POPULATION LEVEL GENOME-WIDE ASSOCIATION STUDIES IN DAIRY CATTLE

XIAOWEI MAO

PhD THESIS · SCIENCE AND TECHNOLOGY · 2016

AARHUS UNIVERSITY

SLU

Department of Molecular Biology and Genetics
Center for Quantitative Genetics and Genomics
Blichers Allé 20
8830 Tjele
Denmark

Department of Animal Breeding and Genetics
Swedish University of Agricultural Sciences
75007 Uppsala
Sweden

Tryk: www.digisource.dk
ISBN: 978-87-93398-53-5
ISSN: 1652-6880
ISBN (printed version): 978-91-576-8743-2
ISBN (electronic version): 978-91-576-8744-9

Acta Universitatis agriculturae Sueciae
2016:120

Cover: Limin Wang (Drawing); Xiaowei Mao (Designed)

The research presented in this thesis was conducted under the joint auspices of Aarhus University and the Swedish University of Agricultural Sciences and is part of the Erasmus Mundus Joint Doctorate Program “EGS-ABG”.

Population level genome-wide association studies in dairy cattle

Xiaowei Mao

PhD THESIS

Main supervisors: Bernt Guldbrandtsen and Dirk-Jan De Koning

Department of Molecular Biology and Genetics

Center for Quantitative Genetics and Genomics

Aarhus University

Blichers Allé 20, 8830 Tjele, Denmark

Department of Animal Breeding and Genetics,

Swedish University of Agricultural Sciences

75007 Uppsala Sweden

Dedication

Dedicated to my grandfather who is always in my heart

& my grandmother for her endless love

List of Publications

This thesis is based on the work contained in the following papers:

1. Mao, X., N. K. Kadri, J. R. Thomasen, D.J. De Koning, G. Sahana, and B. Guldbrandtsen. 2015. Fine mapping of a calving QTL on *Bos taurus* autosome 18 in Holstein cattle. J. Anim. Breed. Genet. 133:207–218.

2. Mao, X., G. Sahana, D.J. De Koning, and B. Guldbrandtsen. 2016. Genome-wide association studies of growth traits in three dairy cattle breeds using whole-genome sequence data. J. Anim. Sci. 94:1426.

3. Mao, X., A. M. Johansson, G. Sahana, B. Guldbrandtsen, and D.J. de Koning. 2016. Imputation of markers on the Bovine X chromosome. J. Dairy Sci. 833:1–6.

4. Mao, X., G. Sahana, A. M. Johansson, A. Ismael, P. Løvendahl, D.J. de Koning, and B. Guldbrandtsen. 2016. Genome-wide association mapping for dominance effects in female fertility using real and simulated data from Danish Holstein cattle. (Manuscript)

Table of Contents

Acknowledgement

This PhD study was performed in the Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University (AU), Denmark and in the Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences (SLU), Sweden. I am enrolled and funded by the Erasmus-Mundus joint doctorate European Graduate School in Animal Breeding and Genetics (EGS-ABG).

If my longevity is luckily and optimistically 100 years, this four-year experience is 1/25 of my life, not too long, not too short. I prefer to call this experience "修行 (self-knowing and expressing)" rather than study or work. I have learned not only through the courses, literatures but also through various ups and downs and open-up times with many people.

Bernt Guldbrandtsen, my main supervisor in AU. I had really difficult times googling your pictures before applying for this PhD program. Besides the curiosity for science, the curiosity for meeting the real people partially helps me on board of this project. Without your kind encouragements and endless help on coding and Linux cluster, this project would be a tremendous task. Apart from work, sometimes even during the work, we could talk about random things for hours as friends. I still remember your kindness of introducing me what Marzipan is and all kinds of Wikipedia type of storytelling. More importantly, instead of pushing me to be a paper machine, you guided me to be an independent thinker.

Dirk-Jan De Koning, my main supervisor in SLU. The warm greeting from you at SLU encouraged me to be quickly involved in this group. Your "Fika time" jokes and easy-going personality tie the quantitative group together nicely. Your efficient project and time management skills also inspired me to be both flexible and productive during the stay in SLU.

Goutam Sahana, my co-supervisor in AU. My rookie period in the journey of gene mapping would not be that smooth without your intuitively clear explanations and tons of help for my analyses. Your responses for academic help are always on time and straightforward.

Anna Maria Johansson, my co-supervisor in SLU. You spared the time to join my meetings and discussions, even though having to take care of your newborn babies. Your valuable input also encouraged me to finish my third project.

Karin Smedegaard, my secretary in AU. You are the most responsible and organized person I have ever met, perfectly balanced between being strictly professional and being personal. There is no chance for me to say thank you face to face. May your soul rest in heaven!

I would like to thank Thu Hong Le, Solomon Antwi Boison, Merina Shrestha, Liesbeth François, Agnese Viluma and Christos Dadousis for the firm friendship ever since the European Master in Animal Breeding and Genetics program six years ago. It is a privilege to know you, especially Thu and Solomon, with whom I feel no cultural barriers and often share deep feelings.

My life in Foulum in Denmark would not be that colorful without Naveen Kadri, Mai Anh Ton Nu, Nuzul Widyas, Smitha Anantharama Ankinakatte, who bore with my complains and warmly involved me in this international group at the beginning of the PhD; Setegn Worku Alemu, Ashutosh Das, Ahmed Ismael Sayed Ismael, Dana Olijhoek, Irene van den Berg, Hadi Esfandyari, Bashar Amer, with whom I enjoyed the chatting, laughing, travelling, nature walking, swimming and sauna competitions; Louise Dybdahl Pedersen and Elise Norberg, who helped a lot with EGS-ABG and other monetary issues.

My life in Uppsala in Sweden would be left with only analyses and writing papers without the PhD "Mafia" group: Chrissy Rochus, Bingjie Li, Shizhi Wang, Sandrine Duchemin, Berihu Welderufael André M. Hidalgo, Nancy Amuzu, Merina Shrestha, Agnese Viluma, Thu Hong Le, Juan Cordero, Amabel Tenghe. You remind me of various music and movie nights, and the famous Flogsta Roar. I also enjoyed nice conversations and jokes with Freddy Fikse, in addition to your academic help.

