome. Functional annotation of the genes was undertaken using DAVID (<http://david.abcc.ncifcrf.gov>).

4.5 Results

4.5.1 Direct calving difficulty

The mean proportion of SNPs estimated not to be associated with direct calving difficulty (i.e. π) was 0.94. The proportion of genetic variance in direct calving difficulty accounted for by all SNPs using the Bayes B algorithm was 0.93 (Table 4.1). The Bayesian posterior probabilities (model frequency) per SNP varied with one SNP included in the model 97.41% of the time (Figure 4.1). Table 4.2 details the Bayes factors, PPAs and minor allele frequency for all 45 SNPs that showed at least 'substantial evidence' (Jeffreys, 1961) of an association with direct calving difficulty. The Bayes factor for the strongest SNP, ss86324977, was 580.21, suggesting 'very strong evidence' of a QTL (Jeffreys, 1961). The SNP ss86324977 accounted for 2.10% of the genetic variance in direct calving difficulty and the surrounding SNPs contributed no more to the genetic variance with the five SNP window containing ss86324977 accounting for 2.11% of the genetic variance. The distance between the flanking SNPs in the window varied either side of ss86324977, with only a distance of ~30 Kb between BTA-97501-no-rs and ss86324977 but ARS-BFGL-NGS-6380 was over 200kb downstream from ss86324977. However, strong LD existed between these flanking SNPs, with a D' value of one (r^2 value of 0.194) between BTA-97501-no-rs and ss86324977 and a D' value of 0.959 (r^2 value of 0.035) between ss86324977 and ARS-BFGL-NGS-6380. The D' was one between these SNPs due to only three of the four possible haplotypes existing in the population. LD analysis undertaken on a 500kb region either side of ss86324977 revealed several small LD blocks (Figure 4.2). The SNP ss86324977 was located in a LD block with the SNP BTA-97501-no-rs. Moderate to strong levels of LD also existed between some of the SNPs in neighbouring LD blocks downstream of ss86324977, with a D' value of 0.711 (r^2 value of 0.013) between ARS-BFGL-NGS-23169 and ARS-BFGL-NGS-100080. Over 83% of the study population were homozygous CC for the SNP ss86324977, whereas only 1% were homozygous AA. The

allele substitution effect for the A allele was +0.576 units (Table 4.2). Animals that were homozygous CC had a lower mean PTA for direct calving difficulty (-3.20 units; s.e 0.06), compared to heterozygote (-1.89 units; s.e 0.08) and homozygous AA (-0.59 units; s.e 0.08).

LD structure upstream and downstream of the top ten SNPs most often included in the model can be seen in Appendix B.1 and genes identified within these LD blocks are listed in Appendix B.2. In total, over 20 genes were identified overlapping with all LD blocks. These genes ranged in function from transcription factors, transporters and membrane proteins. The ss86324977 SNP was located in an intron in the gene sialic acid binding Ig-like lectin 5 (SIGLEC5).

Gene annotation on all 45 SNPs that showed 'substantial evidence' for an association with direct calving difficulty (Table 4.2) revealed that 538 genes or gene-products were identified within a 500kb distance up and downstream. A total of 454 genes mapped to 428 human orthologs. The most significantly over represented KEGG pathway was the riboflavin metabolism signalling pathway ($p=1.54 \times 10^{-3}$). The protein binding GO term, GO:0005515, was the most associated term (12.6%) for genes associated with direct calving difficulty.

Table 4.1 Proportion of genetic variance explained by SNPs and the number of SNPs found in three different Bayesian posterior probability ranges for all traits.

Trait	π	σ^2_g	Posterior Probability		
			>0.95	0.50-0.95	0.15-0.50
Direct Calving Difficulty	0.94	0.93	1	1	77
Maternal Calving Difficulty	0.94	0.96	0	3	55
Perinatal Mortality	0.96	0.96	0	0	5

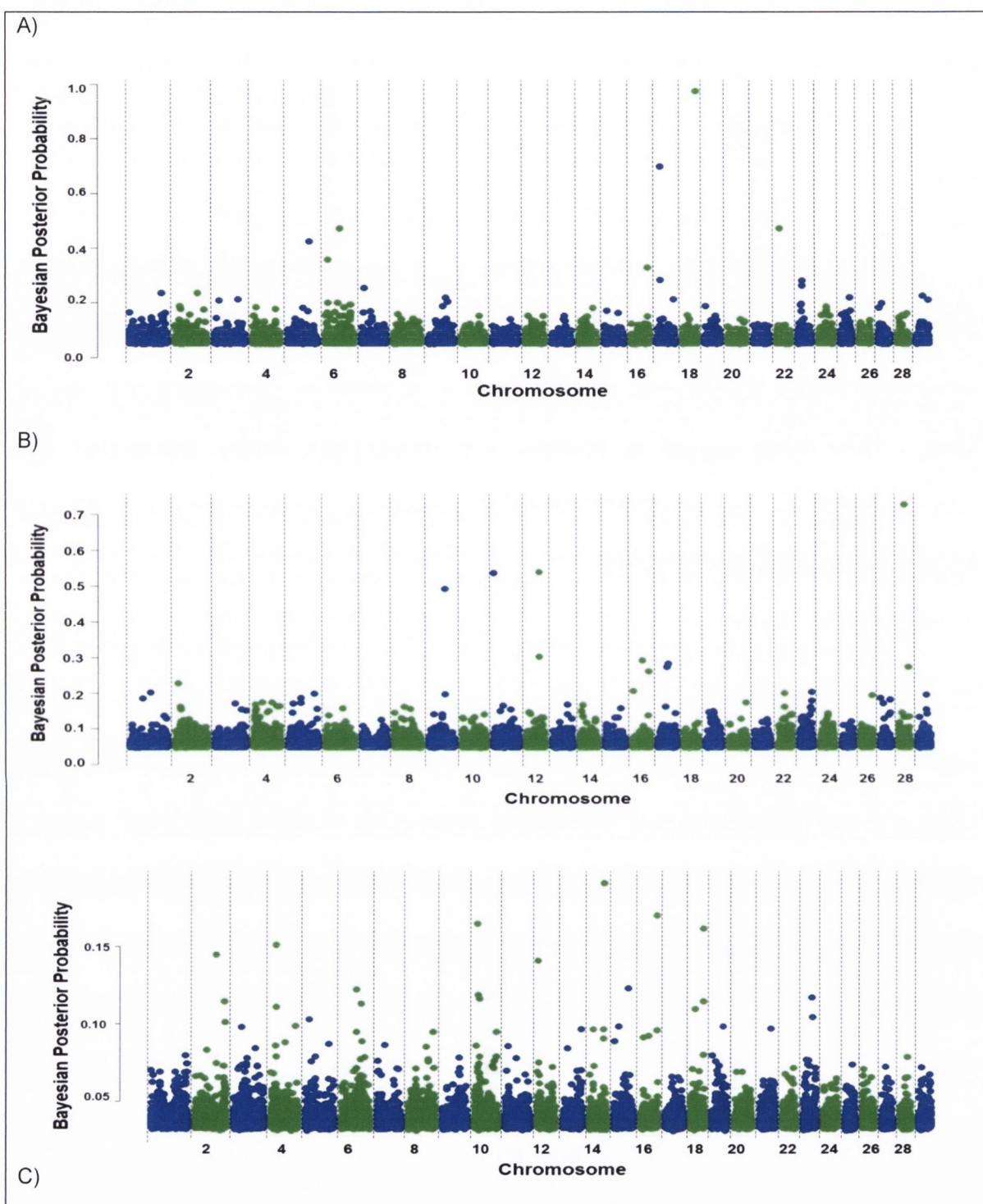


Figure 4.1 The Bayesian posterior probability of model frequency from a Bayes B analysis for calving traits. A) Manhattan plot for direct calving difficulty; BTA 18 had a Single Nucleotide polymorphism (SNP) with the highest model frequency appearing in the Bayesian model 94.7% of the time. B) Manhattan plot for maternal calving difficulty; BTA 28 had a SNP with a model frequency value 73.4%. C) Manhattan plot for direct perinatal mortality; BTA 14 had a SNP with a model frequency value 19.3%.

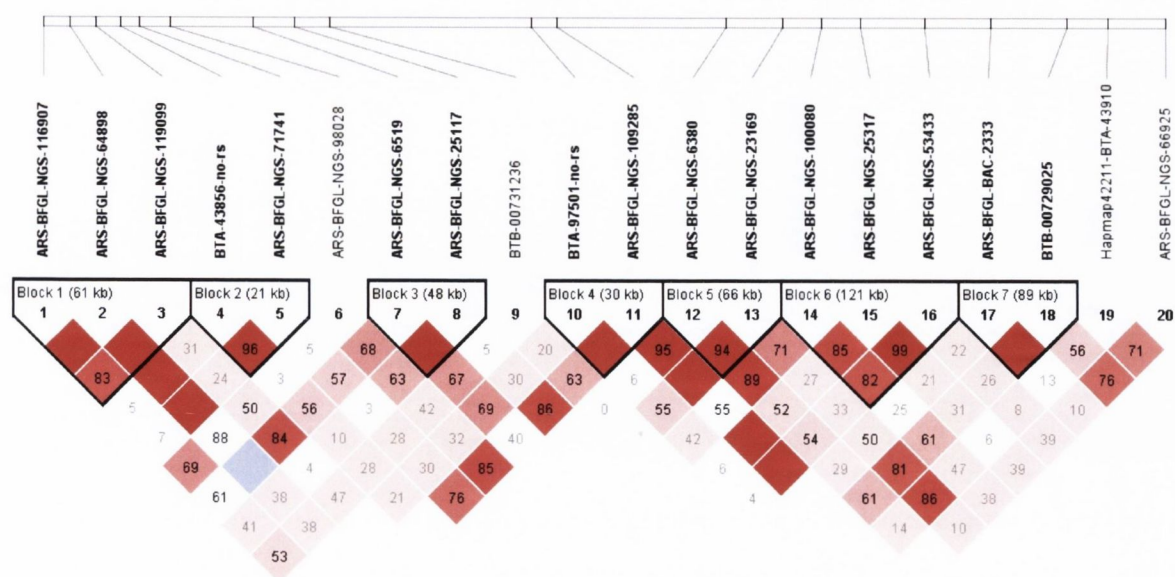


Figure 4.2 Haplotype analysis of a 500kb region up and downstream surrounding ARS-BFGL-NGS-109285 (ss86324977) on chromosome 18. Seven haplotype blocks are outlined with block 4 containing ARS-BFGL-NGS109285 which was the most significantly associated SNP with direct calving difficulty.

Table 4.2 Bayes Factors and posterior probability of associations (PPA) for all SNPs that showed 'substantial evidence' (Bayes factor >3.1) of an association with direct calving difficulty ordered by PPA.

SNP	Chromosome	Btau 4.0 Position	Minor Allele Frequency	Effect of minor allele	Bayes Factor	PPA
ss86324977	18	57125868	0.07	0.576	580.21	0.974
ss117972188	17	14409896	0.28	0.174	35.67	0.698
rs42803295	6	66383308	0.44	-0.084	13.84	0.473
ss86297882	22	12830183	0.38	0.076	13.70	0.472
rs41575151	5	96742248	0.32	0.070	11.44	0.426
ss86284183	6	13429879	0.34	0.049	8.63	0.359
rs41602564	16	73229434	0.19	0.052	7.59	0.330
ss86289496	17	13327070	0.12	0.062	6.11	0.284
ss86339077	23	15477503	0.31	0.035	6.08	0.283
ss86328687	23	16175876	0.18	-0.037	5.53	0.264
ss86329209	7	10305959	0.36	0.027	5.29	0.255
ss86293375	1	138836143	0.31	0.027	4.79	0.237
rs41564989	2	111736094	0.33	-0.026	4.76	0.236
rs41584940	29	26623360	0.48	-0.022	4.49	0.225
rs41623541	9	79690071	0.36	-0.022	4.35	0.220
rs41651840	25	38166957	0.46	-0.022	4.34	0.220
ss86298242	17	70687787	0.06	0.049	4.20	0.214
rs43710473	3	97711220	0.33	-0.023	4.19	0.214
ss86325256	29	51089108	0.42	0.020	4.12	0.211
rs43715268	3	18070495	0.25	0.023	4.07	0.209
ss86336830	9	88580635	0.32	-0.020	3.96	0.204
rs43451309	6	13459802	0.32	0.020	3.90	0.202
ss86324290	6	55916486	0.42	-0.018	3.88	0.201
ss86298059	27	13208360	0.44	0.018	3.84	0.199
ss86339627	23	11627827	0.39	0.019	3.78	0.197
ss86295170	6	109396634	0.12	0.027	3.74	0.195
ss86321501	23	10543971	0.41	-0.018	3.66	0.192
ss86309669	2	20549824	0.35	0.018	3.57	0.188
ss86302661	19	3547083	0.14	0.022	3.56	0.188
rs42059669	9	60721222	0.45	0.017	3.56	0.188
ss86339794	24	33228289	0.31	-0.018	3.51	0.186
rs41610991	6	87904281	0.38	0.018	3.51	0.185
rs41611285	4	13926497	0.45	-0.016	3.50	0.185
rs42239052	27	1423384	0.39	-0.015	3.45	0.183
rs41643203	5	73297801	0.25	-0.018	3.44	0.182
ss86320620	14	62006529	0.36	0.016	3.42	0.182
ss86335870	2	23652490	0.23	0.018	3.34	0.178
ss86314958	4	104575404	0.26	-0.016	3.33	0.178
ss86340336	2	127172772	0.27	0.016	3.31	0.177
rs41589422	15	10052464	0.32	0.015	3.19	0.171
rs29020396	24	32814374	0.43	0.015	3.19	0.171
ss86306426	25	42223581	0.36	-0.014	3.18	0.171
ss117963979	7	39069059	0.36	0.015	3.16	0.170
ss86303481	5	88252891	0.39	-0.014	3.15	0.170
ss61545995	23	11572150	0.32	0.016	3.14	0.169

4.5.2 Maternal calving difficulty

The proportion of SNPs not associated with maternal calving difficulty (i.e. π) was 0.94. The proportion of the genetic variance in maternal calving difficulty accounted for by all SNPs was 0.96 (Table 4.1). The Bayes factors, PPAs and minor allele frequency for all 44 SNPs that showed 'substantial evidence' (Jeffreys, 1961) of an association with maternal calving difficulty are in Table 4.3. A moderate level of heterozygosity and moderate MAF (ranging from 0.13 to 0.50) existed for all 44 SNPs associated with maternal calving difficulty. Only one SNP on chromosome 28, rs41652463, exhibited 'very strong evidence' for association with maternal calving difficulty whereas three of the remaining SNPs showed 'strong evidence' (Jeffreys, 1961) (Figure 4.1). The SNP rs41652463 was included in the Bayesian model for 73% of the iterations, accounting for 0.69% of the genetic variance in maternal calving difficulty with the additional four flanking SNPs in the window containing rs41652463 contributing no additional genetic variance. LD analysis carried out on a 500kb region flanking rs41652463 revealed that although the SNP was not in a LD block, three surrounding LD blocks were assigned (Figure 4.3). Weak LD existed between SNPs flanking rs41652463, suggesting the possibility of recombination hotspot but also that the causal mutation is likely to be very near the rs41652463 SNP.

Two SNPs, ss61478550 and rs41628739 on chromosome 12 were in close proximity, just over 80kb apart and were both associated with maternal calving difficulty; the LD D' value between both SNPs was of 1 (r^2 value was 0.954). Combined, both of these SNPs were included in the model 84.38% of the time suggesting a QTL influencing maternal calving difficulty to be located in the vicinity. Both SNPs were located within the same LD block (Appendix B.3), however no genes were identified within this region but one microRNA did exist, bta-mir-1256, a stemloop microRNA involved in post-transcriptional regulation of gene expression. Several other genes were identified within

LD blocks of the remaining seven most often included SNPs in the model and are detailed in Appendix B.3.

Gene annotation on the 44 SNPs that showed 'substantial evidence' for an association with maternal calving difficulty (Table 4.3) revealed that 313 genes were identified within a 500kb distance up and downstream. A total of 290 genes mapped to 283 human orthologs. The glycosaminoglycan degradation pathway was the most significantly over represented KEGG pathway ($p=1.57 \times 10^{-3}$) and the protein binding GO term, GO:0005515, was the most associated term again (9.8%).

Table 4.3 Bayes Factors and posterior probability of associations (PPA) for all SNPs that showed 'substantial evidence' (Bayes factor >3.1) of an association with maternal calving difficulty ordered by PPA.

SNP Name	Chromosome	Btau4.0 Position	Minor Allele Frequency	Effect of minor allele	Bayes Factor	PPA
rs41652463	28	24176808	0.27	0.245	44.48	0.728
ss61478550	12	57275021	0.34	0.156	19.59	0.541
ss86336751	11	882225	0.41	0.119	19.31	0.537
rs41612473	9	66769980	0.20	0.144	16.15	0.492
rs41628739	12	57357710	0.35	0.070	7.25	0.303
ss86274519	16	46461911	0.29	-0.053	6.87	0.292
ss86307824	17	46074871	0.46	0.046	6.60	0.284
ss105262945	17	39505881	0.33	0.045	6.32	0.275
rs42151848	28	41763375	0.39	-0.045	6.28	0.274
rs41576933	16	72861402	0.33	0.041	5.91	0.262
rs41606071	2	12764123	0.20	0.039	4.94	0.229
rs41583489	16	2962354	0.24	-0.029	4.37	0.208
rs42507912	23	52268620	0.29	-0.028	4.27	0.204
rs43246622	1	91761868	0.26	0.031	4.22	0.202
ss86309427	22	29209152	0.34	-0.029	4.20	0.201
ss86274830	5	116338745	0.50	-0.025	4.15	0.199
rs41662440	9	70354793	0.13	0.042	4.11	0.198
ss117965965	29	34052178	0.35	-0.024	4.06	0.196
ss86315829	26	50729434	0.36	-0.025	4.03	0.195
rs42239052	27	1423384	0.39	0.022	3.83	0.187
rs41637723	5	59419350	0.45	0.024	3.82	0.187
rs43239430	1	54591712	0.13	0.041	3.80	0.186
rs42232531	27	44022775	0.49	0.021	3.71	0.182
ss86274925	23	50924060	0.13	0.032	3.66	0.180
ss86303523	20	72450059	0.43	-0.020	3.55	0.176
rs43385720	4	41929497	0.47	-0.020	3.54	0.175
rs29024319	4	86283586	0.46	-0.021	3.52	0.175
rs41637710	5	59482668	0.47	0.021	3.49	0.173
rs41761090	3	85411459	0.38	-0.020	3.46	0.172
ss86288891	5	21817803	0.36	-0.019	3.46	0.172
rs41649179	27	13757810	0.26	0.021	3.45	0.172
ss86325528	13	63434278	0.30	-0.021	3.40	0.170
rs42598610	4	9865959	0.23	0.022	3.40	0.170
ss117970981	14	11748067	0.45	0.019	3.36	0.168
rs41572779	11	43726129	0.50	-0.018	3.32	0.166
ss86338974	4	122418701	0.46	-0.017	3.29	0.165
ss86288831	4	107126535	0.45	-0.018	3.28	0.165
ss86325459	22	28464322	0.44	0.019	3.25	0.164
ss86336145	2	21388968	0.24	0.020	3.24	0.163
rs42949046	8	44951017	0.50	-0.017	3.24	0.163
ss86291522	17	32171713	0.29	0.019	3.23	0.163
rs41589734	23	47107418	0.49	0.017	3.20	0.161
rs42839025	15	66077414	0.15	-0.023	3.13	0.158
ss86303133	6	74080182	0.33	-0.019	3.13	0.158

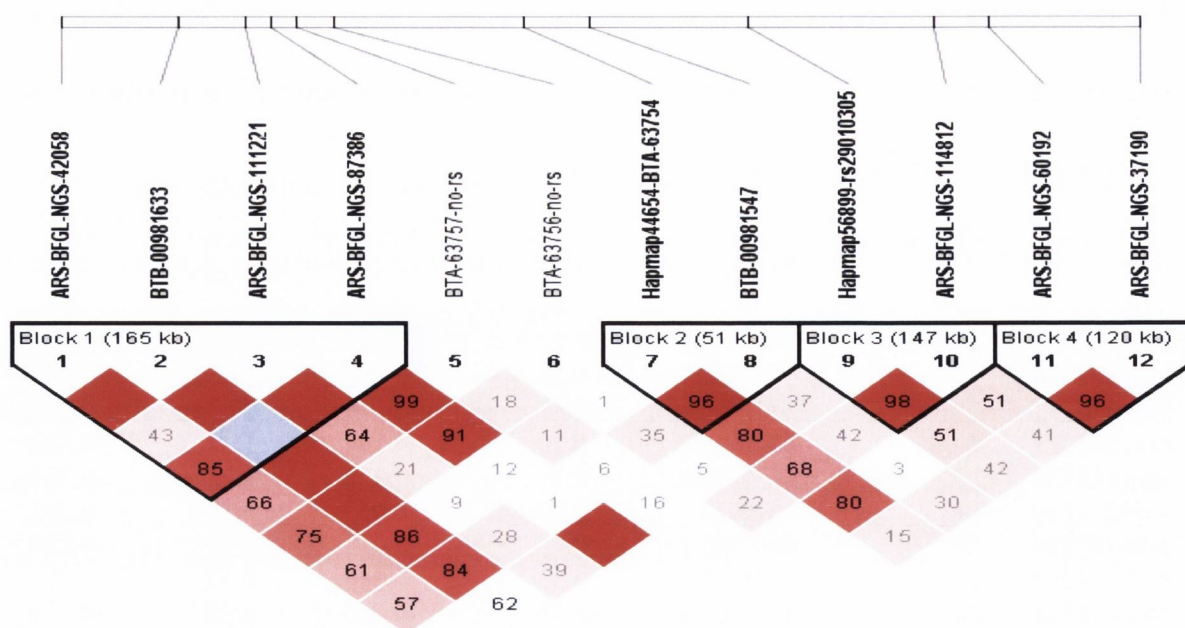


Figure 4.3 Haplotype analysis of a 500kb region up and downstream surrounding BTA-63756-no-rs (rs41652463) on chromosome 28. BTA-63756-no-rs was found to have low LD with adjacent SNPs but is surrounded by two haplotype blocks in the vicinity.

4.5.3 Perinatal Mortality

The mean proportion of SNPs not associated with perinatal mortality (i.e. π) was 0.96. The proportion of genetic variance accounted for by all SNPs was 0.96. Bayesian posterior probabilities were low for perinatal mortality with the highest value of 0.19 for SNP ss86296129 on chromosome 14 (Figure 4.1). Bayes Factors and PPAs were consequently low for this trait with no SNP showing a Bayes Factor value >10.1 , indicating there was no evidence for a ‘strong association’ (Jeffreys, 1961) for this trait (Table 4.4). Although chromosome 14 had the SNP with the highest posterior probability, accounting for 0.0287% of the genetic variance in perinatal mortality, a five-SNP window on chromosome 18 contributed the greatest to the genetic variance (i.e. 0.0338% of the genetic variation). Excluding the SNP on chromosome 18 (MAF of 0.07), a moderate to high level of heterozygosity and consequently high MAF (ranging from 0.17 to 0.49)

existed for all 17 SNPs that showed 'substantial evidence' for an association with perinatal mortality. LD structure of the region surrounding the ss86296129 SNP on chromosome 14 revealed that it was located in a LD block (Figure 4.4). Strong LD existed between the SNP ss86296129 and both ARS-BFGL-NGS-8890 and BTA-110405-no-rs from the neighbouring LD block, with a D' value of 1; therefore gene annotation was undertaken on both LD blocks. No genes were in the immediate vicinity of ss86296129 with the nearest gene (SLC26A7) located >300kb from this SNP. LD structure for the remaining nine SNPs most often included in the model can be seen in Appendix B.5 and several genes identified within these LD blocks and their functions are listed in Appendix B.6

A total of 200 genes were identified within a 500kb distance up and downstream for all SNPs that showed 'substantial evidence' of an association with perinatal mortality. A total of 164 genes mapped to 146 human orthologs. The most significantly over represented KEGG pathway was the systemic lupus erythematosus pathway ($p = 1.64 \times 10^{-5}$) and again the protein binding GO term, GO:0005515, was the most associated gene ontology term (9.6%).

Table 4.4 Bayes Factors and posterior probability of associations (PPA) for the ten most significant SNPs associated with direct perinatal mortality ordered by PPA.

SNP Name	Chromosome	Btau 4.0 Position	Minor Allele Frequency	Effect of minor allele	Bayes Factor	PPA
ss86296129	14	71416566	0.42	-0.019	6.16	0.191
rs41627212	16	74940930	0.49	-0.015	5.33	0.170
rs43616983	10	12799367	0.48	-0.015	5.13	0.165
ss86324977	18	57125868	0.07	0.040	5.02	0.162
ss86316303	4	25183065	0.17	0.019	4.63	0.151
ss86303585	2	108670605	0.49	0.012	4.39	0.145
rs41617236	12	6281991	0.35	0.012	4.27	0.141
ss86315311	15	6092280	0.45	-0.009	3.65	0.123
rs43469091	6	66481495	0.43	0.010	3.63	0.122
ss86299461	10	16734592	0.47	-0.009	3.50	0.119
rs41643479	23	39529264	0.23	-0.010	3.45	0.117
ss105237380	10	22642189	0.20	0.008	3.42	0.116
ss86334674	2	132168987	0.39	0.012	3.36	0.114
ss86340659	18	56515084	0.18	-0.011	3.36	0.114
ss86333023	6	92101327	0.21	-0.007	3.31	0.113
rs43022042	4	25473056	0.44	0.007	3.24	0.111
ss86338216	18	21907405	0.44	-0.010	3.20	0.110

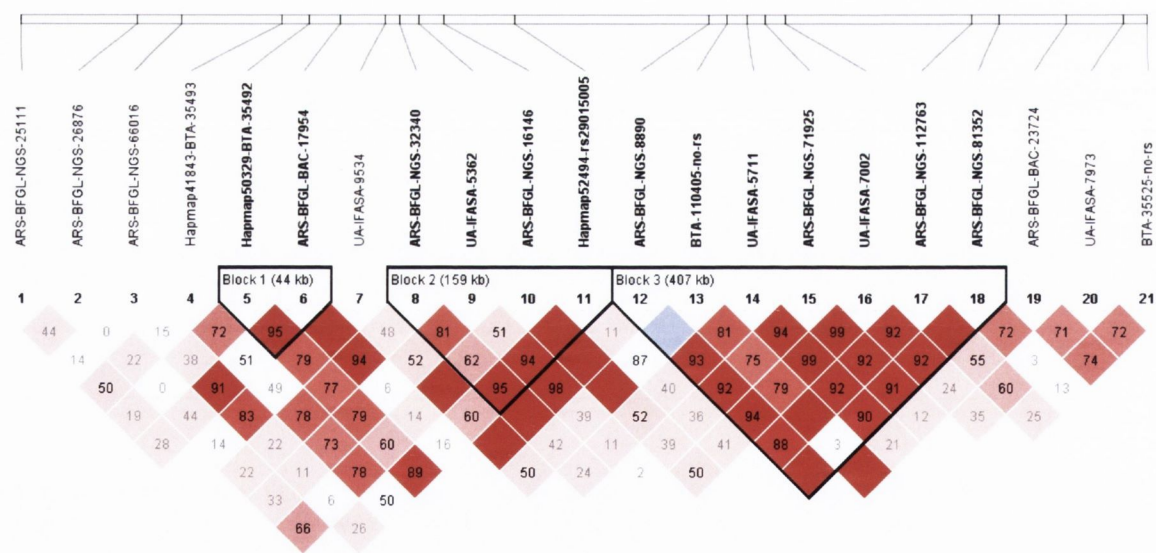


Figure 4.4 Haplotype analysis of a 500kb region surrounding ARS-BFGL-NGS-16146 (ss86296129) on chromosome 14. Three haplotype blocks are outlined with block 2 containing ARS-BFGL-NGS-16146 which was the most significantly associated SNP with perinatal mortality.