Without the Chinese community in Foulum and Uppsala, I would not be able to refill myself with energy and passion to face the challenges in life abroad. You make me feel home! I specially say thank you to Aoxing, Zhencai, Yingqian for your selfless support and encouragement during this four-year PhD.

My Mama Angela Ewing from Brisbane and Auntie Ping Ren from Uppsala, thank you for keeping me in faith and helping me to grow regardless of my stubbornness.

Above all, I cannot finish without the love from my families.

Xiaowei Mao

September 2016 in Foulum Denmark

Summary

In recent years, genome-wide association studies (GWAS) has become a dominant tool for detecting genetic architectures for complex traits. Thousands of associated genetic variants have been reported. However, the resolution of these studies was limited by the available marker density for the quantitative trait loci (QTL) region. Moreover, the X chromosome and non-additive genetic effects have often been excluded from GWAS, despite of their potentially important biological functions. This thesis carried out the finer mapping of functional (calving and female fertility) and production (growth) traits in dairy cattle utilizing high-density SNP chip (HD) and imputed whole-genome sequence (WGS) data, and explored the genotype imputation of the X chromosome and the mapping of variants exhibiting dominance effects for female fertility.

In **chapter 2**, fine-mapping of a previously reported QTL in Holstein cattle on *Bos taurus* autosome 18 (BTA18) for calving traits was performed, using imputed HD genotypes followed by imputed WGS variants. BTA18 was analyzed for seven direct calving traits in 6,113 bulls with imputed HD genotypes. Single nucleotide polymorphism (SNP) rs136283363 (BTA18:57,548,213) was consistently the most significantly associated SNP across all seven traits. Then WGS variants within the targeted QTL region were tested for associations with direct calving traits and with three conformation traits. Genes *SIGLEC12*, *CD33* and *CEACAM18* were proposed as candidate genes. In addition, pleiotropic effects of this QTL region were observed on direct calving traits and conformation traits. However, the extent of linkage disequilibrium (LD), lack of complete annotation and potential errors in the *Bos taurus* genome assembly hampered our efforts to pinpoint the causal mutation.

In **chapter 3**, we performed a GWAS for growth traits in Nordic Holstein, Jersey, and Red Dairy Cattle. First, GWAS was performed within breeds using WGS variants. Then a meta-analysis was performed to combine information across these three breeds. Several QTL were identified to have large effects on growth traits in Holstein and Red Dairy Cattle, but only one QTL located nearby gene *CYP19A1* on chromosome 10 was shared between Holstein and Red Dairy Cattle. Another QTL near 25 Mb on chromosome 14 was very significantly associated with growth traits in Red Dairy Cattle, consistent with the previously reported gene *PLAG1*, which affect growth in beef cattle and humans. No QTL was statistically significant in Jersey, which might be due to the low power of detection with the small sample size. Meta-analysis of the three breeds enhanced the power to detect QTL.

In **chapter 4**, we performed the imputation of markers on the X chromosome in Holstein cattle for non-genotyped animals and animals genotyped with low density (Illumina BovineLD) chips, using animals genotyped with medium density (Illumina BovineSNP50) chips. The program *FImpute V2.2* and genotypes of 26,884 Holstein individuals genotyped with medium density chips were used in this study. We found that the imputation accuracy of markers on the X chromosome was improved by treating the pseudo-autosomal region (PAR) as autosomal and by increasing the proportion of females in the reference group. We also found imputation for non-genotyped animals in general had lower accuracy compared to animals genotyped with the low density SNP array. Besides, higher cumulative pedigree relationships between the reference group and the target animals led to higher imputation accuracy. Better marker coverage of the X chromosome should be developed to facilitate genomic studies involving the X chromosome in future studies.

In **chapter 5**, we aimed to detect dominance effects on female fertility traits in Danish Holstein cattle using Illumina BovineSNP50 data, and evaluate the power, precision, and type 1 error of detecting dominance effects through simulations. Female fertility data (number of inseminations, days from calving to first insemination, and days from the first to last insemination) were recorded from 3,040 genotyped heifers and 4,483 genotyped cows from Danish Holstein population. Firstly, the additive and dominance genetic variances were quantified using GBLUP for the fertility traits. Secondly, the association analyses for heifers and cows were performed separately. In the end, a simulation study was carried out to test the power, precision and type 1 error rate for detection of dominance effects. The dominance genetic variance was larger than additive genetic variance in heifers, but had a large standard error. Four QTL were detected for IFL in heifers, while one QTL was detected for cows. All these five QTL were detected with significant additive and dominance effects. Simulations showed that the current sample size had limited power to detect dominance effects for female fertility in cattle. In the future, more females need to be genotyped and/or imputed to map the genetic variants with dominance effects on female fertility traits.

Sammendrag

I de senest år er genomiske associationsstudier blevet det fremherskende værktøj til at under søge komplekse egenskabers genetiske arkitektur. Tusinder af associerede, genetiske varianter er blevet rapporteret. Desværre er disse studiers præcision med hensyn til stedfæstelsen af de identificere kvantitative egnskabsloci (*quantitative trait loci*, QTL) begrænset af den tilgængelige markørtæthed. Desuden er X-bundne og ikke-additive genetiske effekter ofte blevet udelukket fra genomiske associationsstudier på trods af at deres potentielt vigitige, biologiske funktioner. Denne afhandling beskriver finkortlægning af funktionelle (kælvning og hunlig frugtbarhed) og produktions- (tilvækst) egenskaber i malkekvæg under udnyttelse af SNP-chip med stor tæthed (*high density*, HD) og imputeret helgenomskevens (whole genome sequence, WGS) data, og udforsker imputation af markører på X-kromosomet, samt kortlægning af varianter, som udviser dominanseffekter for hunlig frugtbarhed.