4.6 Discussion

In the present study, we identified several SNPs that were ‘strongly associated’ (Jeffreys, 1961) with calving performance traits in Irish dairy cows. Associations existed on 16 different chromosomes, some of which have previously been documented to be associated with calving traits (i.e. ss86324977) while others are novel previously unreported associations but have been reported to be associated with growth, skeletal development, birth weight and stature (Cole et al., 2009, McClure et al., 2010, Meyers et al., 2010). QTL regions that were solely associated with one trait were also identified facilitating direct genomic selection for these regions thereby possibly aiding in resolving known genetic antagonisms between the calving performance traits (Luo et al., 1999, Steinbock et al., 2003). However, caution has to be taken when interpreting these QTL regions, as the power of detecting QTLs associated with perinatal mortality in comparison to the other traits was reduced due to the lower quantity of records.

4.6.1 Chromosome 18

The SNP, ss86324977 on chromosome 18 which was most strongly associated with direct calving difficulty in the present study has previously been documented to be associated with calving difficulty in US (Cole et al., 2009) and both Danish and Swedish (Höglund et al., 2012, Sahana et al., 2011) dairy cattle populations. This SNP is located in an intron of the sialic acid binding Ig-like lectin 5 gene (Siglec-5). Cole et al (2009) reported this gene to be associated with a variety of calving traits such as sire and daughter calving ease, rump width, stature and body depth. In humans, a similar gene Siglec-6, is expressed in the placenta and may play a role in the initiation of parturition (Brinkman-Van der Linden et al., 2007). Cole et al (2009) hypothesized that animals homozygous 'AA' for the SNP ss86324977 may experience a delayed parturition due to a lectin deficiency caused by high

levels of sialic acid binding Ig-like lectin 5. The high frequency in the present population of the favourable allele associated with reduced difficulty calving implies past breeding programs have selected against the AA genotype of this SNP or that it has been subjected to natural selection pressure. Qanbari et al (2011) confirmed the presence of a strong selection signature using both iHS and F_{ST} analyses in the vicinity of the Sialic acid binding Ig-like lectin 5 gene. Qanbari et al (2011) suggested that this selection signature is the outcome of strong and recent selective pressure due to the long run of homozygosity observed in this region. This SNP ss86324977 was also shown to be associated with perinatal mortality in the present study thereby substantiating the known genetic association between both traits (Luo et al., 1999); the allele substitution effect for the A allele of this SNP was positive for both direct calving difficulty and perinatal mortality. This suggests this genomic region harbours either a gene with pleiotrophic effects on both traits or harbours two or more linked (i.e. co-inherited) genes affecting both traits.

4.6.2 Chromosome 17

Four of the top ten SNPs for direct and maternal calving difficulty were located on chromosome 17. The SNP ss117972188, the second most associated SNP with direct calving difficulty, was located in the gene HHIP whose orthologue has been shown to be associated with human stature (Liu et al., 2010). However, this HHIP gene has not previously been identified as a significant SNP controlling stature in cattle (Pryce et al., 2011). Stature has been previously correlated with the direct effects of calving difficulty, with larger calves more likely to experience a difficult birth (Bennett and Gregory, 2001). The remaining three SNPs (ss86289496, ss86307824 and ss105262945) have not previously been documented to be associated with calving traits in cattle. However, SNP ss86289496 was located in close proximity to zinc finger protein 827 and these proteins,

which are well known transcription factors, have recently been reported to play an important role in the specific regulation of mammalian bone and skeletal development with zinc deficiencies known to result in skeletal growth retardation (Cole et al., 2009, Ganss and Jheon, 2004); this therefore is a likely positional candidate gene.

The allele effects for these four SNPs (ss117972188, ss86289496, ss86307824 and ss105262945) corroborate the negative genetic correlation that is known to exist between direct and maternal calving difficulty (Steinbock et al., 2003) with the allele effects of ss117972188 and ss86289496 having a positive effect on direct calving difficulty and a negative effect on maternal calving difficulty and vice a versa for SNP ss86307824.

4.6.3 Chromosomes 2, 16 & 12

Birth weight is known to be non-linearly associated with dystocia and perinatal mortality, with lighter and heavier than average calves tending to have an increased risk of mortality and dystocia (Berger et al., 1992). Putative QTL regions affecting birth weight are well documented on chromosome 2; Grosz & MacNeil (2001) reported that approximately 2.8 kg difference of birth weight could be accounted for by a QTL residing at the 114 cM region. In the present study, the SNP ss86303585 on chromosome 2 associated with perinatal mortality was in close proximity to this QTL region described by Grosz & MacNeil (2001), suggesting this SNP may be in a region influencing birth weight and therefore perinatal mortality. Gene analysis also revealed that the Insulin-like growth factor-binding protein 2 Precursor gene (IGFBP-2) was located just 140kb downstream of the ss86316303 SNP. These IGF-binding proteins inhibit insulin like growth factors with IGFBP-2 believed to play an important role as a growth promoting hormone during gestation (Fowden, 2003) therefore making it a logical candidate gene. McClure et al., (2010) previously associated several QTL regions with birth weight, two of which

overlapped with regions identified in the present study; the QTL region around 72-74Mb on chromosome 16 associated with all three traits and the 57 Mb QTL region on chromosome 12 associated only with maternal calving difficulty.

4.6.4 Chromosome 23 & 28

Evidence of a strong QTL for maternal calving difficulty containing the SNP rs41652463 existed on chromosome 28, in close proximity to a QTL documented previously to be associated with direct but not maternal calving difficulty by Olsen et al., (2010). Similarly, a QTL associated with direct calving difficulty on chromosome 23 in the present study was located 1Mb downstream of a QTL documented in Danish and Swedish cattle populations (Sahana et al., 2011, Höglund et al., 2012) that influenced maternal calving traits such as the calving index, maternal calving ease and maternal still birth in first parity animals. An earlier study by Seidenspinner et al (2009) also identified a QTL influencing direct calving difficulty within this region. This suggests QTL influencing calving traits are located within these genomic regions on chromosome 23 and chromosome 28. The QTL on chromosome 28 was not located in a LD block but several genes associated with the nucleus and transportation were located within a 500kb distance with a small nucleolar RNA belonging to the SNORA36 family and the graves disease carrier protein being in the closest proximity. Similarly on chromosome 23, no genes were identified in the LD block containing SNP ss86328687 however, several genes associated with mRNA enzymes and triggering receptors on myeloid cells were identified within 500kb of the QTL region surrounding both ss86328687 and ss896339077.

4.6.5 Chromosome 10 & 14

Three possible candidate genes (SLC26A7, TMEM55A & TMEM64) were identified within the LD block containing ss86296129 on chromosome 14, the SNP most strongly associated with perinatal mortality. Two of these genes, SLC26A7 and TMEM64 have been documented to be associated with cartilage and bone development in humans (Kim et al., 2013, Lohi et al., 2002) suggesting that these genes may be responsible for the many morphological abnormalities associated with perinatal mortality. In cattle, a deletion in a similar gene (SLC4A2) has been shown to be associated with osteoporosis in Red Angus cattle (Meyers et al., 2010). Osteoporosis is a bone disorder observed in humans and many other animals and can be inherited as a recessive defect. Loss of SLC4A2 function induces premature cell death and cytoplasmic alkalinization of osteoclasts (Meyers et al., 2010). The death of osteoclasts which are large multinucleated cells that reabsorb bone, can result in the formation of extremely dense fragile bones and affected calves are typically born stillborn, slightly premature and morphologically small with a flat skull (Meyers et al., 2010). Selecting on the causative mutation tagged by the ss86296129 SNP could aid in reducing both direct calving difficulty and perinatal mortality; a moderate MAF exists for this SNP signifying scope for selection.

The two QTL on chromosome 10 for perinatal mortality were not in the vicinity of QTL previously reported to be associated with perinatal mortality on this chromosome (Höglund et al., 2012, Kühn et al., 2003, Sahana et al., 2011, Schnabel et al., 2005, Seidenspinner et al., 2011). Two candidate genes (RAB11A and MEGF11) were identified in the LD block of rs43616983. The gene RAB11A which encodes the Rab11A small GTPase, has been shown to play an important role in mediating transferrin recycling back to the plasma membrane and may play a role in innate immunity (Ullrich et al., 1996). A weak innate immune system has previously been documented to contribute to perinatal mortality (Kirkbride, 1993).

4.6.6 Pathway analysis

Several KEGG pathways were associated with calving performance traits of which the systemic lupus erythematosus pathway was the most over represented for perinatal mortality. Systemic lupus erythematosus is an autoimmune disease that affects connective tissues suggesting that the genes associated with perinatal mortality are primarily of an immune function. Systemic lupus erythematosus has been associated with perinatal mortality in humans where an increased risk of spontaneous abortion, intrauterine foetal death, preclampsia, intrauterine growth retardation and preterm birth has been documented (Clowse et al., 2005). The riboflavin metabolism pathway was the most over represented pathway for direct calving difficulty. Riboflavin, an important micronutrient, is required for the metabolism of fats, ketone bodies, carbohydrates, and proteins, and consequently has been shown to impact growth (Chan et al., 2010). This impact on growth may be why it was the most over represented KEGG pathway for this trait.

Glycosaminoglycans have been documented to be associated with the reproductive cycle in many species. Their composition and concentration vary at different stages throughout the cycle and they have been documented to be important regulators of cervical function as they contribute to the visco-elastic properties of the connective tissue present there (Osmers et al., 1993). The glycosaminoglycan degradation pathway was the most over represented pathway for maternal calving difficulty possibly due to the continuous role they play throughout pregnancy particularly with cervical ripening and dilatation being associated with vast changes in glycosaminoglycan pattern (Osmers et al., 1993).

4.7 Conclusion

This study confirmed several different QTL regions previously documented to be associated with calving traits but also several novel QTL regions. Identification of QTLs that are only associated with one calving traits have also been identified allowing for direct selection through genomic breeding programs while overcoming the genetic antagonism that exists between calving traits. Two genes of particular interest identified in this study, Siglec-5 and SLC26A7, have been shown to impact calving traits, suggesting that there is the existence of influential calving trait genes.

Chapter 5

Genome-wide association study for calving performance
using high density genotypes in dairy and beef cattle.

5.1 Preface

At the time of thesis submission this chapter was submitted to Genetics Selection Evolution.

Deirdre Purfield was primary author and performed the GWAS and bioinformatics analysis and drafted the manuscript. Donagh Berry and Daniel Bradley conceived the study, participated in the design and co-ordination of this study and helped draft the manuscript. Francis Kearney and Ross Evans estimated the EBVs for analysis in this study.

5.2 Abstract

Calving difficulty and perinatal mortality are prevalent in modern-day cattle production systems. It is well-established that there is a genetic component to both traits yet little is known about their underlying genomic architecture, particularly in beef breeds. Therefore, the objective of this study was to perform a genome-wide association study using Illumina bovine high-density beadchip genotypes to elucidate the genomic architecture and to identify regions of the genome associated with these traits. Regions of the bovine genome associated with calving difficulty (direct and maternal) and perinatal mortality were detected in this study using two statistical approaches: 1) single SNP regression and 2) a Bayesian approach. Data included high-density genotypes on 2,660 Holstein-Friesian (n=770), Charolais (n=927) and Limousin (n=963) bulls. Several novel and previously identified genomic regions were detected; however associations often differed by breed. A genomic association in the Holstein-Friesian population on chromosome 18 contributed 2.49% of the genetic variance in direct calving difficulty whilst an association on chromosome 2 explained 3.13% in the Charolais population. Analysis of imputed genome sequence data was used to refine significant associations in the Holstein-Friesian sample. Several candidate genes on chromosome 18 and four highly significant missense variants ($p < 2.5 \times 10^{-8}$) located within three genes (SIGLEC12, CTU1, ZNF615) were identified. However, only CTU1, contained a missense variant which may impact direct calving difficulty (SIFT=0.06; Polyphen=0.95). Imputed sequence data refined an association for maternal calving difficulty on chromosome 4 in the Holstein population where the strongest association was an intronic variant in the PCLO gene. A meta-analysis was also performed for each calving performance trait to identify common genomic regions of association in the three breeds and results suggest that a proportion of the genetic variation in calving performance is common across breeds. The genomic architecture of calving performance is complex and primarily influenced by many polymorphisms each of small

effect; associations of moderate effect were identified, however these were primarily breed specific. This suggests different breed allele effects for calving performance exist or else the linkage phase between the genotyped allele and the causal mutation varies per breed.

Keywords: calving difficulty, perinatal mortality, genomic, sequence, dairy, beef, GWAS

5.3 Introduction

Dystocia, more commonly known as calving difficulty, is defined as a prolonged or difficult parturition, often with assistance required during delivery. The adverse effects of calving difficulty have been well documented including an increased risk of both dam and calf mortality, as well as a reduction in dam and calf performance which cumulatively impact on herd profit (Bicalho et al., 2007, Lombard et al., 2007, Mee et al., 2011). The importance of dystocia and perinatal mortality in modern day cattle production systems has been demonstrated by their inclusion in breeding programmes (Eriksson et al., 2004, Mee et al., 2011). The international prevalence of both calving difficulty and perinatal mortality, however, still remains unacceptably high with a range from 2% to 14% of difficult calvings and a prevalence of perinatal mortality of between 2% and 10% (Mee et al., 2008, Meyer et al., 2001, Steinbock et al., 2003). Genetic variation nonetheless exists for calving performance traits with heritability estimates ranging from <0.01 to 0.17 for calving difficulty and <0.01 to 0.12 for perinatal mortality (Hickey et al., 2007, Johanson et al., 2011, Steinbock et al., 2003).

Calving difficulty and perinatal mortality are both complex quantitative traits, believed like all traits of this nature, to be influenced by many genomic polymorphisms of individually small effects (Hayes et al., 2010). The accuracy of genomic predictions depends in part on the genomic architecture of the trait such as the number of polymorphisms affecting the

trait and the distribution of the size of their effects (Hayes et al., 2010). Accurate across-breed genomic predictions rely on common underlying genomic variations of common allele substitution effect. Consequently, it is important to determine the number of polymorphisms and the size of their effects on calving difficulty and perinatal mortality to facilitate potentially more accurate genomic predictions. Genome-wide association studies (GWAS) in cattle have detected QTL associated with calving difficulty and perinatal mortality and these were generally concentrated on chromosomes 6, 11, 12, 18 and 28 (Cole et al., 2009, Olsen et al., 2010, Purfield et al., 2013); these studies, however, have mainly focused on dairy breeds.

The objective of this study was to perform a GWAS using Illumina bovine high-density beadchip genotypes (i.e., 777,962 SNPs) to identify regions of the genome associated with three calving performance traits; (i) direct calving difficulty, (ii) maternal calving difficulty and (iii) direct perinatal mortality, in three cattle breeds (Holstein-Friesian, Charolais and Limousin) and to define genomic associations using whole genome sequence data.

5.4 Methods

5.4.1 Genotypic data

Illumina bovine high density SNP genotypes were available on 2,660 dairy and beef bulls. The 2,660 animals consisted of 927 Charolais, 770 Holstein-Friesian, and 963 Limousin. All animals had a genotype call rate >95%; where genotypes were available on both sire and son(s), the integrity of the animal genotype identity was confirmed. Mendelian inconsistencies between sire-son pairs were determined; 1,929 sire-son pairs existed. A total of 3,554 SNPs had >2% mendelian inconsistencies and were therefore discarded. The genotype of both the sire and son were set to missing for the remaining SNPs where

sporadic mendelian inconsistencies existed. A total of 1,574 SNPs with a GenTrain score <0.55 were discarded as were 40,934 non-autosomal SNPs; duplicate SNPs on the Illumina HD manifest were also discarded. Of the remaining SNPs, 17,273 with a call rate of $<95\%$ were discarded and all SNPs with a MAF <0.02 within each breed were discarded within that breed but retained in the other breeds. Finally, any SNP that deviated ($P < 0.1 \times 10^{-8}$) from HWE within breed were discarded. Following all edits, 605,718 autosomal SNPs remained for inclusion in the analysis in the Holstein-Friesian breed, 602,372 autosomal SNPs in the Charolais breed, and 600,980 autosomal SNPs in the Limousin breed. Missing genotypes were imputed using Beagle (Browning and Browning, 2009, Browning and Browning, 2007).

5.4.2 Phenotypic data

Estimated breeding values (EBVs) and their associated reliabilities for all 2,660 dairy and beef bulls were obtained from the Irish multi-breed national genetic evaluation in December 2013. The direct and maternal component of calving difficulty and the direct component of perinatal mortality were obtained from this evaluation; there is currently no maternal component estimated for perinatal mortality in the Irish national genetic evaluations. In Ireland, calving difficulty is subjectively scored by producers on a linear scale of 1 to 4, where 1 = No calving assistance; 2 = Slight assistance (assistance by one person, without needing to use a calf puller); 3 = Considerable assistance (assistance by one person using a calf puller or more than one person); 4 = Veterinary assistance (including caesarean). Perinatal mortality is recorded as a binary variable by producers, indicating whether the calf was dead within a 24 hour period. It is a legal requirement in Ireland to record all animal deaths, including stillbirths. Phenotypic data were edited as described in detail by Purfield et al., (2013). EBVs for each trait were deregressed and only

sires with reliability, less parental contribution, hereon in referred to as adjusted reliability, of >30% were retained. After edits 2,531 animals remained for analysis of direct calving dystocia, 1,275 for maternal calving difficulty and 2,244 animals for perinatal mortality.

5.4.3 Whole genome association analysis

Association analyses were undertaken within each breed separately using two methods; single SNP regression and Bayesian methods.

5.4.3.1 Single SNP regression

Single SNP regression (SSR) analysis was undertaken in Wombat (Meyer, 2007) using a mixed model framework. Each SNP, as allele count, was individually included one at a time as a fixed effect covariable in the model. Relationships among all animals were accounted for via the numerator relationship matrix. The weight on the dependent variable (Garrick et al., 2009) was

$$w_i = \frac{1 - h^2}{[c + (1 - r_i^2)/r_i^2]h^2}$$

where h^2 is the heritability of the trait, r_i^2 is the adjusted reliability of the animal and c is the proportion of genetic variation for which genotypes cannot account, valued at 0.9 for the SSR analyses. Several values of c were tested (i.e 0.1, 0.2, 0.8, 0.9) with minimal impact on the results, so a value of 0.9 was chosen for SSR which allowed a single SNP to be weighted to attribute to 10% of the genetic variance in calving performance traits. Boddihiredy et al., (2014) also reported no difference in prediction accuracy when comparing c values of 0.1, 0.5 and 0.7. Test statistics for all SNPs were obtained. Multiple testing correction was applied using the Bonferroni correction method and p-values were transformed into their corresponding q values assuming a false discovery rate of 5% (Dabney and Storey, 2004).

Within-breed SSR p-values for each SNP were combined, within trait, using the weighted Z-score method. The combined weighted Z-score was calculated as:

$$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}}$$

where the weight w_i was the square root of sample size of breed_{*i*} and $Z_i = \Phi^{-1}(1 - \frac{p_i}{2}) * (\pm \text{one for the effect direction for each breed})$ where Φ is the standard normal cumulative distribution function and p_i is the p-value for that SNP in breed_{*i*}. All SNP Z-statistics per breed (Z_i) were corrected for multiple testing prior to combining all values. This multiple correction test involved dividing each Z-statistic (Z_i) per breed by a lambda factor, computed as the median of all Z-statistics in breed_{*i*} divided by the expected median of the Z-statistics under the null hypothesis of no association in breed_{*i*}. Weighted Z-scores were then transformed into their corresponding p-values.

5.4.3.2 Bayesian analyses

The Bayesian approach was undertaken by fitting all SNP simultaneously using a two-step Bayes procedure, using two different Bayesian models: 1) Bayes C (Habier et al., 2011) followed by 2) Bayes B (Meuwissen et al., 2001). Bayes C was implemented first to provide an accurate estimate of σ_g^2 and σ_e^2 for use in the Bayes B model as Bayes B is sensitive to the prior of the genetic variance and an inaccurate estimate could impact the results. The BayesB method assumes that many SNPs will have no association with the phenotype. Therefore, BayesB is dependent on the prior probability of a SNP having no association with the phenotype under investigation (π) and will fit a mixture distribution that assumes any SNP might come from a continuous distribution or distribution towards zero (Meuwissen et al., 2001). The π value was estimated as one minus half the number of animals in the population divided by the total number of SNPs included in the analysis.

Bayes B was invoked using the posterior genetic and residual variance estimates from Bayes C and the calculated π value. The chain length for all Bayesian analyses was 50,000 iterations, with the first 10,000 discarded as burnin. The dependent variable was weighted as described previously but the value of c (proportion of genetic variation for which genotypes cannot account) was assumed to be 0.1. As a Bayesian method simultaneously fits all SNPs into the model, a value of 0.1 was chosen as it suggests genotypes should account for 90% of the genetic variation. All Bayesian algorithms were undertaken using GenSel, a web based program (<http://big.s.ansci.iastate.edu/bigsgui/login.html>) developed by Fernando and Garrick (Fernando and Garrick, 2011) within each breed separately.

Bayes factors (BF) were used to quantify the strength of the posterior QTL probabilities for all loci (Kass and Raftery, 1995) using the formula;

$$BF = \frac{\frac{\Pr(H_1|y)}{1 - \Pr(H_1|y)}}{\frac{\Pr(H_1)}{1 - \Pr(H_1)}}$$

$\Pr(H_1)$ was calculated as $1-\pi$ (i.e., the probability of a SNP having no association with the phenotype) and the proportion of post-burnin iterations that included the SNP in the model known as model frequency, was used as evidence for an association. Bayes factors can indicate the strength of an association based on their range of values; a Bayes factor >3.1 is indicative of ‘substantial evidence’ that the SNP is associated with a QTL, whereas values >10.1 indicates ‘strong evidence’ and values greater than >30.1 indicates ‘very strong evidence’ for association with a QTL (Jeffreys, 1961). The proportion of genetic variance that each SNP contributed was calculated by dividing the genetic variance attributed to that SNP by the posterior mean estimate of the genetic variance of the trait.

5.4.4 Bioinformatics

Haplotype were identified in a 100kb region up and downstream of each SNP that had a Bayes factor in either of the breeds of >200 for direct calving difficulty and maternal calving difficulty, or >60 for perinatal mortality. These thresholds were chosen, as they limited our haplotype analysis to the strongest associations identified for each trait. Gene search using ensemble (<http://ensembl.org>) and NCBI map viewer (<http://www.ncbi.nlm.nih.gov/mapview/>) on the UMD 3.1 genome build was then completed with emphasis on linkage-disequilibrium (LD) blocks that contained the SNPs of interest. QTL regions were compared to previously reported QTL in CattleQTLdb for calving difficulty or perinatal mortality as well as likely correlated traits such as birth weight and animal size.

5.4.5 Sequence analysis

Whole genome sequence data completed as part of the 1000 Bull Genomes Project (Daetwyler et al., 2014; <http://www.1000bullgenomes.com/>) was available on a total of 311 Holstein-Friesian bulls, of which 64 were HD genotyped within this study. A total of 35.2 million SNP variants were identified across the genome and all sequenced animals were imputed using Beagle software to provide a consensus SNP density across all animals by the 1000 bull genomes project. The average genome coverage was 11.0X. Significant areas of the genome identified to be associated with calving performance, from the HD GWAS in our Holstein-Friesian sample were further investigated by imputation to sequence density and re-analysing using SSR. These regions were chosen from clear peaks of association identified in the HD GWAS with log p-values >8 (chromosomes 18, 2 and 4) or else from results suggested from previous studies which indicate that a QTL impacting calving performance is located on this chromosome (chromosomes 6 and 10)

(Olsen et al., 2010, Seidenspinner et al., 2011). Therefore where possible a 10Mb region surrounding the strongest associations on each of these chromosomes (chromosome 2: 6.6Mb, chromosome 4: 37.85Mb, chromosome 6: 72.02Mb, chromosome 10: 101.72Mb and chromosome 18: 57.58Mb) was imputed to sequence depth and re-analysed. The integrity of the animal genotype identity between the HD genotype and sequence genotype where possible was confirmed and on average across all genomic regions the concordance >99.6%. Post Beagle imputation accuracy was on average >95% and all SNPs were retained for analysis. Due to the limited number of Charolais and Limousin bulls with whole genome sequence data available, imputation to sequence density was unsuccessful in these breeds. This was deemed unsuccessful as an excessively large proportion of SNPs were imputed as monomorphic within these breeds (~10% of SNPs within an imputed region were polymorphic). Any potential missense variants identified within genomic regions of strong association were analysed using the bioinformatics tools, sorting intolerant from tolerant (SIFT) (Ng and Henikoff, 2003) and polymorphism phenotyping (PolyPhen) (Adzhubei et al., 2010) to predict the pathogenicity of these variants. The variant effect predictor from Ensembl (<http://www.ensembl.org/info/docs/tools/vep/index.html>) was used to provide SIFT scores where possible and PolyPhen scores were computed where estimated in the present study.

5.4.6 Copy Number Variation

Copy number variation (CNV) in all three breeds was also analysed to assess the impact of structural variants on calving performance. To identify CNVs, all signal intensity data (Log-R ratio; LRR) and allelic intensity (B-allele frequency; BAF) ratios generated from the Illumina Beadstudio 3.2 software were retained for every SNP on each animal. A total of 344 Holstein-Friesian, 696 Charolais and 725 Limousin had this signal intensity data

available. The BAF is a normalised measure of the intensity of the 'B' allele signal when the beadchip is scanned whereas the LRR is a measure of the overall signal intensity measured from the beadchip array. All animals passed quality control thresholds which included >95% call rate and removal of non-autosomal SNPs (42,669) and SNPs with a LRR deviation <30% (39,306). After edits a total of 695,987 SNPs remained on 1,762 animals.