I kapitel 2 udførtes finkortlægning af et tidligere rapporteret QTL i *Bos taurus* autosom 18 (BTA18), som påvirker kælvningsegenskaber. Dette sker ved hjælp af HD-genotyper imputeret til WGS-niveau. BTA18 blev undersøgt for 7 direkte kælvningsegenskaber for 6.113 tyre med HD-genotyper. SNP rs136283363 (BTA18:57,548,213) var konsekvent den mest signifikant associerede SNP på tværs af de 7 egenskaber. Herefter undersøgtes WGS-varianter i det undersøgte QTL-område for association med kælvningsegenskaber og eksteriøregenskaber. Gener *SIGLEC12*, *CD33* and *CEACAM18* foreslås som kandidatergener. Desuden observeredes pleiotrope effekter på kælvnings- og eksteriøregenskaber. Imidlertid gjorde omfattende koblingsuligevægt, mangelfulde annotationer af genomet og potentielle fejl i *Bos taurus*-genomassembly det vanskeligt med sikkerhed at identificere den specifikke variant, som var årsagen.

I kapitel 3 udførte vi en GWAS for tilvækst i Holstein, Jersey og Rødt Malkekvæg. If første omgang der en GWAS indenfor hver race. Herefter gennemførtes en meta-analyse, som kombinerede informationen på tværs af de tre racer. Adskillige QTL med store effekter på tilvækstegenskaber identificeredes i Holstein og Rødt Malkekvæg, men kun et QTL tæt på *CYP19A1* på kromosome 10 deltes mellem Holstein og Rødt Malkekvæg. Et andet QTL omkring 25 Mb på kromosom 14 var signifikant associeret med tilvækst i Rødt Malkekvæg, hvilket er i overensstemmelse med tidligere rapporter om genet *PLAG1*, som påvirker tilvækst i kødvæg og i mennesker. Der var ingen signifikante QTL i Jersey, hvilket muligvis skyldes lav statistisk styrke

på grund af den mindre stikprøvestørrelse. En meta-analyse af de tre racer øgede den statistiske styrke til at finde QTL.

I kapitel 4 udførte vi en imputation af markører på X-kromosomet i Holsteinkvæg for ikke-genotypede dyr og dyr, som er genotypet chips med lav tæthed (Illumina BovineLD eller LD) ved hjælp af dyr, som var genotypet med mellemtætheds (Illumina BovineSNP50 eller 50k) chips. Programmet *FImpute V2.2* og genotyper for 26.884 Holsteinindivider genotypet med 50k chips blev benyttet hertil. Vi fandt, at imputationssikkerheden forbedredes ved at behandle den pseudoautosomale region som autosomal og ved at forøge andelen af hundyr i referencegruppen. Vi fandt også, at imputation for dyr uden genotyper generelt var lavere end for dyr med LD-genotyper. En større kumulativ slægtskabsgrad til referencegruppen medførte en højere imputationssikkerhed. En bedre markørdækning burde udvikles for at understøtte fremtidige genomiske studier af X-kromosomet.

I kapitel 5 bestræbte vi os på at finde dominanseffekter for hunlige frugtbarhedsegenskaber i danske Holsteinkvæg på baggrund af Illumina BovineSNP50 data, og at bedømme statistisk styrke, præcision og type 1 fejl i sporingen af dominanseffekter ved hjælp af simulationer. Hunlige frugtbarhedsdata (antal inseminationer, NINS, dage fra kælvning til første insemination, ICF, og dage fra første til sidste insemination, IFL) var registreret for 3.040 kvier og 4.483 køer med 50k-genotyper i den danske Holsteinpopulation. Først estimerede vi den additive genetiske varians og dominansvariansen for frugtbarhedsegenskaber ved hjælp af GBLUP. Dernæst gennemførtes associationsstudier særskilt for hhv. kvier og køer. Til slut gennemførtes et simulationsstudie for at undersøge den statistiske styrke, præcisionen og type 1-fejlraten ved detektionen af dominanseffekter. Dominansvariansen var større end den additive varians i kvier med havde store estimationsusikkerheder. Fire QTL kunne påvises for IFL i kvier, mens et QTL kunne påvises i køer. Alle disse fem QTL havde både additive effekter og dominanseffekter. Simulationer viste, at den aktuelle stikprøvestørrelse kun giver begrænset styrke til at påvise dominanseffekter. For at kunne påvise dominanseffekter for hunlig frugtbarhed, så må der i fremtiden genotypes eller imputeres flere hundyr.

Sammanfattning

På senare år har GWAS (associationsstudier med hjälp av markörer spridda över hela genomet) blivit det dominerande redskapet för att beskriva den genetiska bakgrunden till komplexa egenskaper. Tusentals associerade genetiska varianter har rapporterats. Dock har dessa studiers förmåga att identifiera loci för kvantitativa egenskaper (QTL) begränsats av den tillgängliga markörtätheten. Dessutom har X-kromosomen och icke-additiva genetiska effekter ofta uteslutits från GWAS trots deras potentiellt viktiga biologiska funktioner. Denna avhandling har genomfört finare kartläggning av funktionella egenskaper (kalvning och honlig fortplantningsförmåga) och produktion (tillväxt) i mjölkkor genom att utnyttjar SNP-chip med hög densitet (HD) och imputerad helgenomsekvens (WGS) data. Imputering innebär att man beräknar sannolika genotyper för loci som ej genotypats. Avhandlingen har utforskat imputering av genotyper på X-kromosomen och kartläggning av varianter som uppvisar dominanseffekter för honlig fertilitet.

I kapitel 2 genomfördes en kartläggning av en rapporterad QTL på Bos taurus autosom (BTA) 18 för kalvningsegenskaper med imputerade HD-genotyper följt av imputerade WGS-varianter. BTA18 analyserades med avseende på sju kalvningsegenskaper i 6113 tjurar med imputerade HD-genotyper. SNP rs136283363 (BTA18: 57.548.213) var genomgående den mest signifikant associerade SNP i alla sju egenskaper. Sedan testades WGS-varianter inom den undersökta QTL-regionen för association med direkta kalvningsegenskaper och med tre kroppsbyggnadsegenskaper. Generna SIGLEC12, CD33 och CEACAM18 föreslogs som kandidatgener. Dessutom har pleiotropa effekter av denna QTL-region observerats på direkta kalvningsegenskaper och kroppsbyggnadsegenskaper. Våra försök att hitta mutationer har dock hindrats av omfattningen av kopplingsojämvikt, bristen på fullständig annotering och potentiella fel i den tillgängliga genomsekvensen.

I kapitel 3 genomförde vi en GWAS för tillväxtegenskaper i Holstein, Jersey, och röda mjölkkor. Först utfördes GWAS inom raser med hjälp av WGS-varianter. Sedan utfördes en metaanalys för att kombinera information över de tre raserna. Flera QTL identifierades som har stora effekter på tillväxtegenskaper i Holstein och röda mjölkkor, men bara en QTL ligger i närheten gen CYP19A1 på kromosom 10 var densamma i Holstein och hos röda mjölkkor. En annan QTL vid positionen 25

Mb på kromosom 14 hade ett mycket signifikant samband med tillväxtegenskaper hos röda mjölkkor, vilket överensstämmer med den tidigare rapporterade genen PLAG1, som påverkar tillväxten hos människor och köttraser av nötkreatur. Ingen QTL var statistiskt signifikant i Jersey, vilket kan bero på låg styrka med den begränsade provstorleken. Metaanalys av de tre raserna förstärkte styrkan att upptäcka QTL.

I kapitel 4 utförde vi imputering av markörer på X-kromosomen i Holsteinboskap för icke-genotypade djur och djur som genotypats med låg densitet (Illumina BovineLD) chips, med hjälp av djur som genotypats med medium densitet (Illumina BovineSNP50) marker. I denna studie användes mjukvaran FImpute V2.2 och 26884 Holsteinindivider genotypade med markörer med medeldensitet. Vi fann att imputeringens noggrannhet för markörer på X-kromosomen förbättrades genom att behandla den pseudoautosomala regionen som autosomal och genom att öka andelen kor i referensgruppen. Vi fann också att imputering för icke-genotypade djur i allmänhet hade lägre noggrannhet jämfört med djur som genotypats med SNP-chip med låg densitet. Dessutom ledde högre ackumulerade släktskap mellan referensgruppen och måldjuren till större noggrannhet. Bättre markörtäckning av X-kromosomen bör utvecklas för att underlätta genomiska studier med X-kromosomen i framtida studier.

I kapitel 5 är syftet att upptäcka dominanseffekter på honliga fertilitetsegenskaper i nordiska Holstein med hjälp av data från Illumina BovineSNP50 och utvärdera styrka, precision, och typ I-fel för att upptäcka dominanseffekter genom simuleringar. Honliga fertilitetsdata (antal inseminationer, dagar från kalvning till första insemination, och dagar från första till sista insemination) registrerades från 3040 genotypade kvigor och 4483 genotypade kor med hjälp av den nordiska Holsteinpopulationen. Först kvantifierades den additiva variansen och dominansvariansen kvantifierats med hjälp av GBLUP för fertilitetsegenskaper. Sedan analyserades associationen för kvigor och kor separat. Slutligen genomfördes en simuleringsstudie för att testa styrka, precision och typ I-fel för att upptäcka dominanseffekter. Dominansvariansen var större än den additiva variansen hos kvigor, men hade stort standardfel. Fyra QTL upptäcktes för IFL hos kvigor, medan en QTL upptäcktes för kor. Alla dessa fem QTL upptäcktes med betydande additiva och dominanseffekter. Simuleringar visade att den aktuella provstorleken hade begränsad styrka att

upptäcka dominanseffekter för honlig fertilitet hos nötkreatur. I framtiden måste fler hondjur genotypas och/eller genotyperna imputeras för att kartlägga de genetiska varianter som har dominanseffekter på honliga fertilitetsegenskaper.

List of abbreviations

GWAS	genome-wide association studies
GS	genomic selection
QTL	quantitative trait loci
LD	linkage disequilibrium
WGS	whole-genome sequence
SNP	single nucleotide polymorphism
EBV	estimated breeding value
HD	high-density SNP chip
MAF	minor allele frequency
CP	cross-phenotype
PAR	pseudo-autosomal region
AU	Aarhus University
SLU	Swedish University of Agricultural Sciences

1

General Introduction

This PhD study "Population level genome-wide association studies in dairy cattle" could be simply illustrated using the following mixed model equation ignoring genotype × environment interaction.

$$\mathbf{P = \mu + G + E = \mu + A + D + I + E}$$

Where **P** represents the population level of phenotypes or responsible variables from one trait or multiple traits, one breed or multiple breeds, **μ** is the common mean, **G** are the genotypic values, **A** is the additive genetic deviation, **D** is the dominance deviation, **I** is the epistasis effect, and **E** is the residual. This general discussion will first cover the equation from the left to the right. Then we will walk through different association mapping methods. In the end, current progress and application of association mapping will be described.

1.1 The P

1.1.1 Response variables

Estimated breeding value (EBV), daughter yield deviations (DYD) and Deregressed EBV (DRP), other than raw phenotypes, have been used as response variables for gene mapping in dairy cattle (VanRaden & Wiggans, 1991; Israel & Weller, 2002; Mao *et al.*, 2016a). The comparisons of using EBV, DYD, and DRP as response variables have been done in genomic predictions (Guo *et al.*, 2010; Gao *et al.*, 2013). EBVs have advantages over raw phenotypes: Firstly, EBVs of bulls have relatively little random error and high reliability, because the EBV of a bull is a standard measure of genetic merit estimated from information of daughters and other relatives. Secondly, EBVs are available from routine genetic evaluations, while the raw phenotypes are often not available or not accessible for analyses. However, taking EBVs as response variables could result in high false-positive rate, due to the included information from relatives. Ekine et al. (2014) have shown that the included family information in EBVs could reduce the power and increase the false-positive rates of GWAS. However, they also noted that using EBVs is less problematic for dairy cattle bulls, as the bulk of the information is derived from progeny testing (Ekine et al. 2014).

In contrast, DYD of a bull are the average of the daughters' performance adjusted for fixed effects, non-genetic random effects, and genetic effects of the daughters' dams. Hence, using DYD as response variables might not have large false positive rate. However, DYD include less data and thus have relatively large random errors and lower reliabilities. Moreover, DYD are often not available from routine genetic evaluation. DRP has been proposed for gene mapping and genomic

prediction (Garrick *et al.*, 2009). The use of DRP, adjusted for parent average effects, not only takes account of the heterogeneous variances of EBVs due to different numbers of daughter records per sire, but also removes information from other relatives. In the future, a comparison of using EBV, DYD, and DRP as response variables should be carried out in gene mapping studies.