Genomic regions of putative CNV were identified using QuantiSNP v2 using the default parameters. QuantiSNP uses an objective-Bayes (OB) hidden Markov-model to infer CNV. QuantiSNP also produces Bayes Factors (reported as the logarithm of Bayes Factors) values which can be used to set a cut-off false positive rate for all CNV detected. A stringent cut off rate of Log (Bayes factors) >30.1 was applied in this study to ensure a false positive rate <1% and only CNVs >10kb in length were retained. Copy number loss or gain, length and frequency were determined. QuantiSNP identifies five different types of CNVs; 0 is full deletion of both copies of the genomic region, 1 is single copy deletion, 2 is normal homozygous/heterozygous, 3 is single copy duplication and 4 is double copy duplication. CNV burden association mapping was undertaken using a 100,000 permutation two-sided tests test. Significant trait associated SNPs identified through the HD GWAS for each trait were examined for the presence of a CNV. The numbers of genes overlapped by common CNVs were also identified in each population.

5.4.7 Runs of homozygosity

Runs of homozygosity (ROH) which are the contiguous lengths of homozygous genotypes that are present in an animal due to parents transmitting identical haplotypes to their offspring were identified in all three breeds as described in the methods in section 3.3.2. The sum of ROH per animal was computed and regressed against their corresponding EBV for each phenotype. To further examine the association between ROHs and calving

performance, overlapping ROH groups will be identified (i.e. all the individual ROHs in a group include a consensus core region) and a Students t-test will be completed on each ROH group to assess their impact. Also significant genomic regions associated with calving performance in the HD panel will be examined for ROH.

5.5 Results

5.5.1 Direct calving difficulty

Direct calving difficulty refers to the characteristic of the calf itself (e.g., body size) and its impact upon parturition. Several SNP associations for direct calving difficulty were detected in each of the three breeds, although the exact location and direction of effect of these associations differed by breed. Irrespective of statistical significance the sign of the estimated allele effects at 283,042 loci differed between the Holstein-Friesian and Charolais populations, while alleles at 285,133 loci differed in the sign of the effect between Holstein-Friesian and Limousin populations, and alleles at 284,509 loci differed in the sign of the effect between the Charolais and Limousin populations. Twenty three SNPs located on chromosomes 18 (10 SNPs) and 6 (13 SNPs) had a q-value less than 0.05 for direct calving difficulty in Holstein-Friesians. The SNP ARS-BFGL-NGS-109285, located on chromosome 18 (Figure 5.1A), exhibited the strongest association with direct calving difficulty in both the SSR and Bayesian analyses (Bonferroni corrected $p = 5.5 \times 10^{-4}$; $q = 2.0 \times 10^{-4}$; BF= 1684.26). Both ARS-BFGL-NGS-109285 and an adjacent SNP, BovineHD180001676 (41kb apart), were cumulatively included in the Bayesian model at a frequency of 94.96% and together accounted for 2.49% of the genetic variance of direct calving difficulty. Alleles in both SNPs were in complete linkage disequilibrium. Only 0.84% of the Holstein-Friesian samples were homozygous 'AA' ARS-BFGL-NGS-109285 and these animals had on average a greater direct calving difficulty EBV (6.42 units; s.d

4.43) in comparison to the heterozygote (4.10 units; s.d 1.22) and homozygote 'CC' (3.42 units; s.d 0.42).

Due to the moderate proportion of genetic variation accounted for by the association on chromosome 18 and its translation into phenotypic difference, this association was further analysed using imputed genome sequence data within (52.58Mb to 62.58Mb), as seen in Figure 5.2. Several peaks of significance are observed across the imputed region, of which the 57.4-58.4Mb interval had the strongest SNP associations ($p < 2.5 \times 10^{-8}$; 100 strongest SNP genome-wide associations). The strongest SNP association ($p = 9.3 \times 10^{-11}$) was a variant located ~150bp from bta-mir-99b, bta-let7e and bta-mir-125a. In total 16 gene/gene-products contained SNPs with a p-value $< 2.5 \times 10^{-8}$, of which four contained significant missense variants. These were distributed among four genes; two were detected in the mRNA sialic acid binding Ig-like lectin 12-like (SIGLEC12), one in the CTU1 gene and one in the zinc finger ZNF615 (Figure 5.2). The missense variant in CTU1 was the stronger association ($p = 8.9 \times 10^{-10}$). The putative consequences of these missense polymorphisms were tested with SIFT and Polyphen. All four variants were classified as 'tolerated' with SIFT, although the missense variant in CTU1 was classified at the margin of this category (SIFT score=0.06). PolyPhen scores estimated the amino acid substitution of alanine to valine in CTU1 to be possibly damaging (PolyPhen score=0.95). The remaining missense variants were classified as 'benign' substitutions by PolyPhen.

Further analysis through imputation of full sequence in genomic intervals around associated markers on chromosomes 6 (67.02Mb to 77.02Mb) and 10 (96.72Mb to 104.30Mb) for direct calving difficulty was also undertaken (Figure 5.3). In total, 548 SNPs on chromosome 6 had a p-value < 0.0001 with a peak detected at 72.02Mb ($p\text{-value} = 9.91 \times 10^{-8}$); all but 25 of these SNPs were intergenic variants. No significant missense variants were detected within this region, although a non-coding exonic variant in

the novel gene ENSBTAG00000004082 was strongly associated with direct calving difficulty ($p=7.02 \times 10^{-6}$). Other possible candidate genes detected included the CLOCK gene, which functions as a maternal mRNA, regulating events in the oocytes and pre-implantation embryos, and CORIN a serine protease. Similarly on chromosome 10, no significant missense variants were detected and only 139 SNPs had a p-value <0.0001 , all of which were intergenic variants.

Both the SSR and Bayesian methods detected SNP associations for direct calving difficulty on chromosome 2 in the Charolais sample in the HD density panel (Table 5.1; Figure 5.1B). The strongest association for direct calving difficulty in both methods was detected within the 5.73Mb region (Bonferroni corrected $p=8.79 \times 10^{-6}$; $q=4.65 \times 10^{-6}$; $BF=3310.78$). In total, eight SNPs from the 5.51-5.73Mb region on chromosome 2 remained significant after Bonferroni correction ($p=0.004$ to 9.97×10^{-6}). The strongest Bayesian association was included in the model 70.17% of the time and accounted for and estimated 3.13% of the genetic variation in direct calving difficulty in the Charolais population. Only 0.9% of the Charolais sample were homozygous 'AA' for this SNP and a greater mean calving difficulty EBV was observed for these animals (9.5 units; s.d. 1.85) in comparison to the heterozygote (8.55 units; s.d. 0.19) and homozygous 'GG' (7.26 units; s.d. 0.09). This SNP was located within an intron of the transmembrane protein 194b (TMEM194B) gene with three other possible candidate genes (MFSD6, INPP1, HIBCH) in close proximity. In the Limousin breed, a strong association with direct calving difficulty was also detected on chromosome 2 (Figure 5.1C) in a similar location to the peak detected in the Charolais breed in the SSR analysis. A total of 49 SNPs from the 5.89 to 6.68 Mb region on chromosome 2 remained significant after correction for multiple testing; the Bonferroni p-values of these SNPs ranged from 7.39×10^{-4} to 0.05 ($q=2.67 \times 10^{-5}$ to 3.41×10^{-3}). Several growth factors and the myostatin gene were identified within this region. Additional Bayesian associations were also identified on chromosome 3 (Table

5.1) within an intron in DCAF6 with a G-protein coupled receptor (GPR161) was also located within 80kb.

5.5.2 Meta-analyses for direct calving difficulty

The strongest association for direct calving difficulty in the meta-analysis of all three breeds was a 1Mb region from 5.6Mb to 6.6 Mb on chromosome 2 (adjusted- $p < 1.78 \times 10^{-6}$) (Figure 5.1C). A total of 11 genes (TMEM194B, INPP1, MFDS6, C2H2orf88, HIBCH, MSTN, PMS1, ORMDL1, OSGEPL1, ANKAR and ASNSD1) were identified within this 1Mb region. Single SNP regression of imputed sequence data in this region of the Holstein-Friesian genotypes showed no clear association for direct calving difficulty (Figure 5.3C), suggesting this association on chromosome 2 segregates solely in the beef breeds and the meta-analysis may have been influenced by the strong associations in the beef breeds.

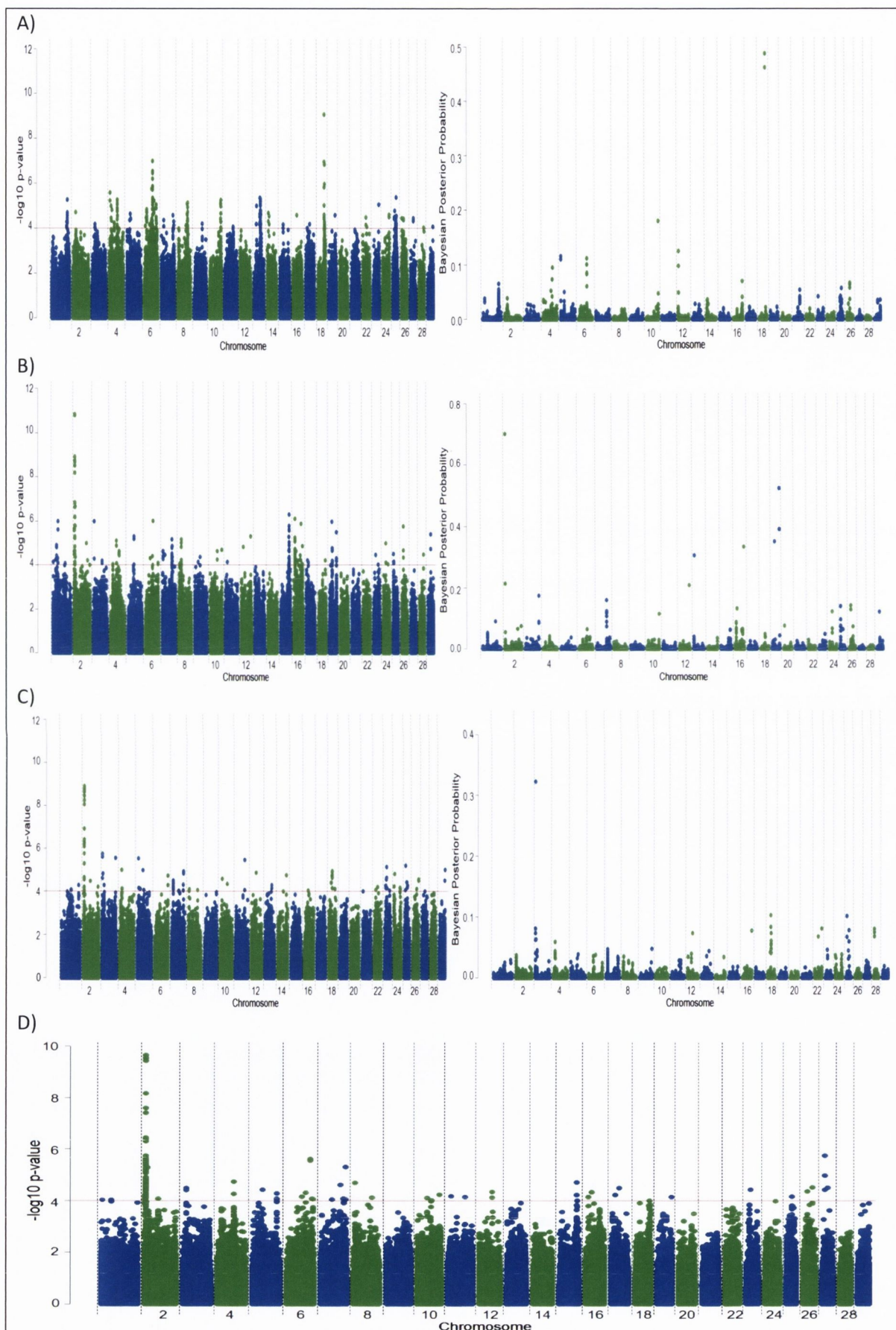


Figure 5.1 Manhattan plots for single SNP regression analysis and Bayesian analysis for A) Holstein-Friesians B) Charolais C) Limousin and D) the meta-analysis for direct calving difficulty.

Table 5.1 Bayes Factors (BF) and single SNP regression p values for all SNPs with a BF >200 for direct and maternal calving difficulty and a BF >40 for perinatal mortality.

	SNP index	SNP name	Chr	Position	MAF	BF	P value
Direct Calving Difficulty							
Holstein-	4011	ARS-BFGL-NGS-109285	18	57589121	0.08	1684.26	8.95×10^{-10}
Friesian	525264	BovineHD1800016761	18	57548213	0.09	1523.56	8.95×10^{-10}
	741177	BovineHD4100008420	10	101721285	0.41	390.31	1.03×10^{-5}
	378664	BovineHD1200001026	12	3213665	0.25	253.84	3.37×10^{-4}
	168233	BovineHD0500000940	5	3768126	0.47	232.98	1.45×10^{-4}
	217768	BovineHD0600020010	6	72022013	0.38	223.73	9.91×10^{-8}
	168231	BovineHD0500000938	5	3764573	0.47	219.46	1.24×10^{-3}
Charolais	67815	BovineHD0200001668	2	5725051	0.13	3310.78	1.59×10^{-11}
	544037	BovineHD1900018321	19	63376818	0.47	1561.23	3.16×10^{-6}
	544036	BovineHD1900018320	19	63376120	0.47	907.81	3.16×10^{-6}
	533956	BovineHD1900007134	19	24561933	0.22	762.86	1.06×10^{-6}
	487931	BovineHD1600022505	16	77709903	0.39	709.33	1.38×10^{-5}
	761092	BTB-01124458	13	6683656	0.39	620.87	1.21×10^{-4}
	67816	BovineHD0200001669	2	5731378	0.13	383.88	1.46×10^{-11}
	398587	BovineHD1200024296	12	83902896	0.14	370.53	4.96×10^{-6}
	133777	BovineHD0300034147	3	117328152	0.39	295.66	5.01×10^{-4}
	253187	BovineHD0700024774	7	84451146	0.27	267.28	1.63×10^{-5}
	645696	BovineHD2600004234	26	16849292	0.36	234.46	1.69×10^{-6}
	631615	BovineHD2500001020	25	4516912	0.17	229.69	8.5×10^{-5}
	471782	BovineHD1600005037	16	18561778	0.37	213.47	5.57×10^{-5}
	645695	BovineHD2600004233	26	16848801	0.36	209.57	1.69×10^{-6}
Limousin	103101	BovineHD0300000138	3	657019	0.47	633.42	1.80×10^{-6}
Maternal calving difficulty							
Holstein-	544242	BovineHD1900018551	19	1942110	0.36	474.13	8.63×10^{-5}
Friesian	416314	BovineHD1300016191	13	56557875	0.14	220.36	4.18×10^{-11}
Charolais	126362	BovineHD0300026067	3	90597805	0.45	769.66	3.85×10^{-6}
	126363	BovineHD0300026068	3	90600368	0.47	700.57	4.85×10^{-6}
	126361	BovineHD0300026066	3	90596779	0.45	438.06	7.18×10^{-6}
	88704	BovineHD0200024474	2	85841490	0.27	306.54	9.47×10^{-7}
	88708	BovineHD0200024478	2	85872885	0.27	260.51	1.38×10^{-6}
	88688	BovineHD0200024458	2	85764305	0.27	241.09	9.47×10^{-7}
	252315	BovineHD0700023816	7	81801829	0.38	215.81	1.09×10^{-5}
	770652	Hapmap42056-BTA-108680	3	91136875	0.38	206.24	1.23×10^{-4}
Direct Perinatal Mortality							
Holstein-	644942	BovineHD2600003337	26	12797274	0.32	40.25	8.90×10^{-7}
Friesian							
Charolais	198553	BovineHD0500035019	5	119896486	0.36	52.13	2.83×10^{-6}
Limousin	750289	BovineHD4100018377	27	37032811	0.24	127.36	6.62×10^{-8}
	664893	BovineHD2700010585	27	37023628	0.23	92.22	1.33×10^{-7}
	664891	BovineHD2700010583	27	37021710	0.23	88.02	1.33×10^{-7}
	664898	BovineHD2700010590	27	37031485	0.23	87.06	1.33×10^{-7}
	664897	BovineHD2700010589	27	37030084	0.23	84.49	1.33×10^{-7}
	664890	BovineHD2700010582	27	37019464	0.23	78.42	1.33×10^{-6}
	664895	BovineHD2700010587	27	37027959	0.23	77.78	1.33×10^{-7}
	664894	BovineHD2700010586	27	37026496	0.24	75.56	1.33×10^{-7}
	664899	BovineHD2700010591	27	37034231	0.23	72.55	1.33×10^{-7}
	664885	BovineHD2700010577	27	37007637	0.24	69.40	7.69×10^{-8}
	664888	BovineHD2700010580	27	37015031	0.24	67.05	7.69×10^{-8}

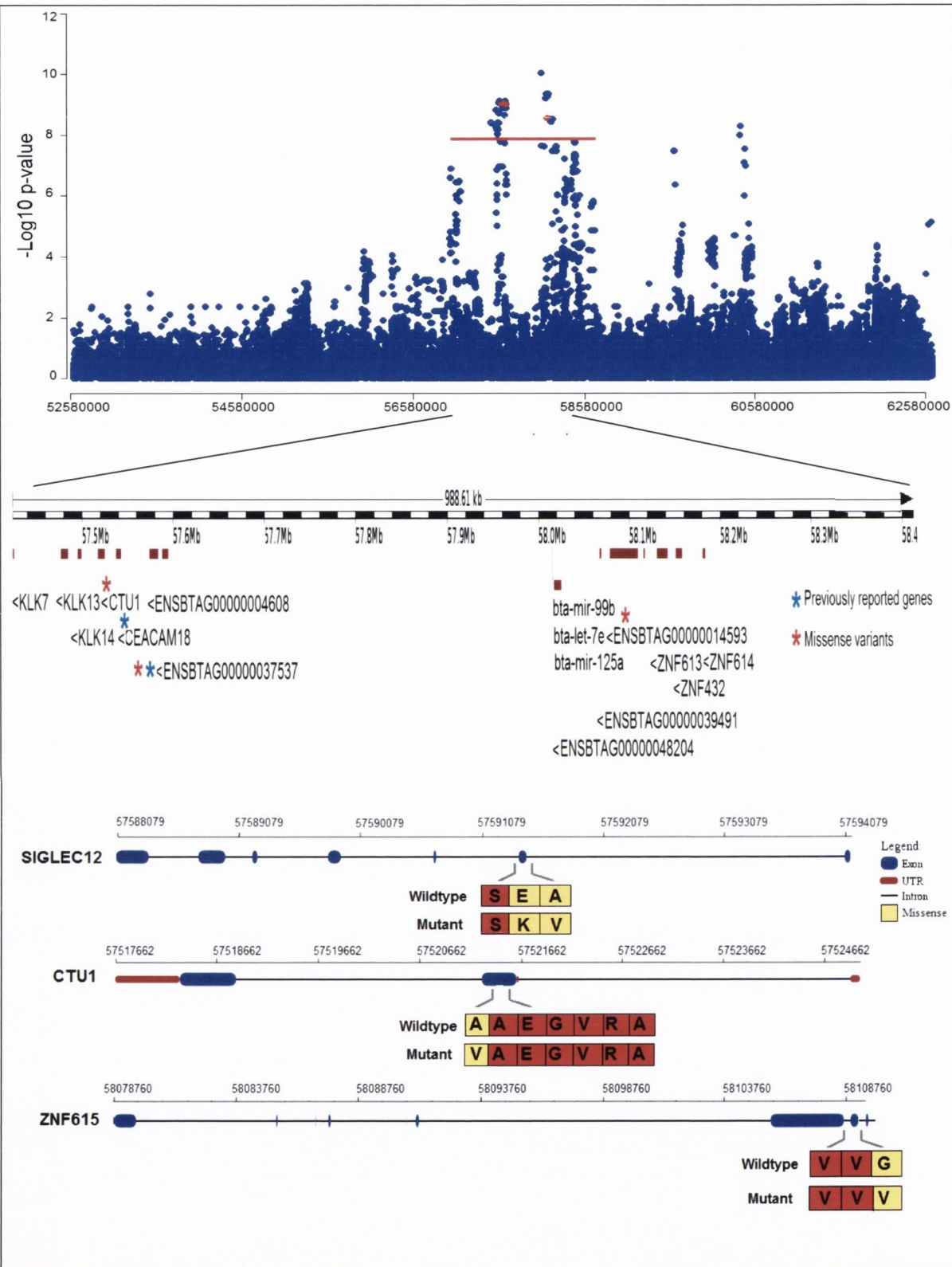


Figure 5.2 Imputation to whole genome sequence single SNP regression results of 10Mb region surrounding ARS-BFGL-NGS-109285 on chromosome 18. Genes annotated are where those identified within the top 100 significant SNPs of which 4 were missense variants. Two of these missense variants (coloured in red in the top plot) were located in *SIGLEC12* (*ENSBTAG000000037537*) and 1 in *CTU1* and *ZNF615* (*ENSBTAG000000014593*).

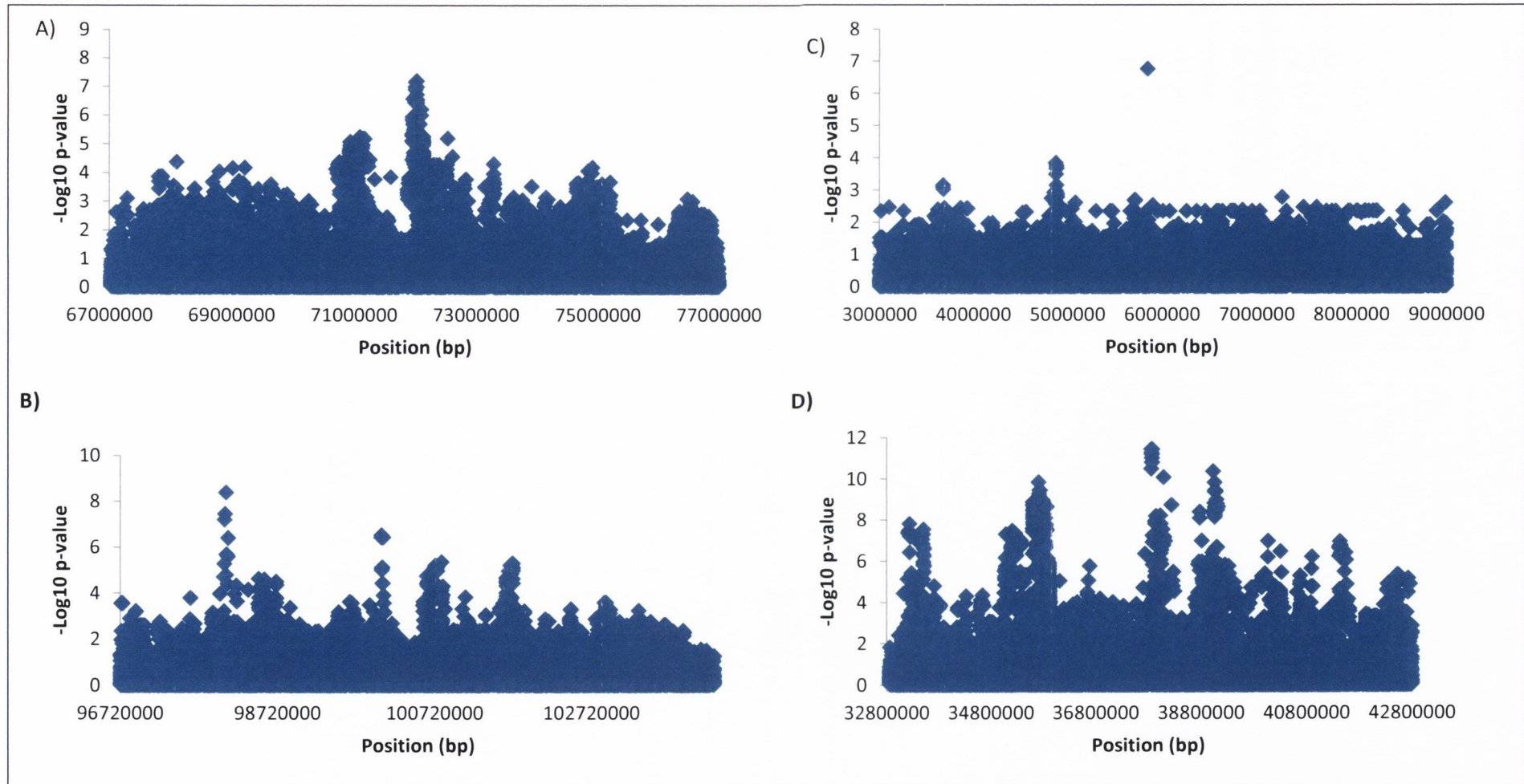


Figure 5.3. Imputed to sequence density single SNP regression (SSR) results for direct calving difficulty in the Holstein-Friesian population. A) Chromosome 6 (10Mb region surrounding HD peak at 72Mb) B) chromosome 10 (8 Mb region surrounding HD peak at 101Mb) C) chromosome 2 (6Mb region surrounding the HD peak at 5.6 to 6.6Mb found in the Charolais and Limousin population) and D) SSR results for maternal calving difficulty on chromosome 4 (10Mb region surrounding peak at 37Mb)

5.5.3 Maternal calving difficulty

Maternal calving difficulty describes the characteristics of the dam giving birth (e.g., pelvic dimensions) and its impact on parturition. Maternal calving difficulty was influenced by many polymorphisms, each of small effect, as the maximum proportion of the genetic variation accounted for by any one SNP was 0.13% (BovineHD1900018551 on chromosome 19 in the Holstein-Friesian population). Several genomic regions associated with maternal calving difficulty were primarily identified on chromosomes 1, 4, 11, 13, and 19 (Figure 5.4A). The same genomic region on chromosome 13 (56.55Mb) was significantly associated with maternal calving difficulty in both association approaches (Bonferroni corrected $p=2.61 \times 10^{-5}$; $q=2.53 \times 10^{-6}$; $BF=220.36$); no candidate gene, however, in the vicinity could be identified. The four strongest associations in the SSR analyses were detected within a 20kb area surrounding location 37.85Mb on chromosome 4 (Bonferroni corrected $p=2.25 \times 10^{-6}$ to 5.99×10^{-6}). Upon using imputed genome sequence data within the interval 32.85Mb to 42.85Mb, the same region at 37.85Mb identified in the HD GWAS continued to exhibit the strongest association (Additional file 2). The strongest SNP association ($p=3.71 \times 10^{-12}$) was located in an intron of the PCLO gene. Additional candidate genes identified within this 10Mb region that contained intronic SNPs with p -values $< 2.51 \times 10^{-7}$ include SEMA3D, HGF, a novel gene ENSBTAG0000019276, CROT and the oestrogen induced gene KIAA1324-like. Two missense variants located in the PCLO gene and one in KIAA124-like had a p -value < 0.005 . SIFT scores were not available on the missense variants in the PCLO gene whereas the variant was classified as tolerated in KIAA125-like (SIFT=0.38).