1.1.2 Cross-phenotype associations

In the past 10 years, many GWAS have been carried out in many species including human, plants and livestock, identifying thousands of genome-wide significant associations (McCarthy *et al.*, 2008; Bergelson & Roux, 2010; Bermingham *et al.*, 2014). Many of these associations were observed to be across multiple traits in the same category, or even distinct traits, and these associations were described as cross-phenotype (CP) associations (Solovieff *et al.*, 2013). CP association has been observed across livestock and humans (Table 1). CP association is defined differently from pleiotropy. Pleiotropy represents a gene or genetic variant affecting more than one trait (Stearns, 2010), whereas CP association could be associations with multiple traits regardless of the underlying genetic cause. Solovieff *et al.* (2013) has classified CP associatons into three categories (Figure 1) : 1. biological pleiotropy which describes one gene or a genetic locus affecting the biology of multiple traits directly; 2. mediated pleiotropy which describes that the CP association occur because the phenotypes are related; 3. spurious pleiotropy which basically means false pleiotropy due to study design, trait misclassification, etc.

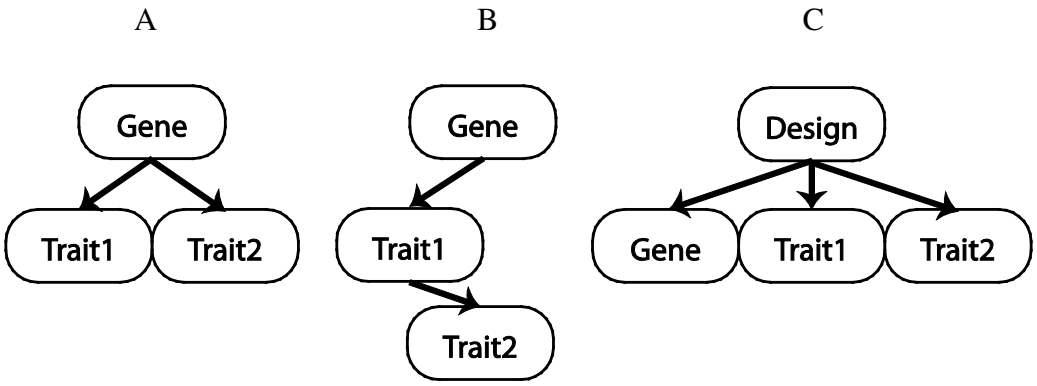

Figure 1. Types of cross-phenotype associations. A: biological pleiotropy. B: mediated pleiotropy. C: spurious pleiotropy.

Table 1. Cross-phenotype associations detected across species.

Variation	Position	Species	Traits	Association	Reference
SNP	39Mb on Chromosome 6	Cattle	Calving and adult stature	LCORL and NCAPG genes in the QTL region for calving traits in dairy cattle, which had been reported to influence fetal growth and adult stature in several species.	(Sahana *et al.*, 2015)
SNP	BTA14:25376827bp	Cattle	Birth weight, human and bovine height	QTL region for birth weight in Nellore cattle harbors genes affecting human and bovine height	(Utsunomiya *et al.*, 2013)
SNP	BTA18: 57,321,450–57,625,355bp	Cattle	Calving and conformation	The QTL shows strong association with calving traits is also strongly associated with conformation traits.	(Mao *et al.*, 2015)
SNP	OAR2:219569259bp	Sheep	Tenderness, meat color, myoglobin, glycogen, unsaturated (omega-3 and -6) fatty acids and saturated fatty acids	Allele near PLCD4 increases tenderness, improves meat color, increases myoglobin, glycogen, and unsaturated (omega-3 and -6) fatty acids and decreases saturated fatty acids.	(Bolormaa *et al.*, 2016)
Copy number variation	16p2.11 duplication	Human	Schizophrenia, autism, intellectual disability, developmental delay, congenital malformations	CNV duplication increases risk for all five disorders	(Helbig *et al.*, 2013)
SNP	rs12720356 (TYK2)	Human	Crohn's disease and psoriasis LDL	The G allele increases risk for Crohn's disease and decreases risk for psoriasis	(Thomas *et al.*, 2008)

Methods to detect CP associations can be generally classified into two groups: multivariate analyses and univariate analyses. Multivariate analyses analyze two or more traits simultaneously to test the associations between one genetic variant and multiple traits. But this methodology requires that all traits are available on the same individuals, which is not always the case. On the other hand, univariate analyses are based on the summary statistics from association between genetics variants and single trait. Univariate analyses are widely applied by large consortia to combine summary statistics from individual research group (Mahajan *et al.*, 2014). Even though different methods have been developed to detect CP associations (Fisher, 1925; Cotsapas *et al.*, 2011; Bhattacharjee *et al.*, 2012), identifying the causal mutations and understanding the underlying biological

mechanisms remains a challenge. As more animals will be sequenced in the future, more sequenced-based GWAS (Höglund *et al.*, 2014a; Mao *et al.*, 2016a) potentially provide us opportunities to detect CP associations more accurately. In the future, univariate analyses could be carried out to combine the single-trait sequence-based GWAS within breed or across breeds, to reveal a more detailed genomic landscape for CP effects.

1.1.3 Meta-analysis

In recent years, GWAS have become a major tool for detecting genetic architectures for complex traits, discovering thousands of associated genetic variants (McCarthy *et al.*, 2008). Even though a single GWAS could identify many common variants, these variants often explain a small fraction of the total genetic variance for the trait. Thus, larger sample sizes are needed to reduce false positives and gain sufficient power to detect associated genetic variants, especially those with small effect sizes (Ioannidis *et al.*, 2006). As a result, meta-analyses were carried out by large consortia to combine the results from multiple independent studies (Mahajan *et al.*, 2014; Replication *et al.*, 2014; van den Berg *et al.*, 2015). Several hundred meta-analyses have been carried out in various species, because meta-analysis does not require individual-level data, but depends on summary data from single GWAS (Evangelou & Ioannidis, 2013). For example, in humans, a meta-analysis was carried out for understanding the genetic basis of type 2 diabetes (T2D) susceptibility (DIAGRAM Consortium et al., 2014). Published meta-analyses of GWAS were aggregated, including 26,488 cases and 83,964 controls of European, East Asian, South Asian and Mexican and Mexican American ancestry. Seven new T2D susceptibility loci were identified and the mapping resolution of the association signals was considered improved. In cattle, Bolormaa et al. (2014) performed a multi-trait meta-analysis to detecting pleiotropic variants for 32 traits categorized in stature, fatness and reproduction in Beef Cattle. They also found that the detection power was increased.