Associations with maternal calving difficulty in the Charolais breed were detected using both the SSR and the Bayesian analysis on chromosomes 2, 3, and 7 (Figure 5.4B, Table 5.1). However, upon adjustment for multiple testing in the SSR analysis no association

remained significant. Similarly in the Limousin sample, no associations remained significant after multiple testing adjustment, although strong SSR associations could be detected on chromosome 5. A total of 3 SNPs on chromosome 3 in the Charolais sample, surrounding the 90.59Mb region, had a $BF > 200$ and cumulatively were included in the Bayesian model 36.43% of the time, accounting for 0.42% of the genetic variation in maternal calving difficulty. No candidate near these loci on chromosome 3 could be identified.

5.5.4 Meta-analysis of maternal calving difficulty

The strongest association in the meta-analysis of maternal calving difficulty was detected on chromosome 13 (adjusted- $p = 1.09 \times 10^{-5}$) (Figure 5.4D), located close ($< 2\text{kb}$) to signal-regulatory protein alpha (SIRPA) and prodynorphin, a hormone involved with signal transduction and cell communication was within 30kb upstream. No significant associations were detected after adjustment for multiple testing.

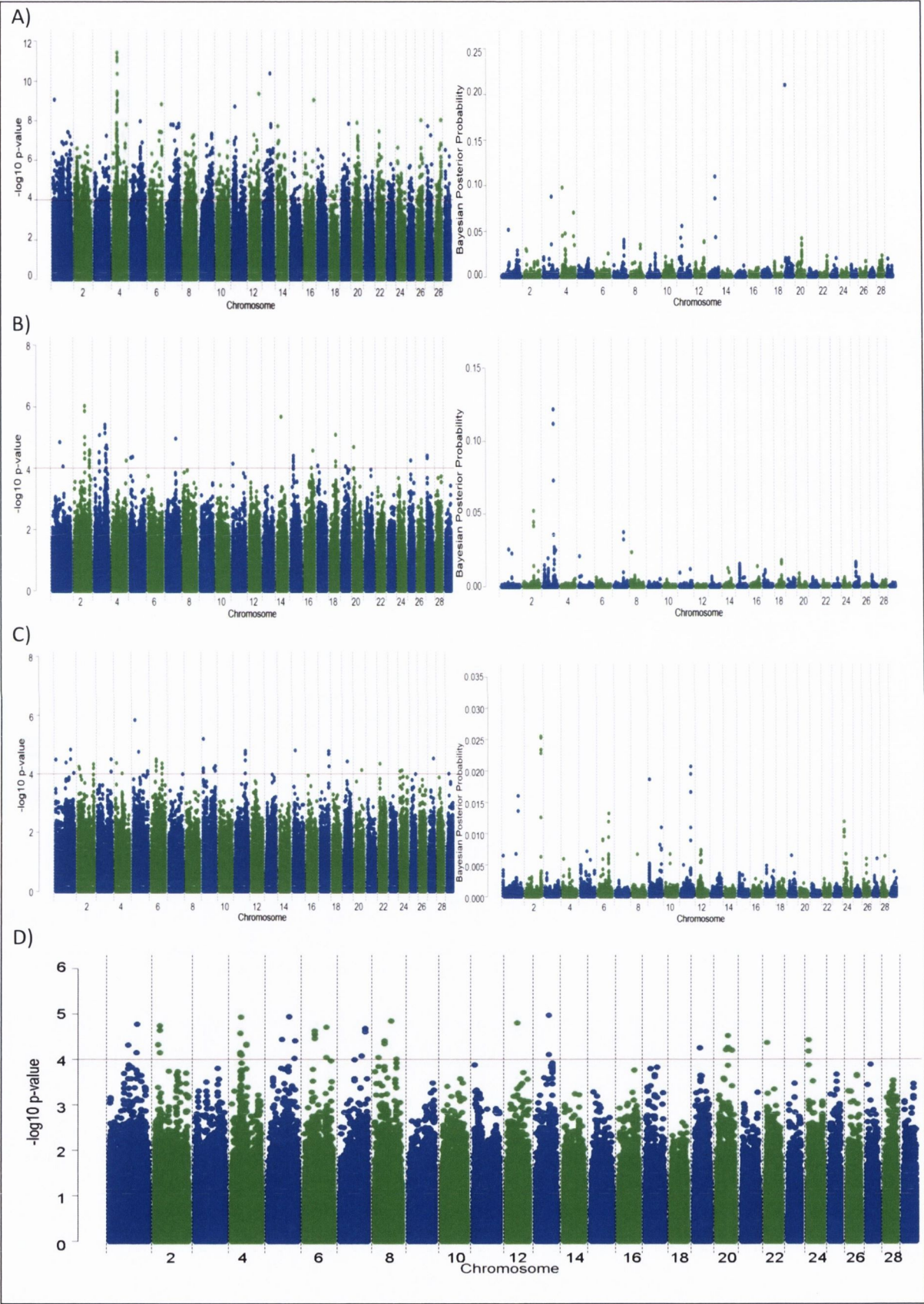


Figure 5.4 Manhattan plots for single SNP regression analysis and Bayesian analysis for A) Holstein-Friesians B) Charolais, C) Limousin and D) the meta-analysis result for maternal calving difficulty.

5.5.5 Perinatal Mortality

Perinatal mortality in cattle is defined as calf mortality shortly before, during, or after parturition with a 24hr period. As no maternal estimate is currently available this study is based on an estimate of the calves direct breeding value for mortality. In both the SSR and Bayesian analysis, chromosomes 4 and 26 had the strongest SNP associations in the Holstein-Friesian sample. (Table 5.1, Figure 5.5A). In total, 9 SNPs had a $BF > 30.1$, suggesting strong evidence of an association and accounted for 0.13% of the genetic variation in perinatal mortality in the Holstein-Friesian population. No analysis with imputed genome sequence data was undertaken as no association remained significant after adjustment for multiple testing in the Holstein-Friesian sample. In the Charolais breed, the strongest SSR and Bayesian association was on chromosome 5 (unadjusted $p = 2.83 \times 10^{-6}$; $BF = 52.13$). Associations were also detected on chromosomes 9 and 18 in the SSR analysis, although after adjustment for multiple testing no association remained significant. A total of six potential candidate genes (HDAC10, MAPK12, MAPK11, PLXNB2, FAM116B, PPP6R2) surrounding the association on chromosome 5 were identified.

Alternatively in the Limousin sample, several strong SNP associations for perinatal mortality were detected on chromosome 27 in both the SSR and Bayesian analysis. A total of 11 SNPs had a $BF > 60$ (Table 5.1) and all were located within the same LD block with D' values > 0.91 for each pairwise combination of two SNPs. Cumulatively all 11 SNPs were included in the model 63.25% of the time and accounted for 0.60% of the genetic variation in perinatal mortality in the Limousin sample. The same SNPs were also highly significant in the SSR analysis, all displaying a corresponding q-value of significance (< 0.05) with three SNPs (BovineHD4100018377, BovineHD2700010577 and BovineHD2700010580) remaining significant ($p < 0.05$) after Bonferroni correction. Gene annotation showed that solute carrier family 20 (SLC20A1), a phosphate transporter was located within the LD block of the twelve SNPs.

5.5.6 Meta-analysis of perinatal mortality

The strongest association for perinatal mortality across all breeds was detected on chromosome 9 (genomic-control $p = 6.91 \times 10^{-6}$; Figure 5.5D), although no SNP associations remained significant after correction for multiple testing. The association on chromosome 9 was located within the gene uronyl-2-sulfotransferase.

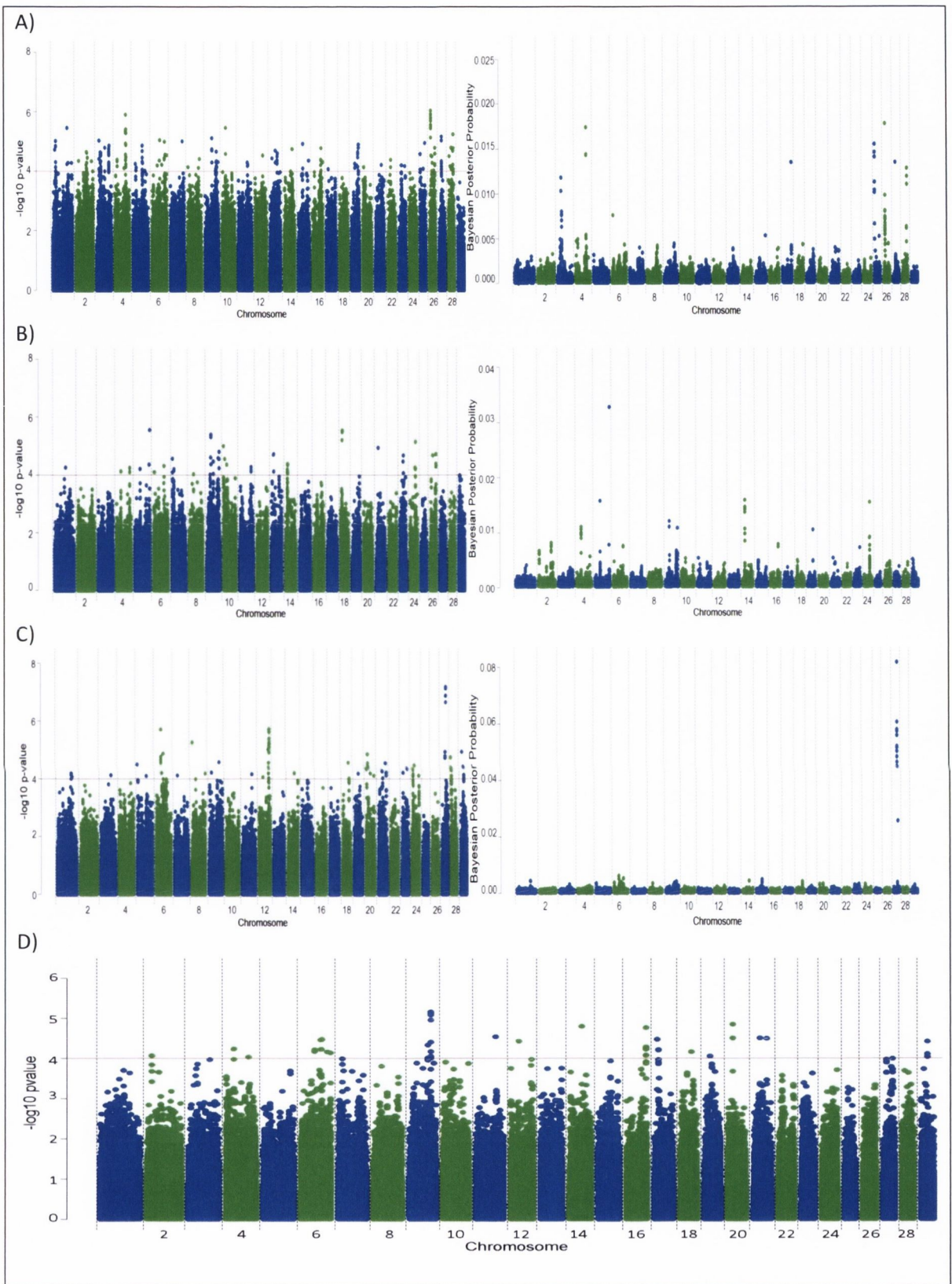


Figure 5.5 Manhattan plots for single SNP regression analysis and Bayesian analysis for A) Holstein-Friesians B) Charolais, C) Limousin and D) the meta-analysis results for direct perinatal mortality.

5.5.7 Copy number variants

A total of 101,180 CNVs were identified across all 1,762 animals with the largest proportion detected in the Holstein-Friesian breed (76,621 CNVs). The majority of these structural variants were between 200 and 500kb in length, although a large percentage (10.59%) were greater than 1Mb in length in the Holstein-Friesian population (Figure 5.6). Single copy deletions were the most common structural variant in the Charolais and Limousin populations whereas double copy duplications were the most prevalent in the Holstein-Friesian breed (Figure 5.6). Chromosomes 7 and 19 had the largest proportion of CNVs in the beef and dairy breeds, respectively. The genome-wide mean coverage of CNVs was 5.95Mb in the Charolais, 5.70Mb in the Limousin and 111.71Mb in the Holstein-Friesians.

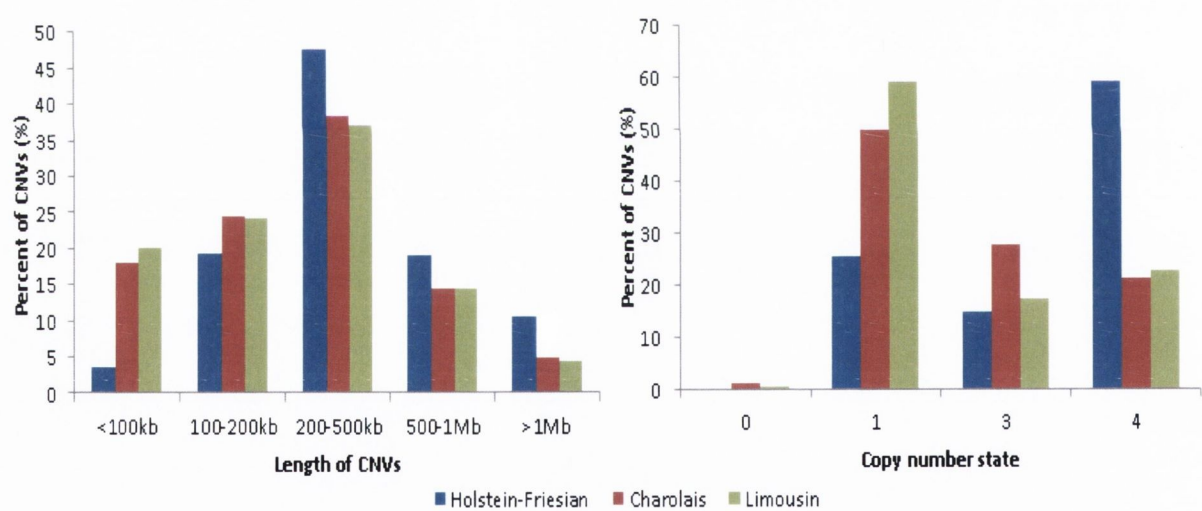


Figure 5.6 Size length distribution of copy number variants (CNV) and percentage of each copy number state identified in each breed. Copy number variation is represented by four states where 0 is double copy deletion, 1 is single copy deletion, 3 is single copy deletion and 4 is double copy duplication.

CNV association analyses showed several significant associations between direct calving difficulty and these structural variants located on chromosomes 11, 16, and 1. Overlap did exist between the significant associations detected in the HD GWAS and CNVs, as the

genomic region on chromosome 18 containing the mRNA SIGLEC12 was structurally dense with CNVs in the Holstein-Friesian population. A total of 77.33% of the Holstein-Friesian population had a double copy duplication surrounding this region in comparison to 1.44% of the Charolais population and 2.21% of the Limousin population. This double copy duplication was significantly associated with direct calving difficulty ($p=0.035$). In the beef breeds, the genomic region on chromosome 2 from 5.6Mb to 6.6Mb that was significantly associated with direct calving difficulty had relatively minimal structural variants with only two animals detected with CNV deletions within this region. No significant associations between maternal calving difficulty and CNVs were detected in any breed, with the same reported for direct perinatal mortality. Analyses of the regions surrounding the significant SNP associations for both maternal calving difficulty and perinatal mortality detected in the HD GWAS revealed very little structural variation across all breeds.

5.5.8 Runs of homozygosity and calving performance

A total of 250,347 ROH were identified across all three populations (84,719 in the Charolais, 97,006 in the Limousin and 68,622 in the Holstein-Friesians). The average genome ROH coverage was 163.02Mb. Significant associations detected for calving performance in the HD GWAS were analysed to detect for ROH. Genomic regions of significant association with direct calving difficulty on chromosomes 18 and 2 were not particularly homozygous as only a small proportion of the population exhibited homozygosity in these regions (1.18% and 8.02%). Only the genomic association for maternal calving difficulty on chromosome 19 in the Holstein-Friesian population was particularly homozygous with a ROH involving this region detected in 16.32% of animals. The genomic associations for direct perinatal mortality were more likely to be found in

ROH in the Holstein-Friesian population with the HD GWAS associations on chromosomes 26, 4, 25 and 17 appearing in a ROH in 5-6% of the population. There was no association between a longer ROH and a greater calving difficulty or perinatal mortality EBV.

5.6 Discussion

This study identified a large number of SNPs that were strongly associated with calving performance traits in dairy and beef cattle and across multiple breeds. The majority of associations differed per breed, implying that a breed effect difference for calving performance exists. Many associations of small effect were also detected for all traits suggesting the infinitesimal model has some validity for calving performance. Nevertheless, strong genomic associations on chromosome 18 and 2 were detected for direct calving difficulty suggesting influential mutations for calving performance do exist which can be possibly selected upon to improve genetic performance.

5.6.1 Direct calving difficulty

The SNP ARS-BFGL-NGS-109285 located on chromosome 18 which was the strongest SNP association in both the SSR and Bayesian analyses in the Holstein-Friesian sample, has been previously documented to be associated with direct calving difficulty in three separate dairy populations genotyped at a lower density (Cole et al., 2009, Purfield et al., 2013, Sahana et al., 2011). This SNP and the adjacent SNP BovineHD18001676 accounted for a large proportion of the genetic variation (2.49%) in direct calving difficulty within the Holstein-Friesian sample in the present study, and suggested an underlying causal mutation was located within this region. Genome sequence data available through the 1000 Bulls Genome Project was used to impute data within this interval and further refine genomic

associations. Several possible candidate genes were identified, of which two have been previously associated with direct calving difficulty; sialic acid binding Ig-like lectin 12-like (SIGLEC12, also known as SIGLEC5) (Cole et al., 2009) and carcinoembryonic antigen-related cell adhesion molecule 18-like (CEACAM18) (Mao et al., 2013). The identification of highly significant missense variants in three genes (SIGLEC12, CTU1 and ZNF615) within this region suggests some possibility of detection of causal mutations. Their impacts on protein function were estimated and, although caution needs to be taken when interpreting their results due to low specificity of Polyphen and SIFT (Flanagan et al., 2010), they do suggest that there may be a mutation in CTU1 impacting direct calving difficulty in the Holstein-Friesian sample. The relevance of CTU1 in calving performance is unknown but it has been suggested that deletion of CTU1 may impact proteins with codons enriched in AAA, GAA and CAA (Dewez et al., 2008). Therefore it may be possible the alteration of this CTU1 gene may impact regulatory proteins or genes of direct calving difficulty. This 1Mb genomic region was enriched for double copy duplications in a large proportion of Holstein-Friesian sample (77.33%) strengthening the suggestion that it is of beneficial importance. Furthermore, the association of intergenic variants on chromosomes 6 and 10 with direct calving difficulty suggests unidentified genes such as non-coding RNAs, promoters and enhancers may be impacting calving performance. Further work is needed to determine the exact causal mutations for direct calving difficulty in the Holstein-Friesian population; however the availability of genome sequence data has allowed several possible candidate genes and regions to be identified in this study.

A strong association located on chromosome 2 (5.6-6.6 Mb) in both the Charolais and the Limousin samples, suggests there is a potential QTL influencing direct calving difficulty in beef breeds within this region. A total of 11 possible candidate genes were identified within this 1Mb region including the well-known double muscling gene, myostatin. Myostatin has long since been associated with calving difficulty with homozygous animals

having an increase of 0.19 (s.e. ± 0.6) in the proportion of calving difficulty (Casas et al., 1999). Additionally, selective sweeps have been previously reported for this QTL in the Blonde d'Aquitaine beef breed which may be symptomatic of the strong recent selection in this breed for growth and development (Boitard and Rocha, 2013). However, the analysis of imputed genome sequence data in the Holstein-Friesian population in the present study suggests this QTL in the 5.6-6.6 Mb region is not associated with direct calving difficulty in this breed group. This suggests that this genomic region is of importance for direct calving difficulty within the beef breeds, where selection has been primarily based on growth and development.

Overall, the identification of polymorphisms that accounted for a large proportion of the genomic variation in direct calving difficulty across all breeds suggests that genomic selection algorithms should facilitate the modelling of alleles with small as well as large effects. The current genetic architecture of this trait suggests Bayesian selection methods may be more suited to the genomic architecture of direct calving difficulty as they facilitate the modelling of heterogeneity in SNP contributions to the genetic variance (Meuwissen et al., 2001).

5.6.2 Maternal calving difficulty

Despite the reduced study population size for maternal calving difficulty the results suggest that maternal calving difficulty is likely to be influenced by many polymorphisms each of small effect since the maximum proportion of the genetic variation accounted for by any one SNP was 0.13% (BovineHD1900018551 in the Holstein-Friesian sample). The accuracy of predicting genetic merit has been proven to be greater in traits affected by large SNP effects, than for polygenic traits (Hayes et al., 2010). This suggests that the

accuracy of genomic predictions for maternal calving difficulty may be impaired due to its underlying genomic architecture.

A clear association with maternal calving difficulty on chromosome 4 in the Holstein-Friesian sample was detected, but in the beef breeds there was no significant regions after adjustment for multiple testing; this is possibly due to the fewer number of animals for this trait (221 Charolais and 357 Limousin). However, the unadjusted significant SSR associations in both the Charolais and Limousin breeds were positioned near or located at the same position as previously reported QTL associated with maternal calving ease, birthweight and calf size on chromosomes 3 and 2, respectively (Grosz and MacNeil, 2001, Purfield et al., 2013, Sahana et al., 2011, Schrooten et al., 2000). This suggests plausible genomic regions impacting maternal calving difficulty in both the Charolais and Limousin samples could still be identified.

The strongest SSR association in the Holstein-Friesian samples in both the HD and imputed whole genome sequence data on chromosome 4, is located 4.3Mb from a QTL previously associated with maternal calving difficulty in Danish and Swedish Holstein cattle (Sahana et al., 2011). Although no missense variants detected were estimated to impact protein structure, strong associations for intronic variants were detected in PCLO suggesting this gene or the surrounding identified genes impact maternal calving difficulty. PCLO has been shown to be involved in developmental malformations in humans (Fenster and Garner, 2002). Other morphological attributes, such as body depth, bone percentage, and birth-weight have been reported to be influenced by this region as well (McClure et al., 2010) suggesting that this region may be structurally important in cattle.

Furthermore, the association of prodynorphin with maternal calving difficulty in all breeds may confirm the perception that calving difficulty is one of the most painful conditions a cow can experience, as it is linked to behaviour, pain perception, and psychological processes, with its expression or up-regulation believed to serve as a biological indicator of

stress in cattle (Huxley and Whay, 2006, Jiang et al., 1997). Association of a gene regulating such a hormone with maternal calving difficulty suggests this trait should remain a welfare concern

5.6.3 Perinatal Mortality

The identification of multiple weak genomic associations for direct perinatal mortality suggests the infinitesimal model hypothesis may indeed hold for this trait. Therefore, it is suggested all loci should be studied collectively when analysing perinatal mortality, and indeed direct and maternal calving difficulty, as they are influenced by the cumulative effect of thousands of loci. The genes identified to be associated with perinatal mortality in the present study (SLC20A1 and UST), have primarily been shown to contribute to morphological abnormalities (Autry and Monteggia, 2012, Beck et al., 2010, Grobe et al., 2005) suggesting structural deformities contribute to the genetic variation in perinatal mortality. However, the proportion of genetic variance accounted for by these genes was minute.

The identification of several breed specific associations of moderate effect for calving performance due to either different breed allele effects or the linkage phase between the genotyped allele and the causal mutation varying per breed, suggests there may be implications for across-breed genomic evaluations. The substantial difference in allele directions between breeds may be the reason why across-breed genomic evaluations have to-date been elusive (Hayes et al., 2009). Additionally, such differences in allele effects suggests the genomic selection for such alleles could be deleterious for breeds not represented, or poorly represented, in reference populations if the allele effect in that population is opposite to the reference population. This is likely the contributor to the sometimes observed negative correlations between direct genomic values for a breed

which were not represented in the reference population (Hayes et al., 2009). Nevertheless, common genomic regions (of the same allele direction) across breeds that significantly associated with calving difficulty were identified possibly facilitating some element of an across-breed genomic evaluation. This suggests that a percentage of genomic variation attributable to calving performance is common across all diverse breeds. Perhaps partitioning genomic prediction algorithms into a within breed component and an across breed component may allow for these within breed effects to be captured.

5.7 Conclusion

Several strong associations were detected in three breeds for all three calving performance traits, although breed-specific SNP effects were found to exist. All traits were found to be polygenic as a large number of SNPs with small effects were identified. The availability of whole genome sequence data enabled imputation in the Holstein-Friesian population allowed defined genomic associations to be refined thus increasing the accuracy of identifying possible candidate genes. Several previous studies have suggested the presence of a causal mutation influencing direct calving difficulty on chromosome 18 in the dairy population, and although no definite mutation was found, a missense variant in CTU1 may be a contributor to the genetic variance in this trait. In addition, despite the strongest associations for calving performance differing per breed, common genomic associations in all three breeds were identified. This suggests that a proportion of the genomic variation attributable to calving performance is common across breeds; which may have implications for across-breed genomic evaluations.

Chapter 6

Discussion

6.1 Summary of thesis

The overall objective of this thesis was to quantify the contribution of genetics (both direct and maternal genetic effects) to phenotypic differences in calving difficulty and perinatal mortality and to elucidate the genomic architecture governing this genetic variation. Both calving difficulty and perinatal mortality are of substantial economic importance to the Irish dairy and beef cattle sector. Calving dystocia is routinely recorded at birth by producers on a scale of one to four; there is a legal requirement in Ireland to record the death of all animals (including stillbirths) on farm.

Because of the complex characteristics and the multifactorial contributors to these traits, limited studies have been performed on dystocia and perinatal mortality; therefore the genetic and genomic variation contributing to calving performance phenotypes has not been thoroughly defined. To determine the potential of breeding for improved calving performance, genetic parameters must be known. The present thesis used calving performance phenotypes from the national database to decompose the observed phenotypic differences among animals in calving performance into its causal components, after adjustment for the environmental risk factors.

Many national dairy cattle genetic evaluations, including Ireland, are supplementing their traditional quantitative genetic evaluations (i.e. BLUP) with genomic data (Wickham et al., 2011; Spelman et al., 2013). Therefore, more accurate genetic evaluations could possibly be achieved by more accurately accounting for the genomic architecture of the trait. This study used high reliability estimated breeding values for calving performance to identify regions of the genome associated with genetic differences in these traits. Also, the contribution of structural variation such as runs of homozygosity and copy number variations to calving performance was assessed.

Results from this study show that genetic variation clearly still exists for both calving difficulty and perinatal mortality and therefore genetic gain is indeed possible. Several genomic regions were significantly associated with genetic differences in calving difficulty and perinatal mortality, although the strongest associations differed by breed. Moreover, there were several genomic regions with large effects but many genomic regions with little or no effects suggesting the assumptions of the infinitesimal model may be true. Furthermore, the sign of association for several alleles differed between breeds suggesting that across-breed genomic selection, using the commercially available genotyping panels, may be difficult.