Various methods of meta-analysis have been proposed and they differ in weighting and their abilities to detect heterogeneity. One of the traditional methods combines p-values from multiple independent studies (Fisher, 1925). This method have been widely used in different scientific fields (Evangelou & Ioannidis, 2013). The disadvantages of meta-analysis based on p-values are that it does not provide an overall estimate of effect size, does not account for heterogeneity between datasets, and ignores the direction of effects. The Z-score method, which was based on the average of z-values, was developed to extend p-value methods to take into account of the direction of the effects (Cooper & Hedges, 2009). Another method named fixed-effect meta-analysis assumes true effect of each variant to be the same in each independent study (Pfeiffer *et al.*, 2009). In the most

common weighting strategy, each study is weighted according by the inverse of its squared standard error. This is called inverse variance weighting (Kavvoura & Ioannidis, 2008). Compared to fixed-effect meta-analysis, random-effect meta-analysis assumes the effects of the same variant differ between independent studies. This takes into account the differences between studies (DerSimonian & Laird, 1986). However, random-effect meta-analysis are generally not used for discovery purposes due to far lower power than fixed-effect meta-analysis, but more often used for predictive purposes (Pereira *et al.*, 2009). Bayesian approaches for GWAS meta-analysis were also proposed and they are straightforward and intuitive. However, these approaches need some knowledge about the prior distributions of parameters of interest, and the computation for the genome-wide data can be intensive (Evangelou & Ioannidis, 2013).

1.1.4 Functional traits

There has been declines in genetic level for many functional traits in dairy cattle, due to the negative genetic correlations between production traits and functional traits, such as the negative genetic correlation between milk yield and fertility (Lucy, 2001). However, in recent years, a stabilization or even increase in genetic trends of functional traits has been observed, due to an active and balanced genetic selection (Figure 2). For example, the genetic progress of female fertility has been improved by 20% in the Nordic cattle population (SEGES, 2015), due to that increased weight has been put on female fertility traits in the breeding goal.

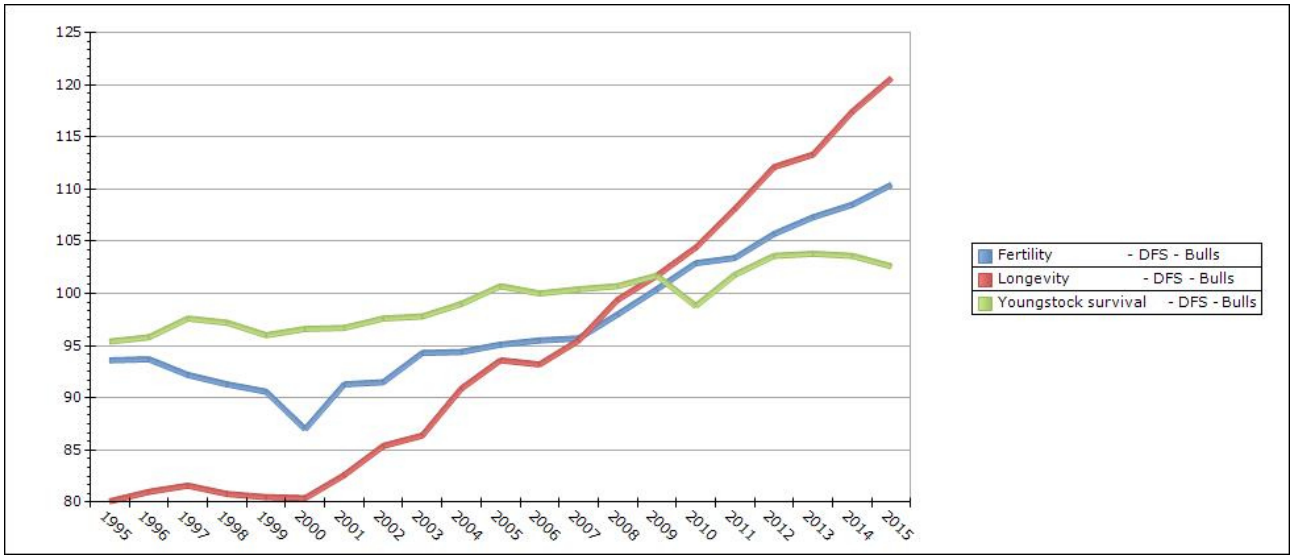

Figure 2. Genetic trend of fertility, longevity, and youngstock survial for dairy cattle in three Nordic countries (Denmark, Finland, and Sweden). On the X axis is the year, on the y axis is the genetic progress (the average breeding value per birth year). (Source: http://www.sweebv.info/ba52nycknav.aspx).

Calving traits were included in this PhD study. Reduced calving ease not only leads to considerable economic losses due to veterinary treatment costs and calf loss but also leads to reduced animal welfare. Three calving traits are included in the Nordic cattle genetic evaluation. They are the birthing process (ease of calving), 24-hour survival after birth (stillbirth) of the calf, and calf size. Calving traits have low heritability, ranging from 0.04 to 0.2 (Lin *et al.*, 1989; Steinbock *et al.*, 2003; Boelling *et al.*, 2007). Several QTL mapping studies for calving traits (Sahana *et al.*, 2011; Höglund *et al.*, 2012; Cole *et al.*, 2014) reported the presence of a QTL with a large effect on calving traits in Holstein cattle at approximately 57 Mb on BTA18. The resolution of these studies, however, was hindered by the limited marker density. Mao *et al.* (2015) carried out a fine-mapping study for the QTL region on BTA18 using imputed HD and WGS data. In this study, significant variants were prioritized with high resolution, and their biological relevance to the traits was interpreted.

Another functional trait category, female fertility, was also included in this PhD study. The traits studied were the number of inseminations (NINS), days from calving to first insemination (ICF), and days from the first to last insemination (IFL). These traits are used as measures of a cow's ability to return to cycling status after calving, to become pregnant after insemination, and maintain pregnancy. Female fertility traits have negative genetic correlation with milk yield (Roxström *et al.*, 2001), and in general have low heritability (from 0.01 to 0.10) (Hou *et al.*, 2009; Sun *et al.*, 2009). QTL mapping studies have been carried out to understand the genetic architecture underlying fertility traits (Palucci *et al.*, 2007; Sahana *et al.*, 2010; Höglund *et al.*, 2014b; Aliloo *et al.*, 2015), utilizing SNP chip and imputed WGS data. Strong associations of SNPs with fertility traits were reported on BTA1, BTA4, BTA7, BTA9, BTA11 and BTA13 in Nordic Holstein, Nordic Red and Jersey dairy cattle.

1.1.5 Production traits

Beef production from dairy production system should not be overlooked, because most of the bull calves from dairy breeds end up in beef production, for example in Nordic countries (Johansson *et al.*, 2008) and Ireland (Hickey *et al.*, 2007). Besides, beef production from dairy cows is more efficient than from suckler cows in terms of climate impact (Johansson et al., 2008). Thus, it is worthwhile to explore the genetic architecture of growth traits, such as daily carcass gain and carcass conformation scores from dairy breeds.

In beef cattle, a large numbers of GWAS have been carried out to identify QTL associated with growth traits (Setoguchi *et al.*, 2009; Lee *et al.*, 2013). For example, a QTL on *Bos taurus* autosome (BTA) 6 was found to be associated with carcass-related traits in Japanese Black and Brown cattle (Setoguchi *et al.*, 2009). However, much fewer QTL mapping studies have been performed on growth traits for dairy cattle than for beef cattle (Elo *et al.*, 1999; Pryce *et al.*, 2011). Dairy cattle might differ in genetic architectures, because beef cattle accumulate nutrients as meat while dairy cattle mainly transform nutrients into milk (Bellmann *et al.*, 2004). Moreover, no definitive findings of causal mutations have been reported for growth in cattle except for the double-muscled phenotype (Grobet *et al.*, 1997).

1.2 The G

1.2.1 Development of genomic tools for Cattle genetics

The cattle played an important role in the development of human civilizations around the world and still the major source of animals protein for nearly 7 billion humans, due to their ability to efficiently convert low-quality forage into energy-dense fat, muscle, and milk (Muers, 2009; The Bovine Genome Sequencing and Analysis Consortium, 2009). In addition, regarding the area of comparative genomics, cattle have an interesting position in the phylogenetic tree: they are in a clade phylogenetically distant from humans and rodents (Muers, 2009). They also provide an important model organism for human genetic research (Andersson, 2016). In consequence, it is important to understand the genetic mechanisms underlying variation in complex traits for cattle.

The continued development of genomic technologies provides a powerful tool to investigate cattle genetics. Like other mammals, cattle genomics developed in the wake of human genomics research and originated in somatic cell genetics (Womack & Moll, 1986). Early on isolated cases of associations between traits and genetic markers were identified (Conneally & Stone, 1965; Hines *et al.*, 1969). In the early 90s, the first "genomic map" produced for cattle consisted of synteny groups assigning loci and markers to specific bovine chromosomes by integrating somatic cell genetics with *in situ* hybridization (Fries *et al.*, 1993). Bovine radiation hybrid (**RH**) maps were developed and then used for high-resolution comparative mapping (Williams *et al.*, 2002). Microsatellite markers in cattle were applied to develop linkage maps to enable QTL mapping for economically important traits (Fries *et al.*, 1990; Barendse *et al.*, 1994; Kappes *et al.*, 1997). In 2009, two assemblies of the bovine genome were published based on the genomic sequence of the Hereford

cow, L1 Dominette 01449. The assembly was initially generated and assembled by the Baylor College of Medicine Human Genome Sequencing Center (The Bovine Genome Sequencing and Analysis Consortium; & Elsik, C. G.; Tellam, R. L.; Worley, 2009). At the same time, Zimin et al. (2009) published an alternative assembly using the primary sequence data. Both assemblies have gone through several iterations of improvement and were upgraded to the current versions, Btau_4.6.1 and UMD3.1.1 (http://bovinegenome.org/). Then, there was a rapid improvements in high-throughput methods for SNPs genotyping and sequencing technologies (Matukumalli *et al.*, 2009; Goodwin *et al.*, 2016). In addition, the cost of sequencing decreases dramatically. Figure 3 illustrates the changes of the cost of sequencing a human genome. Due to the reduced cost, followed by an initiative human genome project known as 1000 Genome project (Consortium *et al.*, 2010), an analogous project called The 1000 Bull Genomes Project was carried out for bovine genetics and genomic research (Daetwyler *et al.*, 2014). These advances in genomic technologies have contributed to the research of GWAS and Genomic Selection (GS) (Meuwissen *et al.*, 2001).

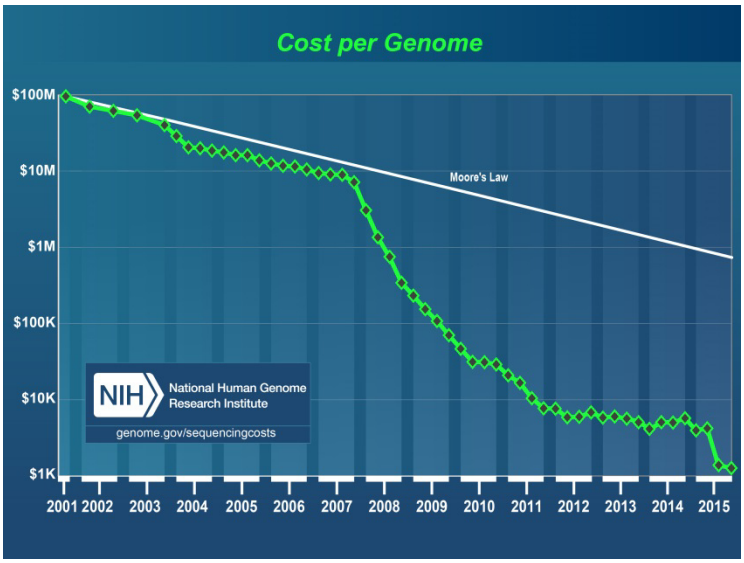

Figure 3. Sequencing cost presented from 2001 to 2015. (https://www.genome.gov/27541954/dna-sequencing-costs-data/)

1.2.2 Imputation

Genotype imputation is defined as predicting genotypes that are not directly typed in samples of individuals based on information from other individuals (Marchini & Howie, 2010). Imputation has been applied in several scenarios: 1. Imputation could be used to impute the low-density genotyped or even non-genotyped individuals using the reference population genotyped with high-density chip,

in order to reduce the genotyping cost (Daetwyler *et al.*, 2011; Pryce *et al.*, 2014); 2. Imputation is often used for combining data from different studies and genotyping platforms, in order to improve the power of analyses. For instance, imputation is widely used in the meta-analysis of GWAS, to detect variants with small to moderate effects (Saxena *et al.*, 2007; WTCCC, 2007); 3. Imputation could be beneficial to the call rate for individuals with sporadically missing genotypes (Marchini & Howie, 2010).