The following are summaries in bullet point form, outlining the objective, primary methods, results and conclusions of each chapter, followed by the overall conclusions of this thesis.

6.1.1 Chapter 2: Genetic parameters of calving performance traits

Objective: To estimate genetic parameters for direct and maternal dystocia and perinatal mortality with particular emphasis on perinatal mortality which was not associated with a dystocia (PMND) event in Irish beef and dairy cows.

- Edited dataset consisted of 100,664 birth events from dairy and beef herds (33,526 primiparous animals and 67,138 multiparous animals).
- (Co)Variance components were estimated using univariate repeatability animal-dam linear mixed models.
- Prevalence of all traits was greatest in primiparous animals (severe dystocia 4.29%, perinatal mortality 4.15%, and perinatal mortality not associated with dystocia 1.57%).
- Direct heritability estimates for dystocia were 0.207 (± 0.026) and 0.104 (± 0.012) in primiparous and multiparous dams, respectively; the respective maternal heritability estimates were 0.034 (± 0.013) and 0.008 (± 0.004). The genetic correlation between direct and maternal dystocia was -0.34 (s.e. 0.138) and -0.68 (s.e 0.079), for primiparous and multiparous cows, respectively.
- Direct heritability estimates for perinatal mortality were low at 0.007 (± 0.008) and 0.001 (± 0.002) in primiparous and multiparous dams, respectively. The maternal heritability estimates were similarly low at 0.017 (± 0.007) and 0.001 (± 0.001).
- Perinatal mortality not associated with dystocia was only heritable for the direct component in primiparous animals (0.019 ± 0.009).
- In conclusion, the prevalence of dystocia and perinatal mortality in Ireland is considerable. Genetic variation in calving performance exists suggesting that exploitation through breeding programs to lower the prevalence even further is possible.

6.1.2 Chapter 3: Runs of homozygosity and population history in cattle

Objective: To characterise runs of homozygosity (ROH) in a variety of breeds with high density (i.e. 777,962 SNPs) genotypes and validate the efficacy of detecting ROH using a lower density (i.e., 54,001 SNPs) genotype panel.

- Illumina HD genotypes were available on 891 sires of multiple breeds. A reduced HD panel of 48,734 SNPs common to both the HD and Illumina BovineSNP50 Beadchip was used to evaluate the lesser *circa* 50k SNP density for identifying ROH. BovineSNP50 genotypes were also available on 1166 animals from 42 breeds.
- The measure of homozygosity per animal was calculated as the sum of ROH per animal above a pre-defined length divided by the total length of autosome covered by SNPs. Pedigree based inbreeding coefficients for all animals were also calculated.
- ROH were frequent across all breeds but between animal variation in patterns of ROH length and quantity illustrate variations in breed origins and recent management.
- All animals genotyped on the HD panel displayed at least one ROH that was between 1–5 Mb in length with certain regions of the genome more likely to be involved in a ROH than others.
- A strong correlation ($r = 0.75$, $p < 0.0001$) existed between the pedigree-based inbreeding coefficient and a statistic based on sum of ROH of length > 0.5 KB
- In conclusion, ROH may be used to infer aspects of recent population history even from relatively few samples. ROH analysis in cattle provides a sufficient predictor of the pedigree inbreeding coefficient and the commonly used lower density (i.e., *circa* 50,000 SNPs) genotyping array may be a sufficient tool to predict ROH.

6.1.3 Chapter 4: Genome-wide association study for calving traits in Holstein-Friesian dairy cattle.

Objective: To identify regions of the bovine genome associated with dystocia and perinatal mortality in Holstein–Friesian dairy cattle using the Illumina BovineSNP50 Beadchip.

- Predicted transmitted abilities (PTAs) for calving performance traits with an adjusted reliability >40% and Bovine SNP50 genotypes (43,204 SNPs after edits) were available on
 - 1,970 male animals for direct calving dystocia,
 - 1,555 for maternal calving difficulty and
 - 740 animals for direct perinatal mortality.
- Association analyses were undertaken by fitting all SNP simultaneously in a Bayesian framework.
- Several SNPs on chromosomes 5, 6, 11, 12, 17, 18 and 28 were detected to be strongly associated (Bayes factors >30.1) with calving performance traits.
- One genomic region on chromosome 18 showed very strong evidence of an underlying causative mutation, accounting for 2.1% of the genetic variation in direct calving difficulty.
- Several KEGG pathways were identified. The riboflavin metabolism pathway was the most over-represented pathway for direct calving difficulty, the glycosaminoglycan degradation pathway was over-represented for maternal calving difficulty, and systemic lupus erythematosus pathway for perinatal mortality.
- In conclusion, this study confirmed the presence of several QTL previously associated with calving performance traits but also identified several novel associations and QTLs that are only associated with one calving trait allowing for direct selection while overcoming the genetic antagonisms that exists between the calving traits.

6.1.4 Chapter 5: Genome-wide association study for calving performance using high density genotypes in dairy and beef cattle.

Objective: To identify regions of the genome associated with calving performance traits in multiple breeds using the Illumina High-Density (HD) beadchip (777,962 SNPs).

- Illumina HD SNP genotypes were available on 2,660 dairy and beef bulls with EBVs and associated adjusted reliabilities >30%. After edits up to 605,718 autosomal SNPs remained.
- Association analyses were undertaken using 1) single SNP regression and 2) Bayesian methods within each breed separately. A weighted meta-analysis was used to combine within breed SNP effects.
- Many significant associations were detected for each trait and although most differed by breed, several common genomic regions with significant effects were detected across breeds.
- A genomic region on chromosome 18 accounted for 2.49% of the genetic variance attributable to direct calving difficulty in the Holstein-Friesians and a region on chromosome 2 accounted for 3.13% of the genetic variance in Charolais.
- Genome sequence data was used to refine associations on chromosome 18. Several significant associations were identified of which four were missense variants.
- The missense variant in CTU1 may impact protein structure (SIFT 0.06; PolyPen 0.95). CTU1 may impact direct calving difficulty through a regulatory protein.
- The strongest association detected for direct calving difficulty in all three breeds was a 1Mb genomic region on chromosome 2, encompassing 11 genes including MSTN.
- In conclusion, common genomic variation with an association with calving performance was detected across breeds, but many regions were not common across breeds. This has implications for across-breed genomic evaluations.

6.2 Overall conclusions and implications

Calving difficulty and perinatal mortality are both prevalent in Ireland. The economic impact of these traits necessitates more in-depth research into their underlying genetic and genomic architecture. Approximately 1,085,864 million calves are born annually in dairy herds (CMMS, 2012). Assuming a prevalence of 16.08%, 2.22%, 0.91% for slight assistance, severe assistance and veterinary assistance (including caesarian) and the respective economic effects on dairy cow performance of €13, €145 and €662 (Berry et al., 2007a) this equates to a cost nationally of €32,735,748. A similar cost is expected in the national beef herd. Moreover, assuming 2,202,451 calves born nationally (CMMS, 2012) and a value per calf of €200, a perinatal mortality prevalence of 2.49% suggests a cost nationally of perinatal mortality of €10,968,205 annually.

Chapter 2 demonstrates that although heritability estimates of calving difficulty and perinatal mortality were relatively low, a reasonable amount of genetic variation existed. Because both traits are economically (Berry et al., 2007a) and socially (Mee, 2013) important, they should be included in national breeding goals. Moreover, because producers record calving dystocia on most births (61.5%) (Evans and Pabiou, 2012) and there is a legal requirement to record perinatal mortality, high reliability genetic evaluations for sires is possible. Selection for improved calving performance will reduce the incidence of these phenomena. Genetic variances, however, differed substantially between primiparous and multiparous dams in particular for calving difficulty; this phenomenon agrees with international literature (Eriksson et al., 2004, Steinbock et al., 2003) and suggests that genetic evaluations should probably account for this difference when estimating breeding values for calving difficulty. Such an approach however would need consideration of which scale to present the EBVs nationally. The negative correlation between the direct and maternal components of calving difficulty is presumably the result

of small easy calving female calves who then, at maturity, may experience difficult calvings due to their reduced size and pelvic opening dimensions (Philipsson et al., 1979). Thus, if calving performance is to be included in national breeding goals, both the direct and maternal components should be simultaneously included and ideally no unfavourable genetic trends expected. Moreover, understanding the genomic architecture influencing the direct and maternal components separately and their correlation may aid in effective simultaneous selection for both traits. This is because genetic correlations can be due to pleiotrophy or genetic linkage and can potentially be resolved during meiosis once the underlying causal mutations are known. Genomic regions that were associated with only one trait were identified in Chapter 4. Also the identification of an influential association on chromosome 18 in both the BovineSNP50 in Chapter 4 and HD density genotype panel in Chapter 5 suggests there is an influential mutation impacting direct calving difficulty in Holstein-Friesian cattle. The use of imputed whole genome sequence data allowed this association on chromosome 18 to be defined to a 1 Mb region where 16 potential candidate gene/gene products were identified. The identification of four highly significant missense variants within this region suggested further evidence of a causal mutation. However, only one of these variants, located in the gene CTU1 was suggested to impact protein structure. Further work is needed on defining this association on chromosome 18 but perhaps an in-vitro study may help explain possible role of CTU1 in direct calving difficulty. Exploitation of the association on chromosome 18, which explained 2.49% of the genetic variation attributable to direct calving difficulty, in breeding programs could be used to achieve genetic improvement in the Holstein-Friesian breed. Although, the frequency of the allele associated with less calving difficulty was high suggesting the previous selection (either artificial or natural) has favoured this allele and that little is to be gained in the direct selection of the favourable allele.

The complexity of calving difficulty was revealed in Chapter 5 in that many genomic regions were associated with this trait and most of them differed by breed. Alleles at 283,042 loci differed in their signed effects between the Holstein-Friesian and Charolais cattle, while alleles at 285,133 loci differed in the sign of their effect between Holstein-Friesian and Limousin cattle; alleles at 284,509 loci differed in the direction of their effect between Charolais and Limousin. This suggests different breed allele effects for calving performance exists or else the linkage phase between the genotyped allele and the causal mutation varies per breed. The substantial difference in allele directions between breeds may be the reason why across-breed genomic evaluations have to-date been elusive (Berry, 2012). Additionally, such differences in allele effects suggests the genomic selection for such alleles could be deleterious for breeds not represented, or poorly represented, in reference populations if the allele effect in that population is opposite to the reference population. This is likely the contributor to the sometimes observed negative correlations between direct genomic values for a breed which were not represented in the reference population (Berry, 2012). Nevertheless, common genomic regions (of the same allele direction) across breeds that were significantly associated with calving difficulty were identified possibly facilitating some element of an across-breed genomic evaluation. This suggests that a percentage of genomic variation attributable to calving performance is common across all diverse breeds. This also suggests that genomic prediction algorithms may be partitioned into a within breed component and an across breed component. Further in depth analysis of sequence data may help refine these models further.

The objectives of Chapters 4 and 5 are similarly focused on finding genomic regions associated with calving performance, and account for a proportion of overlap within this thesis. Chapter 5 was undertaken as a continuation study of Chapter 4 in an attempt to refine the genomic associations identified in this previous chapter in the Holstein-Friesian population whilst also identifying associations for calving performance in two additional

beef breeds. Despite the 13-fold increase in SNP density between the BovineSNP50 panel and HD panel, the strongest genomic association on chromosome 18 was identical in both approaches. This suggests that the BovineSNP50 density chip was competent in identifying accurate genomic regions associated with calving performance in the Holstein Friesian population, presumably due to the high levels of LD detected within this breed (McKay et al., 2007). The use of the HD panel however did refine this association as significant neighbouring additional SNPs were identified, and the panel also permitted the identification of stronger associations on additional chromosomes such as chromosome 6, presumably due to the greater SNP coverage depth. The identification of an association on chromosome 18 in two SNP density panels, combined with its high significance in two different GWAS approaches provided strong evidence for an underlying causative mutation. The availability of whole-genome sequence data from the 1000 Bulls Project in Chapter 5, allowed this genomic region to be interpreted whilst also refining additional strong genomic associations, permitting the accurate identification of possible candidate genes which was limited in Chapter 4.

The identification of polymorphisms that accounted for a large proportion of the genomic variation in calving difficulty across all breeds suggests that genomic selection algorithms should facilitate such a distribution of allelic effects. Genomic evaluations for all traits in Ireland are currently undertaken using a G-BLUP which assumed equal variances for all SNPs. Bayesian approaches, in particular Bayes B, have been shown to have a higher accuracy than G-BLUP when the number of QTL influencing the trait is low (Daetwyler et al., 2010). This approach allows for heterogeneity among SNP contributions to the variance, permitting some large allelic effects while the remaining effects are small or possibly zero (Meuwissen et al., 2001). Habier et al., (2010) concluded that for the genomic evaluation of milk, fat and protein yield and somatic cell score, Bayes B was more accurate than GBLUP, especially as the training population size increase. Such

Bayesian approaches may therefore be better suited to genomic evaluations for direct calving difficulty as large influential SNP contributions per breed were identified in this thesis.

The association between dystocia and perinatal mortality in cattle is well known (Mee et al., 2008). Little however is known about the prevalence of perinatal mortality not associated with dystocia although anecdotal evidence suggests it is on the increase. This was the first study to estimate the prevalence and genetic parameters for such a trait in Ireland. In total over 53% of perinatal mortality in Ireland between the years 2008 to 2012 were not associated with calving difficulty. There are many possible causes for such mortality events including congenital defects, placental dysfunction, infections, precalving nutrition or possibly management issues such as age at first calving. Although a heritability was only detected in primiparae for the direct component of perinatal mortality not associated with dystocia, sufficient genetic variation existed for selection. The lack of genetic variation in later parities may be because the genetic variation has been purged out through selection, especially in beef herds since a primiparous cow with no calf may be slaughtered immediately. Further research is needed to fully comprehend the genetic variation of this trait, although, the recent emphasis of the breeding industry to identify animals that are carriers of genetic defects may help to lower the prevalence of this phenomenon. The identification of genomic regions for perinatal mortality primarily associated with morphological abnormalities in Chapters 4 and 5 may explain some of the genetic variation causing these deaths. Producers can now also voluntarily record a description of morphological abnormalities in animals, which could subsequently be used to detect and purge out unfavourable (and possibly semi-lethal) mutations thus also possibility reducing the incidence of perinatal mortality not associated with dystocia.

Custom-made genotyping SNP chips are becoming increasingly popular. In Ireland, a custom platform (IDB SNP chip; Mullen et al., 2013) comprising over 19,000 SNPs has been developed to decrease the cost of genomic selection (retail value of €30). As well as including SNPs for accurate imputation to higher density, genomic mutations of four prevalent lethal recessive diseases (CVM, DUMPs, BLAD, Brachyspina) are also included on the genotyping panel in addition to mutations with large effects on performance such as growth differentiation factor-8. Several hundred structural variants under investigation as part of national molecular genetic projects are also included on the platform facilitating subsequent association analyses once phenotypes exist on the genotyped animals. Because of the national recording system in Ireland, phenotypes will be available on animals (including stillbirths) and fine-mapping of likely causal mutations can be undertaken particularly now as whole genome sequence data is available. The inclusion of the causative mutation identified in CTU1 associated with direct calving difficulty in the present study, on this custom-made genotyping panel is a possibility; however its functional impact should be confirmed. Inclusion of this mutation on the panel may provide a cheap and readily available method to validate this mutation before prospective in-vitro analysis is carried out.

In conclusion, genetic improvement for calving performance is possible in Irish dairy and beef populations and the wide availability of low-cost genomic tools will be beneficial to achieve this more rapidly. An understanding of the genomic architecture will help facilitate this improvement by allowing for a targeted breeding approach and alternative statistical algorithms that facilitate the inclusion of SNPs with large effects in genomic prediction equations should be investigated for calving performance traits. More importantly however, genomic predictions for breeds not adequately represented in the population to derive the genomic prediction equations should be undertaken with caution.

References

- Allen, A., G. Minozzi, E. Glass, R. Skuce, S. McDowell, J. Woolliams, and S. Bishop. 2010. Bovine tuberculosis: the genetic basis of host susceptibility. *Proc Biol Sci* 277(1695):2737-2745.
- Amer, P. R., G. Simm, M. G. Keane, M. G. Diskin, and B. W. Wickham. 2001. Breeding objectives for beef cattle in Ireland. *Lives Prod Sci* 67(3):223-239.
- Ashwell, M. S., D. W. Heyen, J. I. Weller, M. Ron, T. S. Sonstegard, C. P. Van Tassell, and H. A. Lewin. 2005. Detection of quantitative trait loci influencing conformation traits and calving ease in Holstein-Friesian cattle. *J Dairy Sci* 88(11):4111-4119.
- Autry, A. E. and L. M. Monteggia. 2012. Brain-Derived Neurotrophic Factor and Neuropsychiatric Disorders. *Pharmacol Rev* 64(2):238-258.
- Azzam, S. M. and M. K. Nielsen. 1987. Genetic-Parameters for Gestation Length, Birth Date and 1st Breeding Date in Beef-Cattle. *J Anim Sci* 64(2):348-356.
- Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev. 2010. A method and server for predicting damaging missense mutations. *Nature Methods* 7(4):248-249.
- Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263-265.
- Beck, L., C. Leroy, S. Beck-Cormier, A. Forand, C. Salaun, N. Paris, A. Bernier, P. Urena-Torres, D. Prie, M. Ollero, L. Coulombel, and G. Friedlander. 2010. The Phosphate Transporter PiT1 (Slc20a1) Revealed As a New Essential Gene for Mouse Liver Development. *PLoS One* 5(2).
- Bellows, R. A., R. E. Short, D. C. Anderson, B. W. Knapp, and O. F. Pahnish. 1971. Cause and effect relationships associated with calving difficulty and calf birth weight. *J Anim Sci* 33(2):407-415.

- Berger, P. J., A. C. Cubas, K. J. Koehler, and M. H. Healey. 1992. Factors affecting dystocia and early calf mortality in Angus cows and heifers. *J Anim Sci* 70(6):1775-1786.
- Berglund, B., L. Steinbock, and M. Elvander. 2003. Causes of stillbirth and time of death in Swedish Holstein calves examined post mortem. *Acta Vet Scand* 44(3-4):111-120.
- Berry, D., L. Shalloo, A. Cromie, V. Olori, R. Veerkamp, P. Dillon, P. Amer, R. Evans, F. Kearney, and B. Wickham, 2007a, The Economic Breeding Index: a generation on, 29/11/10, http://www.icbf.com/publications/files/economic_breeding_index.pdf
- Berry, D. P. 2012. Across-breed genomic evaluations in cattle. in *Proc. European Federation of Animal Science (63rd EAAP)*, Bratislava, Wageningen Academic Publishers.
- Berry, D. P., J. W. Bastiaansen, R. F. Veerkamp, S. Wijga, E. Wall, B. Berglund, and M. P. Calus. 2012. Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. *Animal* 6(8):1206-1215.
- Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2002. Genetic parameters for level and change of body condition score and body weight in dairy cows. *J Dairy Sci* 85(8):2030-2039.
- Berry, D. P., B. Coughlan, B. Enright, S. Coughlan, and M. Burke. 2013. Factors associated with milking characteristics in dairy cows. *J Dairy Sci* 96(9):5943-5953.
- Berry, D. P. and A. R. Cromie. 2009. Associations between age at first calving and subsequent performance in Irish spring calving Holstein-Friesian dairy cows. *Livest Sci* 123(123):44-54.
- Berry, D. P. and R. D. Evans. 2014. Genetics of reproductive performance in seasonal calving beef cows and its association with performance traits. *J Anim Sci* 92(4):1412-1422.
- Berry, D. P. and F. Kearney, 2009, *Genomic Selection Procedures for Ireland* (version 1.1). Accessed 11/12/2010, http://www.icbf.com/publications/files/genomic_selection_procedures_v1.pdf

Berry, D. P., J. F. Kearney, and J. R. Roche. 2011a. Evidence of genetic and maternal effects on secondary sex ratio in cattle. *Theriogenology* 75(6):1039-1044.

Berry, D. P., J. M. Lee, K. A. Macdonald, and J. R. Roche. 2007b. Body condition score and body weight effects on dystocia and stillbirths and consequent effects on postcalving performance. *J Dairy Sci* 90(9):4201-4211.

Berry, D. P., K. G. Meade, M. P. Mullen, S. Butler, M. G. Diskin, D. Morris, and C. J. Creevey. 2011b. The integration of 'omic' disciplines and systems biology in cattle breeding. *Animal* 5(4):493-505.

Bicalho, R. C., K. N. Galvão, S. H. Cheong, R. O. Gilbert, L. D. Warnick, and C. L. Guard. 2007. Effect of stillbirths on dam survival and reproduction performance in Holstein dairy cows. *J Dairy Sci* 90(6):2797-2803.

Bicalho, R. C., K. N. Galvão, L. D. Warnick, and C. L. Guard. 2008. Stillbirth parturition reduces milk production in Holstein cows. *Prev Vet Med* 84(1-2):112-120.

Bickhart, D. M., Y. L. Hou, S. G. Schroeder, C. Alkan, M. F. Cardone, L. K. Matukumalli, J. Z. Song, R. D. Schnabe, M. Ventura, J. F. Taylor, J. F. Garcia, C. P. Van Tasse, T. S. Sonstegard, E. E. Eichler, and G. E. Liu. 2012. Copy number variation of individual cattle genomes using next-generation sequencing. *Gen Res* 22(4):778-790.

Boddhireddy, P., M.J. Kelly, S. Northcutt, K. C. Prayaga, J. Rumph, and S. DeNise. 2014. Genomic predictions in Angus cattle: comparisons of sample size, response variables and clustering methods for cross-validation. *J Anim Sci* 92:485-497

Boitard, S. and D. Rocha. 2013. Detection of signatures of selective sweeps in the Blonde d'Aquitaine cattle breed. *Ani Genet* 44(5):579-583.

BordBia., 2012, Factsheet on the Irish Agirculture and Food & Drink Sector. <http://www.bordbia.ie/industryinfo/agri/pages/default.aspx>. Accessed 01/04/2014.

- Brinkman-Van der Linden, E. C., N. Hurtado-Ziola, T. Hayakawa, L. Wiggleton, K. Benirschke, A. Varki, and N. Varki. 2007. Human-specific expression of Siglec-6 in the placenta. *Glycobiology* 17(9):922-931.
- Brøndum, R.F., Su, G., Lund, M.S., Bowman, P.J., Goddard, M.E., and Hayes B.J. 2012. Genome position specific priors for genomic prediction. *BMC Genomics* 13:543
- Browning, B. L. and S. R. Browning. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am J Hum Gen* 84(2):210-223.
- Browning, S. R. and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81(5):1084-1097.
- Bush, W. S. and J. H. Moore. 2012. Chapter 11: Genome-Wide Association Studies. *Plos Comput Biol* 8(12).
- Cameron, N. D. 1997. Selection Indices and Prediction of Genetic Merit in Animal Breeding. C. International, ed. Biddles Ltd Guilford and Kings Lynn, UK.
- Cantor, R. M., K. Lange, and J. S. Sinsheimer. 2010. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am J Hum Genet* 86(1):6-22.
- Carnier, P., A. Albera, R. Dal Zotto, A.F Groen, M. Bona, and G. Bittane. 2000. Genetic parameters for direct and maternal calving ability over parities in Piedmontese cattle. *J Anim Sci*(78):2532-2539
- Casas, E., J. W. Keele, S. C. Fahrenkrug, T. P. L. Smith, L. V. Cundiff, and R. T. Stone. 1999. Quantitative analysis of birth, weaning, and yearling weights and calving difficulty in Piedmontese crossbreds segregating an inactive myostatin allele. *J Anim Sci* 77(7):1686-1692.

Casas, E., S. D. Shackelford, J. W. Keele, M. Koohmaraie, T. P. Smith, and R. T. Stone. 2003. Detection of quantitative trait loci for growth and carcass composition in cattle. *J Anim Sci* 81(12):2976-2983.

Casas, E., R. M. Thallman, and L. V. Cundiff. 2012. Birth and weaning traits in crossbred cattle from Hereford, Angus, Norwegian Red, Swedish Red and White, Wagyu, and Friesian sires. *J Anim Sci* 90(9):2916-2920.

Chan, J., L. Y. Deng, L. G. Mikael, J. Yan, L. Pickell, Q. Wu, M. A. Caudill, and R. Rozen. 2010. Low dietary choline and low dietary riboflavin during pregnancy influence reproductive outcomes and heart development in mice. *Am J Clin Nutr* 91(4):1035-1043.

Charlier, C., J. S. Agerholm, W. Coppieters, P. Karlsvkov-Mortensen, W. B. Li, G. de Jong, C. Fasquelle, L. Karim, S. Cirera, N. Cambisano, N. Ahariz, E. Mullaart, M. Georges, and M. Fredholm. 2012. A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina. *PLoS One* 7(8).

Chassagne, M., J. Barnouin, and J. P. Chacornac. 1999. Risk factors for stillbirth in Holstein heifers under field conditions in France: A prospective survey. *Theriogenology* 51(8):1477-1488.

Chen, G. K. and J. S. Witte. 2007. Enriching the analysis of genomewide association studies with hierarchical modeling. *Am J Hum Genet* 81(2):397-404.

Clowse, M. E. B., L. S. Magder, F. Witter, and M. Petri. 2005. The impact of increased lupus activity on obstetric outcomes. *Arthritis Rheum* 52(2):514-521.

CMMS, 2012, AIM Bovine Statistics Report,
<https://www.agriculture.gov.ie/media/migration/animalhealthwelfare/animalidentificationandmovement/AIMBOVINESTATISTICSREPORT2012220513.pdf>. Accessed 02/06/2013.

Cole, J. B., P. M. VanRaden, J. R. O'Connell, C. P. Van Tassell, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and G. R. Wiggans. 2009. Distribution and location of genetic effects for dairy traits. *J Dairy Sci* 92(6):2931-2946.

Collery, P., J. Bradley, J. Fagan, P. Jones, E. Redehan, and E. Weavers. 1996. Causes of perinatal calf mortality in the Republic of Ireland. *Vet Jour* 49:491-496.

Conneely, M., D. P. Berry, R. Sayers, J. P. Murphy, I. Lorenz, M. L. Doherty, and E. Kennedy. 2013. Factors associated with the concentration of immunoglobulin G in the colostrum of dairy cows. *Animal* 7(11):1824-1832.

Crosse, S. and N. Soede. 1988. The Incidence of Dystocia and Perinatal-Mortality on Commercial Dairy Farms in the South of Ireland. *Vet Jour* 42(1):8-12.

Crump, R. E., N. R. Wray, R. Thompson, and G. Simm. 1997. Assigning pedigree beef performance records to contemporary groups taking account of within-herd calving patterns. *Ani Sci* 65:193-198.

CSO, 2013, Livestock Survey December, April 1st, 2014, <http://www.cso.ie/en/releasesandpublications/er/lld/livestocksurveydecember2012/#.UzqNsPldV50>. Accessed 05/06/2013.

Cymbron, T., A. R. Freeman, M. Isabel Malheiro, J. D. Vigne, and D. G. Bradley. 2005. Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. *Proc Biol Sci* 272(1574):1837-1843.

Dabney, A. and J. D. Storey. 2004. qvalue: Q-value estimation for false discovery rate control. R package version 1.34.0. ed. <http://www.bioconductor.org/packages/release/bioc/html/qvalue.html>.

Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185(3):1021-1031.

Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, RF R, Liao X, Diari A, Rodriguez SC, Grohs C *et al.* 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, 46:858-865.

Devlin, B. and K. Roeder. 1999. Genomic control for association studies. *Biometrics* 55(4):997-1004.

Dwyer, C. M., S. K. Calvert, M. Farish, J. Donbavand, and H. E. Pickup. 2005. Breed, litter and parity effects on placental weight and placentome number, and consequences for the neonatal behaviour of the lamb. *Theriogenology* 63(4):1092-1110.

Eaglen, S. A. E., M. P. Coffey, J. A. Woolliams, and E. Wall. 2012. Evaluating alternate models to estimate genetic parameters of calving traits in United Kingdom Holstein-Friesian dairy cattle. *Genet Sel Evol* 44.

Echternkamp, S. E. 1993. Relationship between placental development and calf birth weight in cattle. *Ani Repro Sci* 32:1-13.

Erb, H. N., R. D. Smith, P. A. Oltenacu, C. L. Guard, R. B. Hillman, P. A. Powers, M. C. Smith, and M. E. White. 1985. Path Model of Reproductive Disorders and Performance, Milk Fever, Mastitis, Milk-Yield, and Culling in Holstein Cows. *J Dairy Sci* 68(12):3337-3349.

Eriksson, S., A. Näisholm, K. Johansson, and J. Philipsson. 2004. Genetic parameters for calving difficulty, stillbirth, and birth weight for Hereford and Charolais at first and later parities. *J Anim Sci* 82(2):375-383.

Evans, R. and T. Pabiou. 2012. The benefits of using farmer scored traits in beef genetic evaluations. Proc of 42nd ICAR session. Ireland.

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

Fan, B., Z. Q. Du, Gorbach D.M., and M. F. Rothschild. 2010. Development and Application of High-density SNP arrays in Genomic Studies of Domestic animals. Pages 833-847. Vol. 23. July ed, *Asian.Aust J.Anim. Sci.*

- Fan, B., S. K. Onteru, Z. Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs. *PLoS One* 6(2):e14726.
- Fenster, S. D. and C. C. Garner. 2002. Gene structure and genetic localization of the PCLO gene encoding the presynaptic active zone protein Piccolo. *Int J Dev Neurosci* 20(3-5):161-171.
- Fernando, R. and D. J. Garrick, 2011, GenSel: User Manual for a Portfolio of Genomic Selection Related Analyses. <http://big.s.ansci.iastate.edu/bigsgui>. Accessed 17/01/2013.
- Fernando, R. L. and D. J. Garrick. 2013. Genome-wide Association Studies and Genomic Prediction. Vol. 1019. *Bayesian Methods Applied to GWAS*.
- Fitzgerald, A. M., D. P. Berry, T. Carthy, A. R. Cromie, and D. P. Ryan. 2014. Risk factors associated with multiple ovulation and twin birth rate in Irish dairy and beef cattle. *Journal of Animal Science* 92(3):966-973.
- Fourichon, C., F. Beaudeau, N. Bareille, and H. Seegers. 2001. Incidence of health disorders in dairy farming systems in western France. *Livestock Production Science* 68(2-3):157-170.
- Fowden, A. L. 2003. The insulin-like growth factors and feto-placental growth. *Placenta* 24(8-9):803-812.
- Fox, J., 2011, Frequentist vs. Bayesian: resources to help you choose. <http://oikosjournal.wordpress.com/2011/10/11/frequentist-vs-bayesian-statistics-resources-to-help-you-choose/>. Accessed 02/04/2014.
- Ganss, B. and A. Jheon. 2004. Zinc finger transcription factors in skeletal development. *Crit Rev Oral Biol Med* 15(5):282-297.
- Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 41:55.

Gautier, M., T. Faraut, K. Moazami-Goudarzi, V. Navratil, M. Foglio, C. Grohs, A. Boland, J. G. Garnier, D. Boichard, G. M. Lathrop, I. G. Gut, and A. Eggen. 2007. Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* 177(2):1059-1070.

Gautier, M., L. Flori, A. Riebler, F. Jaffrézic, D. Laloë, I. Gut, K. Moazami-Goudarzi, and J. L. Foulley. 2009. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10:550.

Gautier, M., D. Laloë, and K. Moazami-Goudarzi. 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One* 5(9).

Gibbs, R. A., J. F. Taylor, C. P. Van Tassell, W. Barendse, K. A. Eversole, C. A. Gill, R. D. Green, D. L. Hamernik, S. M. Kappes, S. Lien, L. K. Matukumalli, J. C. McEwan, L. V. Nazareth, R. D. Schnabel, G. M. Weinstock, D. A. Wheeler, P. Ajmone-Marsan, P. J. Boettcher, A. R. Caetano, J. F. Garcia, O. Hanotte, P. Mariani, L. C. Skow, T. S. Sonstegard, J. L. Williams, B. Diallo, L. Hailemariam, M. L. Martinez, C. A. Morris, L. O. Silva, R. J. Spelman, W. Mulatu, K. Zhao, C. A. Abbey, M. Agaba, F. R. Araujo, R. J. Bunch, J. Burton, C. Gorni, H. Olivier, B. E. Harrison, B. Luff, M. A. Machado, J. Mwakaya, G. Plastow, W. Sim, T. Smith, M. B. Thomas, A. Valentini, P. Williams, J. Womack, J. A. Woolliams, Y. Liu, X. Qin, K. C. Worley, C. Gao, H. Jiang, S. S. Moore, Y. Ren, X. Z. Song, C. D. Bustamante, R. D. Hernandez, D. M. Muzny, S. Patil, A. San Lucas, Q. Fu, M. P. Kent, R. Vega, A. Matukumalli, S. McWilliam, G. Sclep, K. Bryc, J. Choi, H. Gao, J. J. Grefenstette, B. Murdoch, A. Stella, R. Villa-Angulo, M. Wright, J. Aerts, O. Jann, R. Negrini, M. E. Goddard, B. J. Hayes, D. G. Bradley, M. Barbosa da Silva, L. P. Lau, G. E. Liu, D. J. Lynn, F. Panzitta, K. G. Dodds, and B. H. Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324(5926):528-532.

Gibson, J., N. E. Morton, and A. Collins. 2006. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet* 15(5):789-795.

Glessner, J. T., K. Wang, G. Q. Cai, O. Korvatska, C. E. Kim, S. Wood, H. T. Zhang, A. Estes, C. W. Brune, J. P. Bradfield, M. Imielinski, E. C. Frackelton, J. Reichert, E. L. Crawford, J. Munson, P. M. A. Sleiman, R. Chiavacci, K. Annaiah, K. Thomas, C. P. Hou,

W. Glaberson, J. Flory, F. Otieno, M. Garris, L. Soorya, L. Klei, J. Piven, K. J. Meyer, E. Anagnostou, T. Sakurai, R. M. Game, D. S. Rudd, D. Zurawiecki, C. J. McDougale, L. K. Davis, J. Miller, D. J. Posey, S. Michaels, A. Kolevzon, J. M. Silverman, R. Bernier, S. E. Levy, R. T. Schultz, G. Dawson, T. Owley, W. M. McMahon, T. H. Wassink, J. A. Sweeney, J. I. Nurnberger, H. Coon, J. S. Sutcliffe, N. J. Minshew, S. F. A. Grant, M. Bucan, E. H. Cook, J. D. Buxbaum, B. Devlin, G. D. Schellenberg, and H. Hakonarson. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459(7246):569-573.

Griffiths, A. J., Miller, Jeffrey H., Suzuki, David T., Lewontin, Richard C., Gelbart, William M.,. 2000. *An Introduction to Genetic Analysis*. 7th ed. New York: W. H. Freeman.

Grosz, M. D. and M. D. MacNeil. 2001. Putative quantitative trait locus affecting birth weight on bovine chromosome 2. *J Anim Sci* 79(1):68-72.

Gulliksen, S. M., K. I. Lie, T. Loken, and O. Osteras. 2009. Calf mortality in Norwegian dairy herds. *J Dairy Sci* 92(6):2782-2795.

Gundelach, Y., K. Essmeyer, M. K. Teltscher, and M. Hoedemaker. 2009. Risk factors for perinatal mortality in dairy cattle: Cow and foetal factors, calving process. *Theriogenology* 71(6):901-909.

Gustafsson, H., H. Kindahl, and B. Berglund. 2007. Stillbirths in Holstein heifers - some results from Swedish research. *Acta Vet Sca* 49.

Gutierrez-Gil, B., J. L. Williams, D. Homer, D. Burton, C. S. Haley, and P. Wiener. 2009. Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle. *J Ani Sci* 87(1):24-36.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinfor* 12:186.

Hanotte, O., C. L. Tawah, D. G. Bradley, M. Okomo, Y. Verjee, J. Ochieng, and J. E. Rege. 2000. Geographic distribution and frequency of a taurine *Bos taurus* and an indicine

Bos indicus Y specific allele amongst sub-saharan African cattle breeds. *Mol Ecol* 9(4):387-396.

Hansen, M., M. S. Lund, M. K. Sørensen, and L. G. Christensen. 2002. Genetic parameters of dairy character, protein yield, clinical mastitis and other disease in the Danish Holstein cattle. *J Dairy Sci* 85(2):445-452.

Hansen, M., M. S. Lund, J. Pedersen, and L. G. Christensen. 2004. Genetic parameters for stillbirth in Danish Holstein cows using a Bayesian threshold model. *J Dairy Sci* 87(3):706-716.

Hansen, M., M. S. Lund, J. Pedersen, and L. G. Christensen. 2004. Gestation length in Danish Holsteins has weak genetic associations with stillbirth, calving difficulty, and calf size. *Livest Prod Sci* 91(2):23-33.

Harris, B. and D. Johnson. 1998. Approximate reliability of genetic evaluations under an animal model. *J Dairy Sci* 81(10):2723-2728.

Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* 13(4):635-643.

Hayes, B.J., P.J Bowman, A.C. Chamberlain, K. Verbyla, and M.E. Goddard. 2009. Accuracy of genomic breeding values in mutli-breed dairy cattle populations. *Genet Sel Evol* 41:51

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. *Plos Genetics* 6(9).

Heins, B. J., L. B. Hansen, and A. J. Seykora. 2006. Calving difficulty and stillbirths of pure Holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and Scandinavian Red. *J Dairy Sci* 89(7):2805-2810.

- Hickey, J. M., M. G. Keane, D. A. Kenny, A. R. Cromie, P. R. Amer, and R. F. Veerkamp. 2007. Heterogeneity of genetic parameters for calving difficulty in Holstein heifers in Ireland. *J Dairy Sci* 90(8):3900-3908.
- Hildebrandt, F., S. F. Heeringa, F. Rüschendorf, M. Attanasio, G. Nürnberg, C. Becker, D. Seelow, N. Huebner, G. Chernin, C. N. Vlangos, W. Zhou, J. F. O'Toole, B. E. Hoskins, M. T. Wolf, B. G. Hinkes, H. Chaib, S. Ashraf, D. S. Schoeb, B. Ovunc, S. J. Allen, V. Vega-Warner, E. Wise, H. M. Harville, R. H. Lyons, J. Washburn, J. Macdonald, P. Nürnberg, and E. A. Otto. 2009. A systematic approach to mapping recessive disease genes in individuals from outbred populations. *PLoS Genet* 5(1):e1000353.
- Hirschhorn, J. N. and M. J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6(2):95-108.
- Höglund, J. K., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2012. Analyzes of genome-wide association follow-up study for calving traits in dairy cattle. *BMC Genet* 13:71.
- Hosseini-Zadeh, N. G. 2014. Effect of Dystocia on the Productive Performance and Calf Stillbirth in Iranian Holsteins. *J Agr Sci Tech-Iran* 16(1):69-78.
- Hou, Y. L., G. E. Liu, D. M. Bickhart, M. F. Cardone, K. Wang, E. S. Kim, L. K. Matukumalli, M. Ventura, J. Z. Song, P. M. VanRaden, T. S. Sonstegard, and C. P. Van Tassell. 2011. Genomic characteristics of cattle copy number variations. *BMC Genomics* 12.
- Huxley, J. N. and H. R. Whay. 2006. Current attitudes of cattle practitioners to pain and the use of analgesics in cattle. *Vet Rec* 159(20):662-668.
- Ibeagha-Awemu, E. M. and G. Erhardt. 2005. Genetic structure and differentiation of 12 African *Bos indicus* and *Bos taurus* cattle breeds, inferred from protein and microsatellite polymorphisms. *J Anim Breed Genet* 122(1):12-20.
- ICBF, 2008, Learn more about ICBF. <http://www.icbf.com/aboutus/index.php>. Accessed 11/12/2010

ICBF, 2012, Introduction of New Terminal and Maternal Beef Indexes. http://www.icbf.com/publications/files/Intro_of_new_terminal_maternal_beef_indexes.pdf. Accessed 10/04/2014.

Illumina, Illumina GenCall Data Analysis Software. http://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf. Accessed 28/11/2010.

Illumina, 2010, Bovine HD Genotyping BeadChip Datasheet, http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf. Accessed 28/11/2010.

Jamrozik, J., J. Fatehi, G. J. Kistemaker, and L. R. Schaeffer. 2005. Estimates of genetic parameters for Canadian Holstein female reproduction traits. *J Dairy Sci* 88(6):2199-2208.

Jeffreys, H. 1961. The theory of probability, 3rd edition. Oxford University Press, United Kingdom.

Jiang, H., G. D. Weesner, and P. V. Malven. 1997. CDNA sequence and expression of bovine prodynorphin. *Gene* 186(2):279-283.

Johanson, J. M. and P. J. Berger. 2003. Birth weight as a predictor of calving ease and perinatal mortality in Holstein cattle. *J Dairy Sci* 86(11):3745-3755.

Johanson, J. M., P. J. Berger, and Kirkpatrick B.W. 2001. Twinning rates for North American Holsteins. *J Dairy Sci* 84(9):2081-2088.

Kass, R. and A. Raftery. 1995. Bayes Factors. *Journal of American Statistics* 90:773-795.
Khatri, P., M. Sirota, and A. J. Butte. 2012. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *Plos Comput Biol* 8(2).

Kemper, K.E., H. D. Daetwyler, P.M. Visscher and M.E. Goddard. 2012. Comparing linkage and association analyses in sheep points to a better way of doing GWAS. *Genetics Research* 94(04):191-203.

- Kim, H., T. Kim, B. C. Jeong, I. T. Cho, D. Han, N. Takegahara, T. Negishi-Koga, H. Takayanagi, J. H. Lee, J. Y. Sul, V. Prasad, S. H. Lee, and Y. Choi. 2013. Tmem64 modulates calcium signaling during RANKL-mediated osteoclast differentiation. *Cell Metab* 17(2):249-260.
- Kirin, M., R. McQuillan, C. S. Franklin, H. Campbell, P. M. McKeigue, and J. F. Wilson. 2010. Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5(11):e13996.
- Kirkbride, C. A. 1993. Bacterial Agents Detected in a 10-Year Study of Bovine Abortions and Stillbirths. *J Vet Diagn Invest* 5(1):64-68.
- Kolbe, J. J., A. Larson, J. B. Losos, and K. de Queiroz. 2008. Admixture determines genetic diversity and population differentiation in the biological invasion of a lizard species. *Biol Lett* 4(4):434-437.
- Krasilnikov, M. A. 2000. Phosphatidylinositol-3 kinase dependent pathways: the role in control of cell growth, survival, and malignant transformation. *Biochemistry-Moscow* 65(1):59-67.
- Kriese, L. A., L. D. Vanvleck, K. E. Gregory, K. G. Boldman, L. V. Cundiff, and R. M. Koch. 1994. Estimates of Genetic-Parameters for 320-Day Pelvic Measurements of Males and Females and Calving Ease of 2-Year-Old Females. *J Anim Sci* 72(8):1954-1963.
- Kühn, C., J. Bennewitz, N. Reinsch, N. Xu, H. Thomsen, C. Looft, G. A. Brockmann, M. Schwerin, C. Weimann, S. Hiendleder, G. Erhardt, I. Medjugorac, M. Förster, B. Brenig, F. Reinhardt, R. Reents, I. Russ, G. Averdunk, J. Blümel, and E. Kalm. 2003. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J Dairy Sci* 86(1):360-368.
- Kuhn, R. and W. Meyer. 2007. Comparative primary hair structure in the lutrinae (Carnivora : Mustelidae). *J Morphol* 268(12):1095-1096.
- Lavergne, S. and J. Molofsky. 2007. Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc Natl Acad Sci U S A* 104(10):3883-3888.

Lee, A. J. 1997. The interplay of feeding and genetics on heifer rearing and first lactation milk yield: a review. *J Anim Sci* 75(3):846-851.

Lencz, T., C. Lambert, P. DeRosse, K. E. Burdick, T. V. Morgan, J. M. Kane, R. Kucherlapati, and A. K. Malhotra. 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A* 104(50):19942-19947.

Li, C., J. Basarab, W. M. Snelling, B. Benkel, B. Murdoch, and S. S. Moore. 2002. The identification of common haplotypes on bovine chromosome 5 within commercial lines of *Bos taurus* and their associations with growth traits. *J Anim Sci* 80(5):1187-1194.

Liu, C. T., A. Calogero, G. Ragona, E. Adamson, and D. Mercola. 1996. EGR-1, the reluctant suppression factor: EGR-1 is known to function in the regulation of growth, differentiation, and also has significant tumor suppressor activity and a mechanism involving the induction of TGF-beta 1 is postulated to account for this suppressor activity. *Crit Rev Oncogenesis* 7(1-2):101-125.

Liu, J. Z., S. E. Medland, M. J. Wright, A. K. Henders, A. C. Heath, P. A. Madden, A. Duncan, G. W. Montgomery, N. G. Martin, and A. F. McRae. 2010. Genome-wide association study of height and body mass index in Australian twin families. *Twin Res Hum Genet* 13(2):179-193.

Loftus, R. T., D. E. MacHugh, D. G. Bradley, P. M. Sharp, and P. Cunningham. 1994. Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A* 91(7):2757-2761.

Lohi, H., M. Kujala, S. Makela, E. Lehtonen, M. Kestila, U. Saarialho-Kere, D. Markovich, and J. Kere. 2002. Functional characterization of three novel tissue-specific anion exchangers SLC26A7, -A8, and -A9. *J Biol Chem* 277(16):14246-14254.

Lund, M. S., B. Guldbrandtsen, A. J. Buitenhuis, B. Thomsen, and C. Bendixen. 2008. Detection of quantitative trait loci in Danish Holstein cattle affecting clinical mastitis, somatic cell score, udder conformation traits, and assessment of associated effects on milk yield. *J Dairy Sci* 91(10):4028-4036.

Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. T. Liu, R. Reents, C. Schrooten, F. Seefried, and G. S. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol* 43.

Lund M. S., G. Su., J. Luc., B. Guldbrandtsen, B. Brøndum, and Froberg, R. 2014. Genomic evaluation of cattle in a multi-breed context. *Livestock Science*, 166:101-110.

Luo, M., P. Boettcher, L. Schaeffer, and J. Dekkers. 2002. Estimation of genetic parameters of calving ease in first and second parities of Canadian Holsteins using Bayesian methods. 74:175-184.

Lynch, M. and B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

MacHugh, D. E., M. D. Shriver, R. T. Loftus, P. Cunningham, and D. G. Bradley. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* 146(3):1071-1086.

Maltecca, C., K. A. Gray, K. A. Weigel, J. P. Cassady, and M. Ashwell. 2011. A genome-wide association study of direct gestation length in US Holstein and Italian Brown populations. *Anim Genet* 42(6):585-591.

Mao, X., G. Sahana, D. J. Koning, and B. Guldbrandtsen. 2013. Population level genome-wide association study for calving traits in Holstein cattle. in *Proc. EAAP-64th Annual Meeting*, Nantes.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* 4(4):e5350.

Mc Hugh, N., A. R. Cromie, R. D. Evans, and D. P. Berry. 2014. Validation of national genetic evaluations for maternal beef cattle traits using Irish field data. *J Ani Sci* 92(4):1423-1432.

Mc Parland, S., J. F. Kearney, M. Rath, and D. P. Berry. 2007a. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. *J Dairy Sci* 90(9):4411-4419.

Mc Parland, S., J. F. Kearney, M. Rath, and D. P. Berry. 2007b. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. *J Anim Sci* 85(2):322-331.

McClintock, S. E., K. T. Beard, A. Gilmour, and M. E. Goddard. 2003. Relationships between Calving Traits in Heifers and Mature Cows in Australia. Pages 102-106. Conference- Interbull. Italy.

McClintock, S. E., K. T. Beard, M. E. Goddard, and D. J. Johnston. 2005. Interactions between gestation length, calf size, dystocia and calf mortality. Pages 275-278 in *Proc. Association for the Advancement of Animal Breeding and Genetics (AAABG)*, Australia.

McClure, M. C., N. S. Morsci, R. D. Schnabel, J. W. Kim, P. Yao, M. M. Rolf, S. D. McKay, S. J. Gregg, R. H. Chapple, S. L. Northcutt, and J. F. Taylor. 2010. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Anim Genet* 41(6):597-607.

McGuirk, B., I. Going, and A. Gilmour. 1999. The genetic evaluation of UK Holstein Friesian sires for calving ease and related traits. *Ani Sci* 68:413-422.

McKay, S. D., R. D. Schnabel, B. M. Murdoch, L. K. Matukumalli, J. Aerts, W. Coppieters, D. Crews, E. Dias Neto, C. A. Gill, C. Gao, H. Mannen, P. Stothard, Z. Wang, C. P. Van Tassell, J. L. Williams, J. F. Taylor, and S. S. Moore. 2007. Whole genome linkage disequilibrium maps in cattle. *BMC Genet* 8:74.

McQuillan, R., A. L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. Macleod, S. M. Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S. H. Wild, M. G. Dunlop, A. F.

- Wright, H. Campbell, and J. F. Wilson. 2008. Runs of homozygosity in European populations. *Am J Hum Genet* 83(3):359-372.
- Mee, J. F. 2013. Why Do So Many Calves Die on Modern Dairy Farms and What Can We Do about Calf Welfare in the Future? *Animals* 3:1036-1057.
- Mee, J. F., D. P. Berry, and A. R. Cromie. 2008. Prevalence of, and risk factors associated with, perinatal calf mortality in pasture-based Holstein-Friesian cows *Animal* 2(4):613-620.
- Mee, J. F., D. P. Berry, and A. R. Cromie. 2011. Risk factors for calving assistance and dystocia in pasture-based Holstein-Friesian heifers and cows in Ireland. *Vet J* 187(2):189-194.
- Meghen, C., D. MacHugh, B. Sauveroché, G. Kana, and D. G. Bradley. 2000. Characterization of the Kuri cattle of Lake Chad using molecular genetic techniques. in *The origin and development of African livestock*. R. M. B. MacDonald, amp, and K.C, ed, London: University College Press
- Meijering, A. 1983. Dystocia and stillbirth in cattle- A review of causes, relations and implications. *Livest Prod Sci* (11):143-177.
- Meuwissen, T. and Z. Luo. 1992. Computing inbreeding coefficients in large populations. *Genet sel evol* 24(4):305-313.
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819-1829.
- Meyer, C. L., P. J. Berger, and K. J. Koehler. 2000. Interactions among factors affecting stillbirths in Holstein cattle in the United States. *J Dairy Sci* 83(11):2657-2663.
- Meyer, C. L., P. J. Berger, K. J. Koehler, J. R. Thompson, and C. G. Sattler. 2001. Phenotypic trends in incidence of stillbirth for Holsteins in the United States. *J Dairy Sci* 84(2):515-523.

Meyer, K. 2007. WOMBAT-A tool for mixed model analyses in quantitative genetics by REML. *Journal of Zhejiang University Science* 8(11):815-821.

Meyers, S. N., T. G. McDanel, S. L. Swist, B. M. Marron, D. J. Steffen, D. O'Toole, J. R. O'Connell, J. E. Beever, T. S. Sonstegard, and T. P. Smith. 2010. A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. *BMC Genomics* 11:337.

Moore, J. H. and J. C. Dunlap. 2010. Statistical Methods for Pathway Analysis of Genome-Wide Data for Association with Complex Genetic Traits. *Computational Methods for Genetics of Complex Traits*, Academic Press, Vol 72,141-179.

More O'Ferrall, G. J. 1990. Phenotypic and Genetic Parameters of Production Traits in Irish Fresian Cows. Pages 95-100. Vol. 29, *Irish J Agric Res*.

Mrode, R. A. and R. Thompson. 2005. Linear Models for the Prediction of Animal Breeding values. 2nd Edition ed. CABI Publishing.

Mullen, M. P., M. C. McClure, J. F. Kearney, S. M. Waters, R. Weld, and P. Flynn. 2013. Development of a Custom SNP Chip for Dairy and Beef Cattle Breeding, Parentage and Research. Pages 58-66 in *Proc. Interbull*, Nantes, France.

Muskens, J. 2008. Stillbirth in heifers: Experiences in the Netherlands. Pages 320-321 in *Proc. 25th World Buiatrics Congress*, Budapest, Hungary.

Nagahata, H. 2004. Bovine leukocyte adhesion deficiency (BLAD): A review. *J Vet Med Sci* 66(12):1475-1482.

Newman, S., M. D. MacNeil, W. L. Reynolds, B. W. Knapp, and J. J. Urick. 1993. Fixed effects in the formation of a composite line of beef cattle: I. Experimental design and reproductive performance. *J Anim Sci* 71(8):2026-2032.

Nielen, M., Y. H. Schukken, D. T. Scholl, H. J. Wilbrink, and A. Brand. 1989. Twinning in dairy cattle: A study of risk factors and effects. *Theriogenology* 32(5):845-862.

- Nielsen, L. B., F. S. Pedersen, and L. Pedersen. 2001. Expression of type III sodium-dependent phosphate transporters/retroviral receptors mRNAs during osteoblast differentiation. *Bone* 28(2):160-166.
- Niskanen., S. and J. Juga. 1998. Calving difficulties and calf mortality in Finnish dairy cattle population. Pages 17-20 in *Proc. Interbull*, New Zealand.
- Ng, P. C. and S. Henikoff. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31(13):3812-3814.
- Nyholt, D.R., 2004. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 74:765-769
- O'Mara, F., 2012, Country Pasture/Forage Resource Profile (Ireland). <http://www.fao.org/ag/agp/AGPC/doc/Counprof/Ireland/Ireland.htm>. Accessed 01/04/2014.
- Olsen, H. G., B. J. Hayes, M. P. Kent, T. Nome, M. Svendsen, and S. Lien. 2010. A genome wide association study for QTL affecting direct and maternal effects of stillbirth and dystocia in cattle. *Anim Genet* 41(3):273-280.
- Onteru, S. K., B. Fan, M. T. Nikkilä, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Whole-genome association analyses for lifetime reproductive traits in the pig. *J Anim Sci* 89(4):988-995.
- Ostersen, T., O. F. Christensen, M. Henryon, B. Nielsen, G. S. Su, and P. Madsen. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genet Sel Evol* 43.
- Pare, G. 2010. Genome-wide association studies--data generation, storage, interpretation, and bioinformatics. *J Cardiovasc Transl Res* 3(3):183-188.
- Pearson, T. A. and T. A. Manolio. 2008. How to interpret a genome-wide association study. *JAMA* 299(11):1335-1344.

Philipsson, J. 1976. Studies on calving difficulty, stillbirth and associated factors in Swedish cattle breeds.II. *Acta Agric Sca* 26:165-174.

Philipsson, J. 1979. Breeding for calving performance. *Calving Problems and Early viability of the calf*. The Hague.

Phocas, F. and D. Laloe. 2004. Genetic parameters for birth and weaning traits in French specialized beef cattle breeds. *Livest Prod Sci* 89(2-3):121-128.

Popescu, S., C. Borda, E. A. Diugan, M. Spinu, I. S. Groza, and C. D. Sandru. 2013. Dairy cows welfare quality in tie-stall housing system with or without access to exercise. *Acta Vet Sca* 55.

Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. *Am J Hum Genet* 67(1):170-181.

Pryce, J., B. Harris, S. Sim, and A. McPherson. 2006. Genetics of stillbirth in dairy calves. Pages 98-102 in *Proc. Proceedings of the New Zealand Society of Animal Production* 66.

Pryce, J. E., B. J. Hayes, S. Bolormaa, and M. E. Goddard. 2011. Polymorphic regions affecting human height also control stature in cattle. *Genetics* 187(3):981-984.

Purfield, D. C., D. G. Bradley, J. F. Kearney, and D. P. Berry. 2013. Genome-wide association study for calving traits in Holstein-Friesian dairy cattle. *Animal*:1-12.

Qanbari, S., D. Gianola, B. Hayes, F. Schenkel, S. Miller, S. Moore, G. Thaller, and H. Simianer. 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Gen* 12.

Ramsbottom, G., A. R. Cromie, B. Horan, and D. P. Berry. 2012. Relationship between dairy cow genetic merit and profit on commercial spring calving dairy farms. *Animal* 6(7):1031-1039.

Sahana, G., B. Guldbrandtsen, L. Janss, and M. S. Lund. 2010. Comparison of Association Mapping Methods in a Complex Pedigreed Population. *Genet Epidemiol* 34(5):455-462.

- Sahana, G., B. Guldbrandtsen, and M. S. Lund. 2011. Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. *J Dairy Sci* 94(1):479-486.
- Schnabel, R. D., T. S. Sonstegard, J. F. Taylor, and M. S. Ashwell. 2005. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Anim Genet* 36(5):408-416.
- Schrooten, C., H. Bovenhuis, W. Coppieters, and J. A. Van Arendonk. 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *J Dairy Sci* 83(4):795-806.
- Schwenger, B., S. Schober, and D. Simon. 1993. Dumps Cattle Carry a Point Mutation in the Uridine Monophosphate Synthase Gene. *Genomics* 16(1):241-244.
- Seidenspinner, T., J. Bennewitz, F. Reinhardt, and G. Thaller. 2009. Need for sharp phenotypes in QTL detection for calving traits in dairy cattle. *J Anim Breed Genet* 126(6):455-462.
- Seidenspinner, T., J. Tetens, D. Habier, J. Bennewitz, and G. Thaller. 2011. The placental growth factor (PGF)--a positional and functional candidate gene influencing calving ease and stillbirth in German dairy cattle. *Anim Genet* 42(1):22-27.
- Sonstegard, T. S., C. Van Tassell, and C. Wolfe. 20 Jun 2011. Cattle from the Isle of Jersey: A perfect storm for change. *Globaldiv Newsletter* (19):15-18.
- Steinbock, L., A. Näsholm, B. Berglund, K. Johansson, and J. Philipsson. 2003. Genetic effects on stillbirth and calving difficulty in Swedish Holsteins at first and second calving. *J Dairy Sci* 86(6):2228-2235.
- Stephens, M. and D. J. Balding. 2009. Bayesian statistical methods for genetic association studies. *Nat Rev Genet* 10(10):681-690.
- Stranger, B. E., E. A. Stahl, and T. Raj. 2011. Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. *Genetics* 187(2):367-383.

- Szkiba, D., M. Kapun, A. Haeseler, and M. Gallach. 2014. SNP2GO:Functional analysis of genome-wide association studies. *Genetics*. Epub 10.1534/genetics.113.160341
- Tenesa, A., P. Navarro, B. J. Hayes, D. L. Duffy, G. M. Clarke, M. E. Goddard, and P. M. Visscher. 2007. Recent human effective population size estimated from linkage disequilibrium. *Genome Res* 17(4):520-526.
- Thomsen, B., P. Horn, F. Panitz, E. Bendixen, A. H. Petersen, L. E. Holm, V. H. Nielsen, J. S. Agerholm, J. Arnbjerg, and C. Bendixen. 2006. A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Res* 16(1):97-105.
- Timmons, B. C., A. M. Fairhurst, and M. S. Mahendroo. 2009. Temporal Changes in Myeloid Cells in the Cervix during Pregnancy and Parturition. *J Immun* 182(5):2700-2707.
- Ullrich, O., S. Reinsch, S. Urbé, M. Zerial, and R. G. Parton. 1996. Rab11 regulates recycling through the pericentriolar recycling endosome. *J Cell Biol* 135(4):913-924.
- Van De Berg, I., S. Fritz., A. Djari., S.C. Rodriguez., D. Esquerre., C. Klopp., D. Rochar., and M. S. Lund.2013. Proceedings of 64TH Annual Meeting of the European Association for Animal Production. Book of Abstracts. No.19 p234.
- Visscher, P. M., S. E. Medland, M. A. Ferreira, K. I. Morley, G. Zhu, B. K. Cornes, G. W. Montgomery, and N. G. Martin. 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* 2(3):e41.
- Wickham, B. W. 2004. Establishing a Shared Cattle Breeding Database-Recent Experiences in Ireland. in Proc. 34th ICAR.
- Wiener, P., D. Burton, and J. L. Williams. 2004. Breed relationships and definition in British cattle: a genetic analysis. *Heredity* 93(6):597-602.

Wiener, P., I. Maclean, J. L. Williams, and J. A. Woolliams. 2000. Testing for the presence of previously identified QTL for milk production traits in new populations. *Anim Genet* 31(6):385-395.

Yun, S., S. H. Lee, S. R. Yoon, M. S. Kim, Z. H. Piao, P. K. Myung, T. D. Kim, H. Jung, and I. Choi. 2011. TOX regulates the differentiation of human natural killer cells from hematopoietic stem cells in vitro. *Immunol Lett* 136(1):29-36.

Zou, K. H., K. Tuncali, and S. G. Silverman. 2003. Correlation and simple linear regression. *Radiology* 227(3):617-622.

Publications from the Thesis

Journal publications

Purfield, D.C., Berry, D.P., McParland, S., & Bradley, D.G. 2012. *Runs of homozygosity and population history in cattle*. BMC Genetics 13:70

Purfield, D.C., Bradley, D.G., Kearney, J.F., & Berry, D.P. 2014. *Genome-wide association study for calving traits in Holstein-Friesian cattle*. Animal 8(2):224-35.

Pertoldi, C., **Purfield**, D.C., Berg, P., Jensen, TH., Bach, OS., Vingborg, R. Kristensen, TN. 2014. *Genetic characterization of a herd of the endangered Danish Jutland cattle*. J Anim Sci. 92:2372-2376

Presentations

Purfield, D.C., Bradley, D.G., Kearney, J.F., & Berry, D.P. 2014. '*Genome wide association study for calving performance in Irish cattle*'. Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. Vancouver, Canada.

Purfield, D.C., Bradley, D.G., Kearney, J.F., & Berry, D.P. 2014. '*Genome wide association study for calving performance in Irish dairy and beef cattle using high-density genotypes*'. Proceedings of Agricultural Research Forum, Tullamore, Ireland. p138

Purfield, D.C., Bradley, D.G., McClure, M., & Berry, D.P. 2014. '*Characterisation of copy number variation in Holstein-Friesian cattle*'. Proceedings of Agricultural Research Forum, Tullamore, Ireland. p138

Purfield, D.C., Bradley, D.G., Kearney, J.F., & Berry, D.P. 2013. '*Genome wide association study for calving performance in Irish cattle*'. Proceedings of the 64th Annual Meeting of the European Federation of Animal Science (EAAP). Book of Abstracts No.19, Nantes, France. p606

Purfield, D.C., Bradley, D.G., Evans, R.D & Berry, D.P. 2013. '*Genetic parameters for calving traits in Irish cattle*'. Proceedings of Agricultural Research Forum, Tullamore, Ireland. p75

Purfield, D.C., Bradley, D.G., Kearney, J.F., & Berry, D.P. 2013. '*Genome wide association study for calving performance in Irish cattle*', Proceedings of Agricultural Research Forum, Tullamore, Ireland. p55

Purfield, D.C. & Coyne, J.M. 2013. '*Genome wide association study for body conformation in Irish dairy cattle*', Proceedings of Agricultural Research Forum, Tullamore, Ireland. p61

Purfield, D.C., Bradley, D.G., Kearney, J.F., Berry, D.P. 2012. '*Genetic parameters for perinatal mortality in dairy and beef herds*', Proceedings of International Committee for Animal Recording Conference, Cork, Ireland. p29

Purfield, D.C., Berry, D.P., McParland, S., & Bradley, D.G. 2012. '*Runs of Homozygosity in several Irish cattle breeds*', Proceedings of Agricultural Research Forum, Tullamore, Ireland. p75

Purfield, D.C., Bradley, D.G., Kearney, J.F., Berry, D.P. 2012. '*Genetic parameters for perinatal mortality in dairy cattle*', Proceedings of Agricultural Research Forum, Tullamore, Ireland p127

Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G. 2011. '*Analysis of Runs of Homozygosity in Cattle Breeds*', Proceedings of International Symposium of Animal Functional Genomics, p59

Purfield, D.C., Berry, D.P., Finlay, E., Mullen M.P., Howard, D.J & Bradley D.G. 2011. '*Population stratification among eight cattle breeds commonly used in Ireland*', Proceedings of Agricultural Research Forum, Tullamore, Ireland p68

Appendix A

A. Runs of homozygosity and population history in cattle

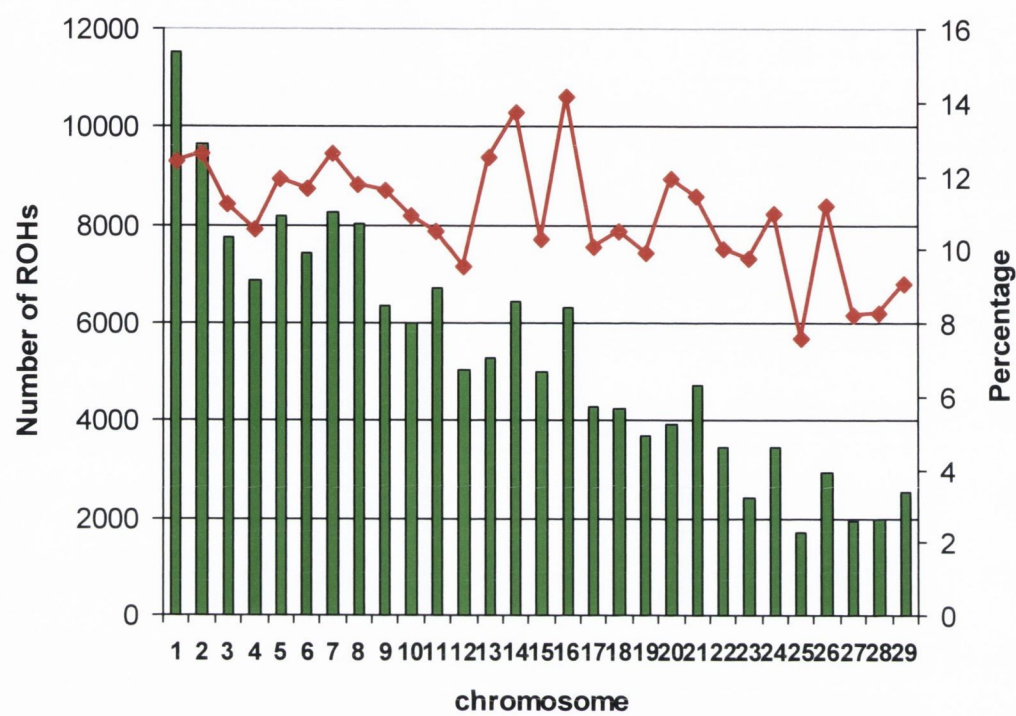
A. 1 Details of animals genotyped with the Bovine SNP50 density panel.

Code	Name of breed	Classification of origin	Number of animals	Data Source
ABO	Abondance	Italian and Eastern France	22	Gautier, Laloë et al. 2010
ANG	Angus	British Isles	62	Matukumalli, Lawley et al. 2009
AUB	Aubrac	Central and Southwest France	22	Gautier, Laloë et al. 2010
BPN	Bretonne Black Pied	Northern Europe	18	Gautier, Laloë et al. 2010
BRU	French Brown Swiss	Italian and Eastern France	18	Gautier, Laloë et al. 2010
BSW	Brown Swiss	Italian and Eastern France	24	Matukumalli, Lawley et al. 2009
CHA	Charolais	Central and Southwest France	20	Gautier, Laloë et al. 2010
CHL	Charolais	Central and Southwest France	26	Matukumalli, Lawley et al. 2009
GAS	Gascon	Central and Southwest France	22	Gautier, Laloë et al. 2010
GNS	Guernsey	Central and Southwest France	21	Matukumalli, Lawley et al. 2009
HFD	Hereford	British Isles	32	Matukumalli, Lawley et al. 2009
HOL	Holstein	Northern Europe	64	Matukumalli, Lawley et al.

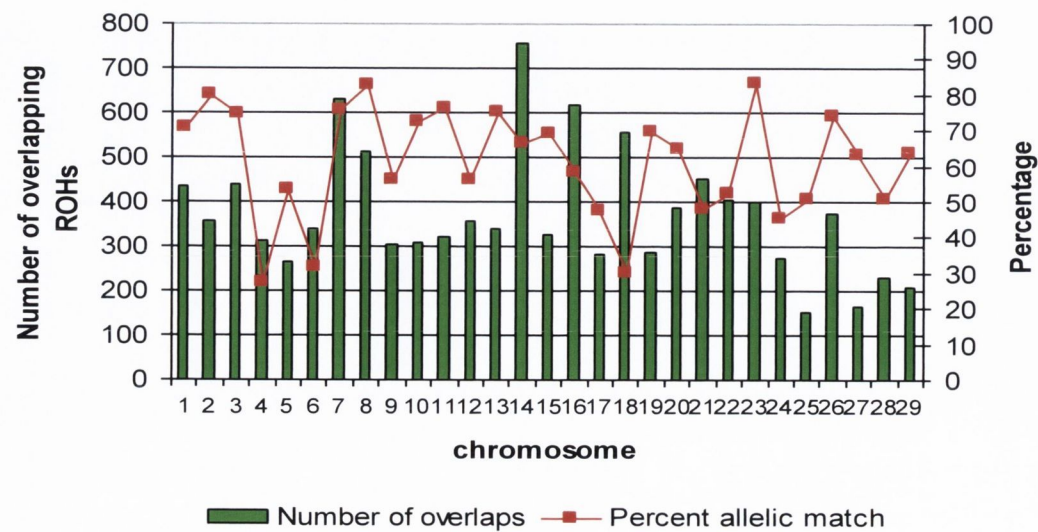
				2009
JER	Jersey	British Isles	28	Matukumalli, Lawley et al. 2009
LMS	Limousin	Central and Southwest France	45	Matukumalli, Lawley et al. 2009
MAN	Maine-Anjou (Rouge des Près)	Northern Europe	16	Gautier, Laloë et al. 2010
MAR	Maraichine (Parthenaise)	Northern Europe	19	Gautier, Laloë et al. 2010
MON	Montbeliard	Italian and Eastern France	30	Gautier, Laloë et al. 2010
NOR	Normande	Northern Europe	30	Gautier, Laloë et al. 2010
NRC	Norwegian Red Cattle	Northern Europe	21	Matukumalli, Lawley et al. 2009
PMT	Piedmontese	Italian and Eastern France	24	Matukumalli, Lawley et al. 2009
PRP	French Red Pied Lowland	Northern Europe	22	Matukumalli, Lawley et al. 2009
RGU	Red Angus	British Isles	15	Matukumalli, Lawley et al. 2009
RMG	Romagnola	Italian and Eastern France	24	Matukumalli, Lawley et al. 2009
SAL	Salers	Central and Southwest France	22	Gautier, Laloë et al. 2010
TAR	Tarine	Italian and Eastern France	18	Gautier, Laloë et al. 2010
VOS	Vosgienne	Italian and Eastern France	20	Gautier, Laloë et al. 2010

BMA	Beef Master	American and Indian zebu	24	Matukumalli, Lawley et al. 2009
SGT	Santa Gertrudis	American and Indian zebu	24	Matukumalli, Lawley et al. 2009
OUL	OulmèsZaer	Africa	27	Gautier, Laloë et al. 2010
BAO	Baoule	Africa	29	Gautier, Laloë et al. 2010
LAG	Lagune	Africa	30	Gautier, Laloë et al. 2010
NDA	N'Dama	Africa	55	Gautier, Laloë et al. 2010
SOM	Somba	Africa	31	Gautier, Laloë et al. 2010
BOR	Borgou	Africa	30	Gautier, Laloë et al. 2010
KUR	Kuri	Africa	30	Gautier, Laloë et al. 2010
ZBO	Zebu Bororo	Africa	23	Gautier, Laloë et al. 2010
ZFU	Zebu Fulani	Africa	30	Gautier, Laloë et al. 2010
SHK	Sheko	Africa	20	Matukumalli, Lawley et al. 2009
ZMA	Zebu from Madagascar	American and Indian zebu	30	Gautier, Laloë et al. 2010
BRM	Brahman	American and Indian zebu	25	Gautier, Laloë et al. 2010
GIR	Gir	American and Indian zebu	24	Gautier, Laloë et al. 2010
NEL	Nelore	American and Indian zebu	24	Gautier, Laloë et al. 2010

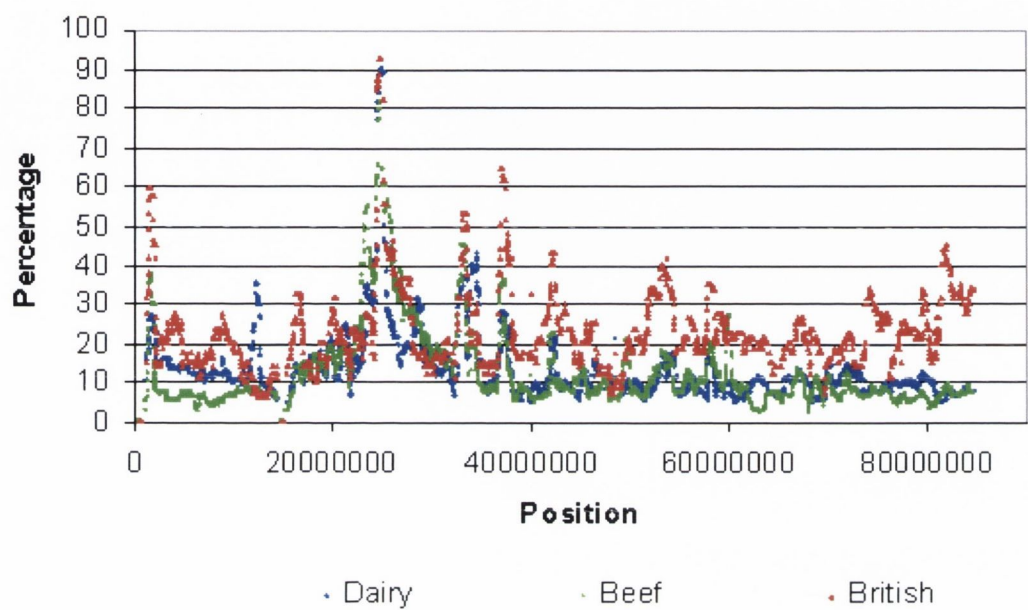
A.2 The number of runs of homozygosity (ROH), in the HD panel population, per chromosome (bars) and the mean per animal percentage coverage of the chromosome covered by ROH.



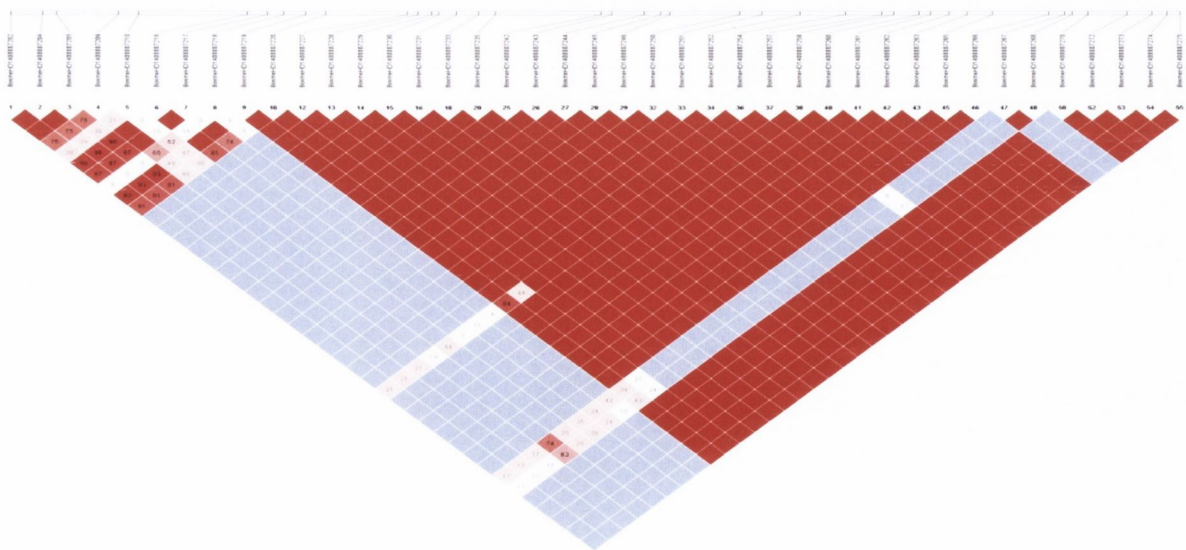
A.3 The maximum amount of overlapping runs of homozygosity (ROH) per chromosome and the percentage of overlapping ROHs that are a >95% allelic match.



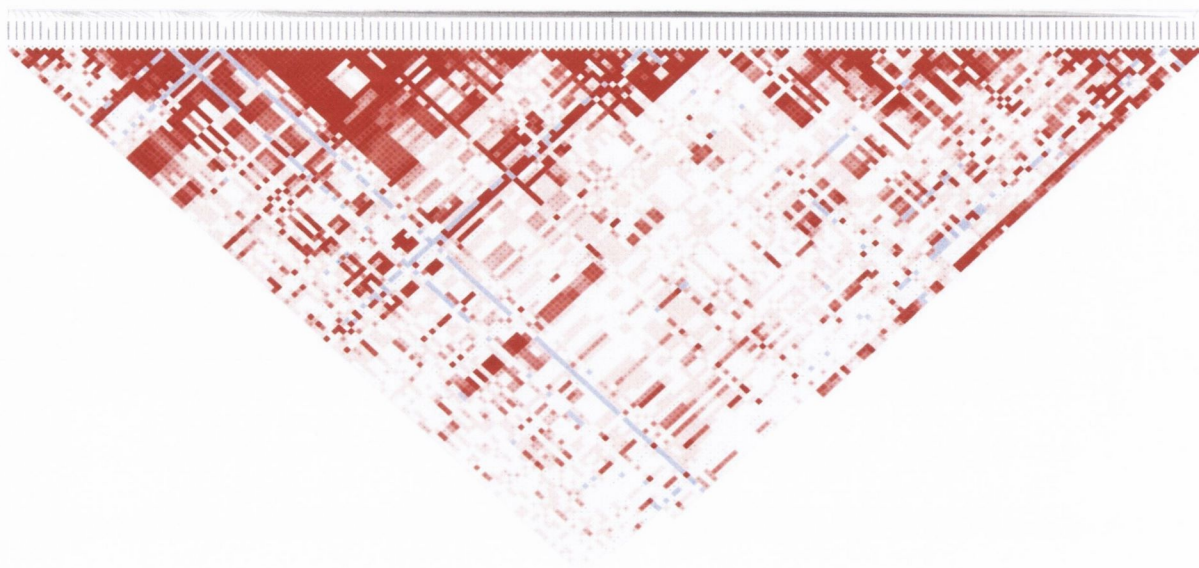
A.4 The proportion of appearances of each single nucleotide polymorphism on chromosome 14 in a ROH across dairy (Holstein, Friesian, Holstein-Friesian), beef (Charolais, Belgian Blue, Limousin, Simmental) and British (Angus, Hereford).



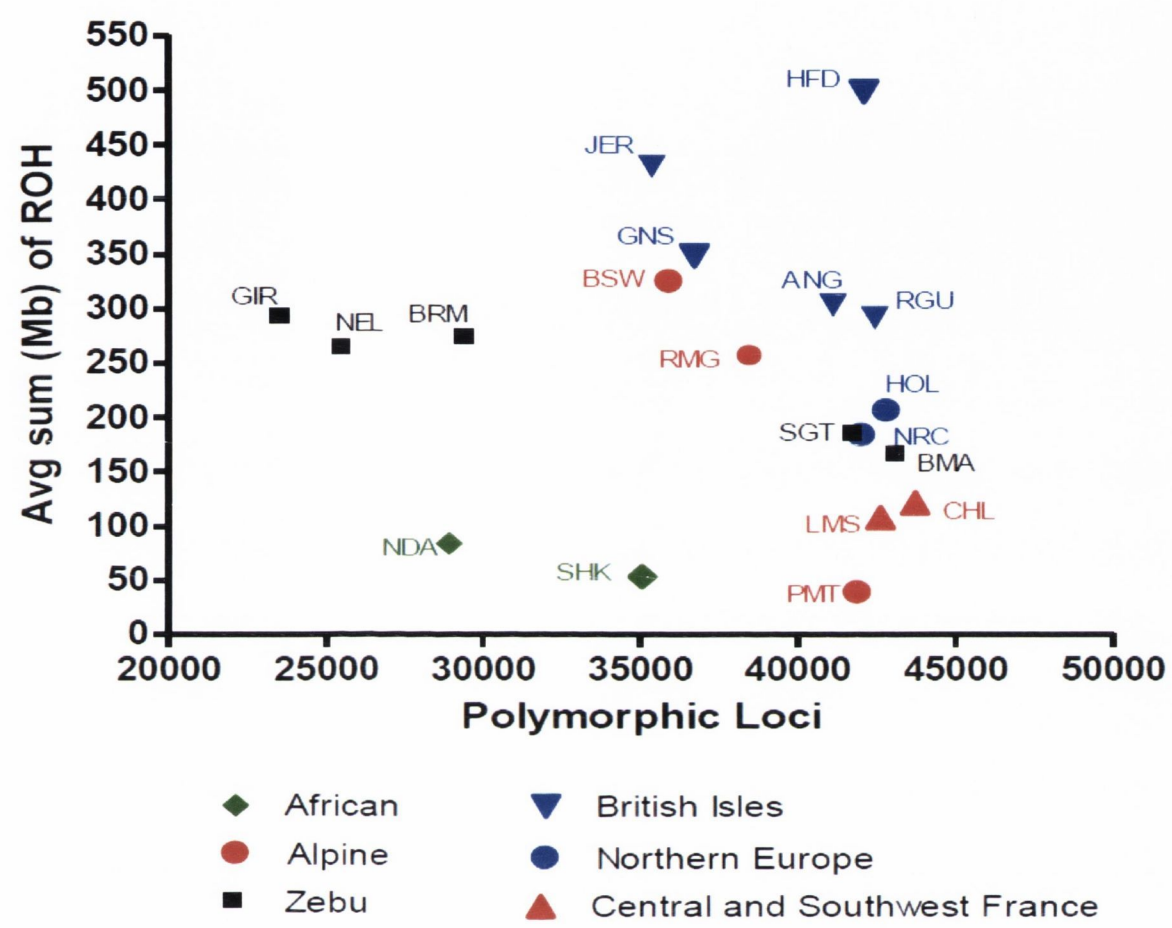
A.5 Linkage disequilibrium plot using Haploview of the consensus overlap segment that contains the TOX gene located on chromosome 14 in 86% of the runs of homozygosity.



A.6 Linkage disequilibrium plot using Haploview of SNPs on chromosome 12 that did not appear in a run of homozygosity.



A.7 Average total sum of Runs of Homozygosity (ROH) per animal within breed vs the number of validated polymorphic loci in the Bovine SNP50 genotyping beadchip graph.

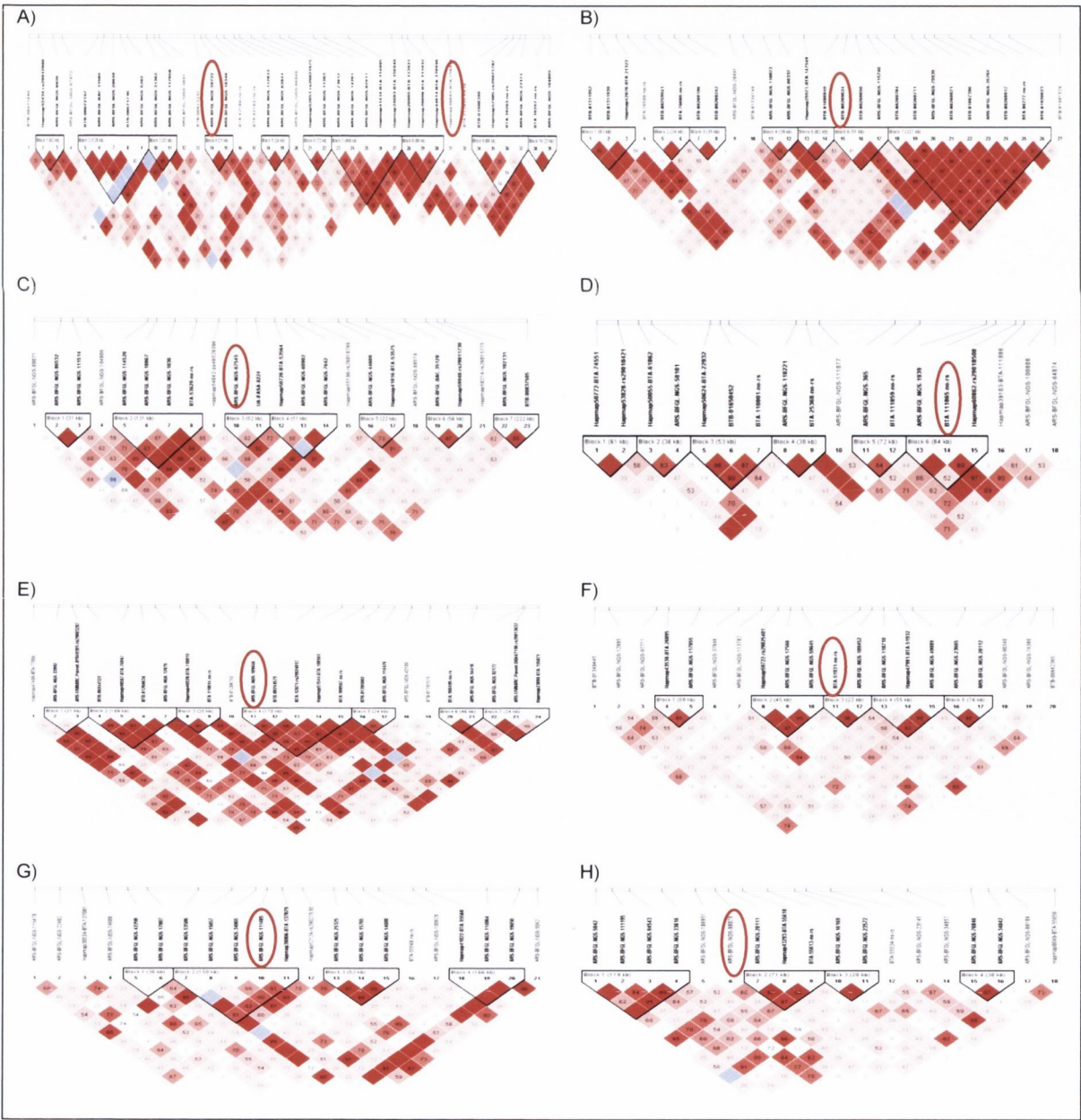


Appendix B

B. Genome-wide association study for calving traits in Holstein-Friesian dairy cattle

B.1 Haplotype analysis of 500kb up and downstream of remaining nine most associated SNPs for direct calving difficulty.

The SNP of interest is outlined in red. A) haplotype plot for ss117972188 and ss86289496 on BTA 17, B) haplotype plot for rs42803925 on BTA 6, C) haplotype plot for ss86297822 on BTA 22, D) haplotype plot for rs41575151 on BTA 5, E) haplotype plot for ss86284183 on BTA 6, F) haplotype plot for rs41602564 on BTA 16, G) haplotype plot for ss86339077 on BTA 23, H) haplotype plot for ss86328687 on BTA 23.



B.2 List of genes, their position and function identified within haploblocks containing the ten most associated SNPs for direct calving difficulty.

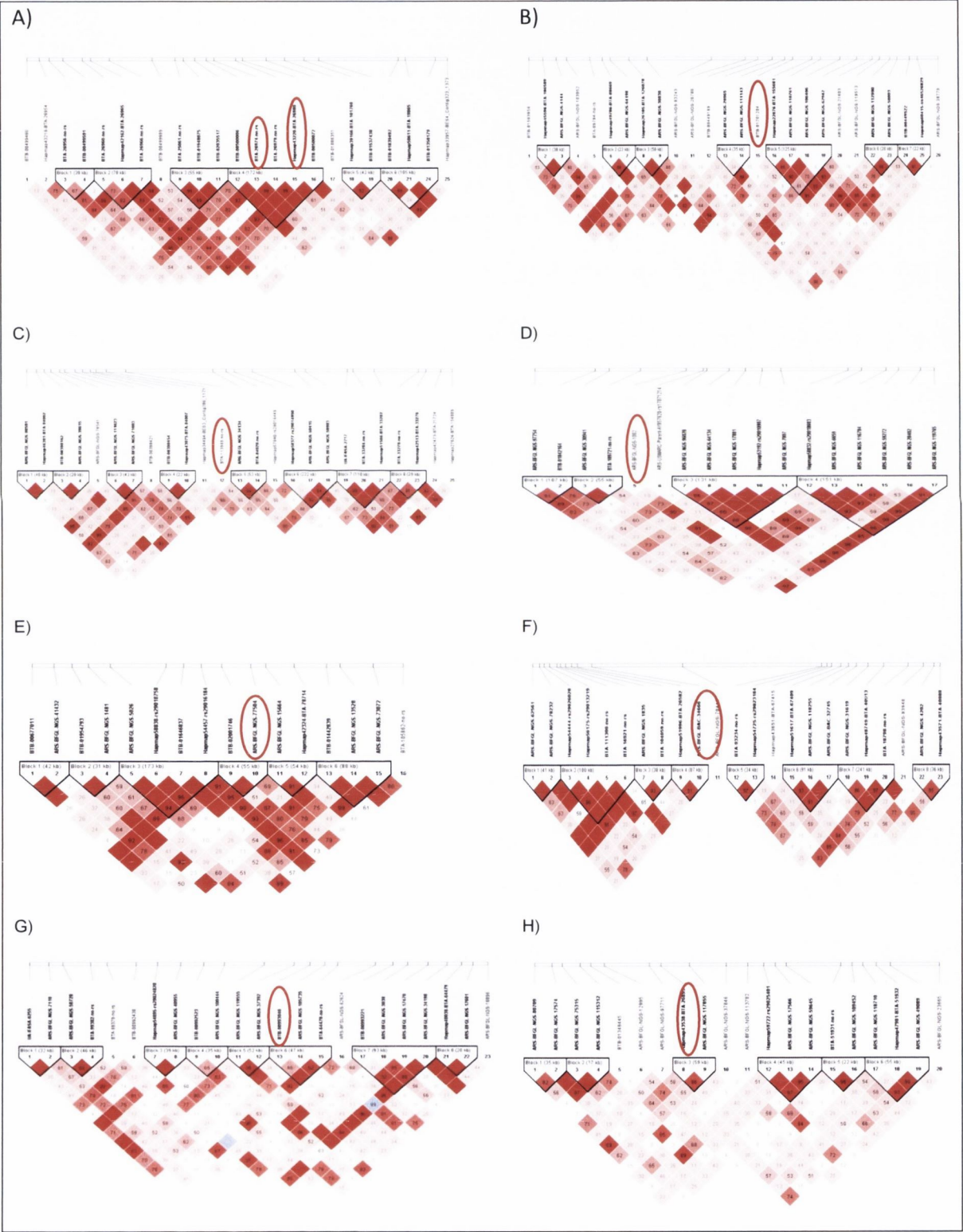
BTA	Haploblock position	Gene	Position	Function
18	56607357-56812286	SIGLEC-5	57,124,825-57,141,257	Also known as SIGLEC12. Sialic acid-binding immunoglobulin-like lectins (SIGLECs) are a family of cell surface proteins belonging to the immunoglobulin superfamily. They mediate protein-carbohydrate interactions by selectively binding to different sialic acid moieties present on glycolipids and glycoproteins.
		CD33	57,110,869-57,119,752	Putative adhesion molecule of myelomonocytic-derived cells that mediates sialic-acid dependent binding to cells.
		C18H19orf7	57,267,927-57,270,979	Member of sialic acid-binding immunoglobulin-like lectins
		IGLON5	57,308,171-57,313,994	Belongs to the immunoglobulin superfamily
		VSIG10L	57,318,113-57,328,795	V-set and immunoglobulin domain containing 10 like. V-set domains are Ig-like domains resembling the antibody variable domain. V-set domains are found in diverse protein families, including immunoglobulin light and heavy chains; in several T-cell receptors such as CD2 (Cluster of Differentiation 2), CD4, CD80, and CD86; in myelin membrane adhesion molecules; in junction adhesion molecules (JAM); in tyrosine-protein kinase receptors; and in the programmed cell death protein 1 (PD1).
		CLDND2	57,346,840-57,348,122	Tight junction forming protein
		NKG7	57,350,257-57,353,010	Natural killer cell protein 7. Found expressed in NK cells.
		ETFB	57,333,562-57,346,053	The electron transfer flavoprotein serves as a specific electron acceptor for several dehydrogenases, including five acyl-CoA dehydrogenases, glutaryl-CoA and sarcosine dehydrogenase. It transfers the electrons to the main mitochondrial respiratory chain via ETF-ubiquinone oxidoreductase (ETF dehydrogenase)
		LIM2	57,357,213-57,364,362	Present in the thicker 16-17 nm junctions of mammalian lens fiber cells, where it may contribute to cell junctional organization. Acts as a receptor for calmodulin. May play an important role in both lens development and cataractogenesis.
		SIGLEC10	57,372,841-57,379,872	Member of the sialic acid-binding immunoglobulin-like lectins
		ZNF175	57,437,817-57,447,909	Down-regulates the expression of several chemokine receptors.
		SIGLEC14	57,562,649-57,566,637	Member of the sialic acid-binding immunoglobulin-like lectins
		HAS1	57,601,946-57,609,339	Hyaluronan or hyaluronic acid (HA) is a high molecular weight unbranched

				polysaccharide. It has variety of functions, including space filling, lubrication of joints, and providing of a matrix through which cells can migrate. HA is actively produced during wound healing and tissue repair to provide a framework for in growth of blood vessels and fibroblasts.
		ZNF577	57,659,498-57,672,731	Zinc finger and may be involved in transcriptional regulation
		ZNF432	57,720,462-57,728,282	Zinc finger and may be involved in transcriptional regulation
17	No haplobock	N/A	N/A	N/A
6	66199733-668099119	NONE	N/A	N/A
22	12830183-12985630	CCR8	12,835,708-12,836,244	This gene encodes a member of the beta chemokine receptor family, which is predicted to be a seven transmembrane protein similar to G protein-coupled receptors. It has been suggested to have a role in regulation of monocyte chemotaxis and thymic cell apoptosis. More specifically, this receptor may contribute to the proper positioning of activated T cells within the antigenic challenge sites and specialized areas of lymphoid tissues.
		THEM4	12,838,604-12,844,384	The protein encoded by this gene negatively regulates protein kinase B activity by inhibiting phosphorylation
		SLC25A38	12,846,582-12,861,550	Mitochondrial carrier required during erythropoiesis. Probably involved in the biosynthesis of heme, possibly by facilitating 5-aminolevulinate (ALA) production. May act by importing glycine into mitochondria or by exchanging glycine for ALA across the mitochondrial inner membrane
		RPSA	12,885,044-12,898,466	Required for the assembly and/or stability of the 40S ribosomal subunit. Required for the processing of the 20S rRNA-precursor to mature 18S rRNA in a late step of the maturation of 40S ribosomal subunits. Also functions as a cell surface receptor for laminin. Plays a role in cell adhesion to the basement membrane and in the consequent activation of signaling transduction pathways. May play a role in cell fate determination and tissue morphogenesis
		LOC515736	12,863,845-12,881,080	Possibly encodes the enzyme caspase 15. Many of these enzymes are part of a proteolytic cascade that plays a central role in cell death by apoptosis.
		MOBP	12,924,547-12,956,464	May play a role in compacting or stabilizing the myelin sheath, possibly by binding the negatively charged acidic phospholipids of the cytoplasmic membrane
5	96358013-96966605	No genes	N/A	N/A
6	13429879-13649338	ANK2	13,341,234-13,554,826	This gene encodes a member of the ankyrin family of proteins that link the integral membrane proteins to the underlying spectrin-actin cytoskeleton. Ankyrins are involved in cell motility, activation, proliferation, contact and the maintenance of

				specialized membrane domains. In humans the protein encoded by this gene is required for targeting and stability of Na/Ca exchanger 1 in cardiomyocytes.
16	73071792- 73506712	PLXNA2	73,453,456- 73,658,974	This gene encodes a member of the plexin-A family of semaphorin co-receptors. Semaphorins are a large family of secreted or membrane-bound proteins that mediate repulsive effects on axon pathfinding during nervous system development. A subset of semaphorins are recognized by plexin-A/neuropilintransmembrane receptor complexes, triggering a cellular signal transduction cascade that leads to axon repulsion.
17	13327070- 13349026	NONE	N/A	N/A
23	15228064- 15504926	NONE	N/A	N/A
23	No haploblock	N/A	N/A	N/A

B.3 Haplotype analysis of 500kb up and downstream of remaining nine most associated SNPs for maternal calving difficulty.

The SNP of interest is outlined in red. A) haplotype plot for ss61478550 and rs41628739 on BTA 12,B) haplotype plot for ss86336751 on BTA 11, C) haplotype plot for rs41612473 on BTA 9 D) haplotype plot for ss86274519 on BTA 16, E) haplotype plot for ss86307824 on BTA 17, F) haplotype plot for ss105262945 on BTA 17, G) haplotype plot for rs42151848 on BTA 28, H) haplotype plot for RS41576933 on BTA 16.

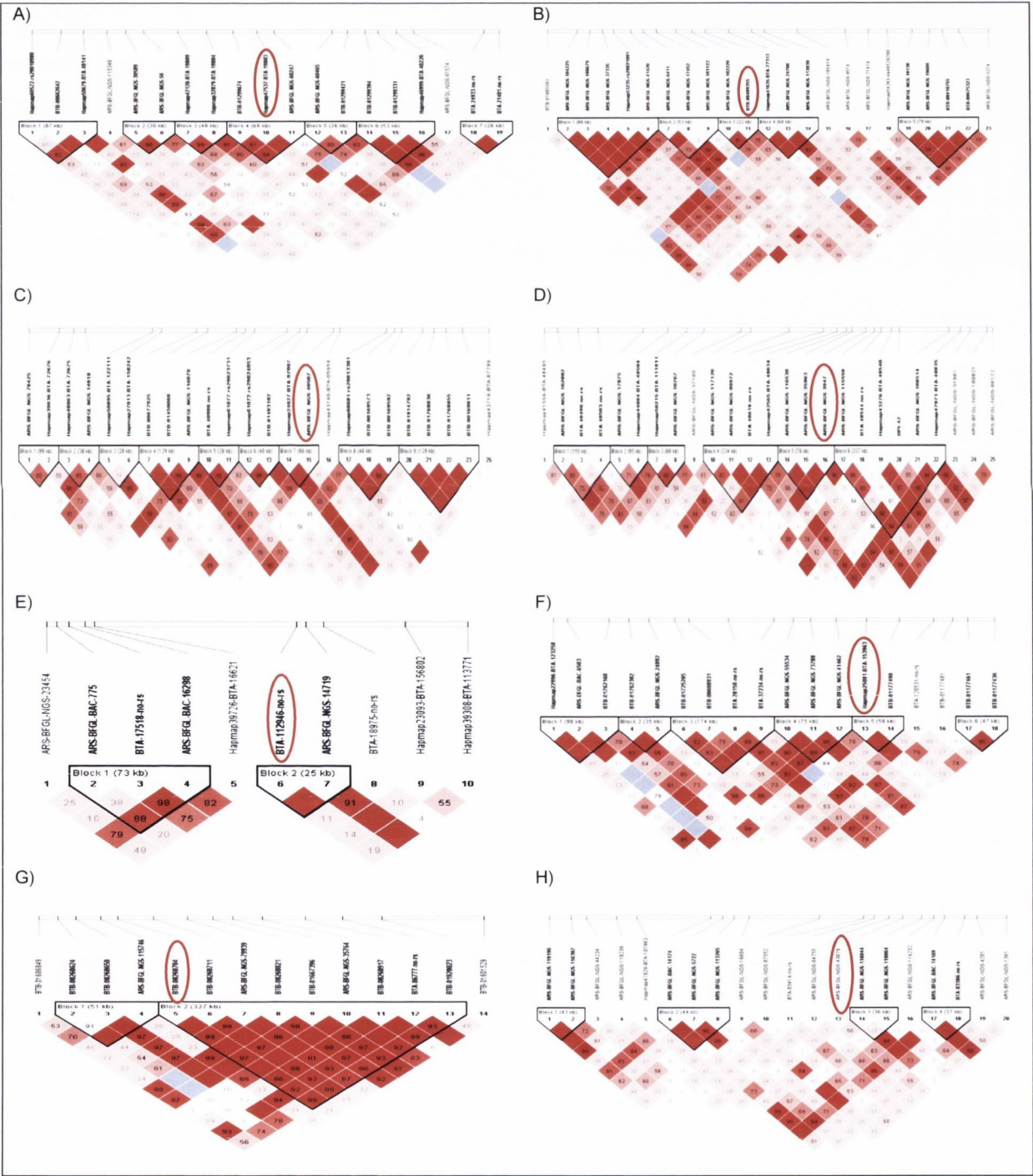


B.4 List of genes, their position and function identified within haploblocks containing the ten most associated SNPs for maternal calving difficulty.

BTA	Haploblock position	Gene	Position	Function
28	No haploblock	N/A	N/A	N/A
12	57164786-57253923	NONE	N/A	N/A
11	846983-882225	NONE	N/A	N/A
9	No haploblock	N/A	N/A	N/A
16	No haploblock	N/A	N/A	N/A
17	45656849-46310121	ACCN5	45,812,197-45,824,166	Acid-sensing (proton-gated) ion channel family member 5.
		GUCY1B3	45,815,868-45,857,540	GUCY1B3 catalyzes the conversion of GTP to the second messenger cGMP and functions as the main receptor for nitric oxide (NO) and nitrovasodilator drugs
		CTSO	45,915,251-45,939,356	Cathepsin O is a cysteine protease and a member of the cathepsin family and papain superfamily.
		TDO2	45,944,647-45,961,246	Incorporates oxygen into the indole moiety of tryptophan.
17	39145131-39505881	GUCY1A1	46,088,197-46,153,264	The protein encoded by this gene is an alpha subunit of the soluble guanylatecyclases and it interacts with a beta subunit to form the guanylatecyclase enzyme, which is activated by nitric oxide.
		NONE	N/A	N/A
		PPYR1	41,791,184-41,792,535	Pancreatic polypeptide receptor 1
		ANXA8	41,833,544-41,849,860	Anxa8 is an anticoagulant protein that acts as an indirect inhibitor of the thromboplastin-specific complex, which is involved in the blood coagulation cascade
16	72861402-72920777	NONE	N/A	N/A

B.5 Haplotype analysis of 500kb up and downstream of remaining nine most associated SNPs for perinatal mortality.

The SNP of interest is outlined in red. A) haplotype plot for rs41627212 on BTA 16,B) haplotype plot for rs43616983 on BTA 10, C) haplotype plot for ss86316303 on BTA 4D) haplotype plot for ss86303585 on BTA 2, E) haplotype plot for rs416117236 on BTA 12, F) haplotype plot for ss86315311 on BTA 15, G) haplotype plot for rs43469091 on BTA 16, H) haplotype plot for ss86299461 on BTA 10. The SNP ss86324977 on BTA 18 is detailed on Figure 4.2.



B.6 List of genes, their position and function identified within haploblocks containing the ten most associated SNPs for perinatal mortality.

BTA	Haploblock position	Gene	Position	Function
14	71356288-72195341	SLC26A7	71,722,604-71,871,605	This gene is one member of a family of sulfate/anion transporter genes.
		TMEM55A	72,033,027-72,105,636	Catalyzes the hydrolysis of the 4-position phosphate of phosphatidylinositol 4,5-bisphosphate.
		TMEM64	72,195,056-72,229,443	Transmembrane protein 64 is a regulator of Ca(2+) oscillation during osteoclastogenesis.
16	74717247-75179403	DENND1B	74,792,846-74,895,655	A members of the connecdenn family, which function as guanine nucleotide exchange factors (GEFs) for the early endosomal small GTPase RAB35 (MIM 604199) and bind to clathrin and clathrin adaptor protein-2.
10	12585173-12899363	RAB11A	12,568,783-12,588,033	Is a small GTPaseRab are key regulators of intracellular membrane trafficking, from the formation of transport vesicles to their fusion with membranes. In humans Rab11a was involved in the recruitment of TLR4 to phagosomes in a process requiring TLR4 signaling.
		MEGF11	12,599,062-12,690,387	In humans MEGF11 is thought to regulate the mosaic spacing of specific neuron subtypes in the retina. Mosaics provide a mechanism to distribute each cell type evenly across the retina, ensuring that all parts of the visual field have access to a full set of processing elements.
4	24191610-25183065	AGMO	24,807,121-25,004,482	The protein encoded by this AGMO is a tetrahydrobiopterin- and iron-dependent enzyme that cleaves the ether bond of alkylglycerols. Variations in this gene have been associated with decreased glucose-stimulated insulin response, type 2 diabetes, and susceptibility to intracranial aneurysms in humans.
		MEOX2	25,059,437-25,135,467	Meox2 encodes a member of a subfamily of non-clustered, diverged, antennapedia-like homeobox-containing genes. The encoded protein may play a role in the regulation of vertebrate limb myogenesis. Mutations in the related mouse protein may be associated with craniofacial and/or skeletal abnormalities
2	108276539-109093402	MARCH4	108,258,480-	MARCH4 is a member of the MARCH family of membrane-bound E3 ubiquitin

			108,284,436	ligases. MARCH enzymes add ubiquitin to target lysines in substrate proteins, thereby signalling their vesicular transport between membrane compartments. MARCH4 reduces surface accumulation of several membrane glycoproteins by directing them to the endosomal compartment
		SMARCAL1	108,615,915-108,666,059	ATP-dependent annealing helicase that catalyzes the rewinding of the stably unwound DNA. Rewinds single-stranded DNA bubbles that are stably bound by replication protein A (RPA). Acts throughout the genome to reanneal stably unwound DNA, performing the opposite reaction of many enzymes, such as helicases and polymerases that unwind DNA.
		RPL37A	108,679,858-108,682,477	RPL37A encodes a ribosomal protein that is a component of the 60S subunit of the ribosome.
		IGFBP2	108,810,787-108,839,273	Inhibits IGF-mediated growth and developmental rates by similarity. IGF-binding proteins prolong the half-life of the IGFs and have been shown to either inhibit or stimulate the growth promoting effects of the IGFs on cell culture. They alter the interaction of IGFs with their cell surface receptors.
		IGFBP5	108,851,172-108,854,511	Same function as IGFBP2.
12	6281991-6307016	NONE	N/A	N/A
15	60424410-61035044	KCNA4	60,436,517-60,443,987	Mediates the voltage-dependent potassium ion permeability of excitable membranes. The protein forms a potassium-selective channel through which potassium ions may pass in accordance with their electrochemical gradient.
		LOC787432	60,620,624-60,626,625	Follicle stimulating hormone that regulates the development, growth, pubertal maturation, and reproductive processes of the body.
		FSHB	60,683,262-60,687,228	Stimulates development of follicle and spermatogenesis in the reproductive organs.
		ARL14EP	60,753,945-60,767,727	Through its interaction with ARL14 and MYO1E, ARL14EP may connect MHC class II-containing cytoplasmic vesicles to the actin network allowing it to control the

				movement of these vesicles along the actin cytoskeleton in dendritic cells
		MPPED2	60,867,597-61,077,511	MPPED2 likely encodes a metallophosphoesterase. The encoded protein may play a role a brain development.
6	66407840-66809119	None	N/A	N/A
10	No haploblock	N/A	N/A	N/A