Current imputation methods rely on two sources of information, thus these methods could be classified into two classes. One class contains population-based methods, which use population linkage disequilibrium (LD) information to predict unobserved marker types based on LD with observed marker types; the other class contains family-based methods, which use linkage information from close relatives.

Population-based methods, which are normally probabilistic based, model haplotype frequencies using LD information. For example, *Beagle*, a commonly used imputation software, uses hidden Markov model (HMM) to infer the haplotype phase and to impute non-genotyped markers (Browning & Browning, 2009). *Beagle* employs a graphical model that constructs a tree of haplotypes from the reference population, and then summarizes it in a directed acyclic graph by haplotype similarity. Then, the probability of a missing genotype is computed by averaging posterior genotype probabilities (Browning & Browning, 2009). *Beagle*4.1 claims to be fast, accurate, and memory-efficient by restricting the probability model to genotyped markers in the target samples and by performing linear interpolation to impute non-genotyped markers (Browning & Browning, 2016). Another commonly used software, *IMPUTE2*, also employs HMM which is based on an approximate coalescent model including mutation and recombination process to infer the haplotypes.

Family-based methods are mainly rule-based methods. These models appear to work better for imputation of very low-density genotyped animals, especially when many close relatives are genotyped (Bouwman *et al.*, 2014). For example, *AlphaImpute* utilizes simple phasing rules, long-range phasing, haplotype library imputation, and segregation analysis (Hickey *et al.*, 2011). *FImpute* relies on efficiently exploiting genealogy or relationships between individuals by searching for haplotypes from the longest to the shortest. This concept assumes that close relatives share longer haplotypes that have lower frequency in the population, and distant relatives share shorter haplotypes which usually have higher frequency (Sargolzaei *et al.*, 2014).

With the rapid reduction in genotyping costs, access to phenotypes becomes a limiting factor especially for traits that are difficult or expensive to measure. Imputing non-genotyped individuals with such phenotype records could be potentially beneficial for GWAS or GS (Bouwman *et al.*, 2014). Currently, only a few family-based imputation programs can impute non-genotyped individuals, such as *AlphaImpute* (Hickey *et al.*, 2011), *FImpute* (Sargolzaei *et al.*, 2014), *FindHap* (VanRaden *et al.*, 2013), and *PedImpute* (Nicolazzi *et al.*, 2013). Another issue that needs more investigation is the imputation for the X chromosome, as most of the imputation studies focused on imputation of autosome (Brøndum *et al.*, 2014; van Binsbergen *et al.*, 2014). However, in dairy cattle for example, ignoring the X chromosome could miss important biological functions and affect genomic evaluation (Sandor *et al.*, 2006; Lyons *et al.*, 2014; Su *et al.*, 2014; de Camargo *et al.*, 2015). Several programs have extended the algorithm to consider the X-chromosome imputation (Hickey *et al.*, 2011; Sargolzaei *et al.*, 2014). It was showed that the Z chromosome in chickens with 25 SNPs could be imputed to a higher density (1,137 SNPs) with relatively high accuracy (approximately 0.9) (Hickey & Kranis, 2013). However, in this study the pseudo-autosomal region (PAR) that recombines with the W chromosome was treated as X-linked in the analysis.

1.2.3 The X chromosome

In mammalian, the females have two X chromosomes while males have one X and one Y chromosomes. Only the pseudo-autosomal region (PAR) is homologous between the X and Y chromosomes, which is required for sex chromosome segregation during meiosis in males (Das *et al.*, 2009). In the bovine genome, the X chromosome is relatively long (148,823,899 bp), constituting approximately 6% of the total physical genome (Zimin *et al.*, 2009). ENSEMBL (release 82) reports 19,981 protein coding genes in the whole bovine genome, of which 833 (4.2%) are found on the X chromosome. It was also reported that the inclusion of the markers on the X chromosome enhanced reliabilities of genomic prediction in dairy cattle. On average markers on the X chromosome accounted for 1.7% of the total additive genetic variance of 15 indices included in the Nordic Total Merit index (http://www.nordicebv.info) (Su *et al.*, 2014). In the human genome, the X chromosome harbors more than 2,300 genes, including coding, noncoding and pseudo genes (Flicek *et al.*, 2014). Around 7% of phenotypes with a known molecular basis are X linked, including autoimmune, cognitive, and behavioral conditions, according to the Online Mendelian Inheritance in Man (OMIM) catalog of human genes and genetic disorders (http://www.omim.org). Considering the potentially important biological functions of the X chromosome in animals and

human, the X chromosome is vital in the genetic analyses. However, the X chromosome has often been excluded from GWAS, as evident from only 242 out of all 743 human GWAS conducted from January 2010 to December 2011 took the X chromosome into account (Wise *et al.*, 2013).

There are particular issues when including the X chromosome in GWAS: 1. Lower marker density on the X chromosome on the current genotyping platforms compared to autosomes for species like cattle. For instance, although the length of the X chromosome is about 6% of the total bovine genome, common genotyping chip Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) (Matukumalli *et al.*, 2009) has only 1.6% (894 out of 55,298) of total SNPs from the X chromosome; 2. Lower genotype calling accuracy compared with that of autosomes. Problems might occur with genotype calling for hemizygous males due to the lower intensity of the X-chromosome variants, because males might cluster differently than females (Wise *et al.*, 2013). 3. Quality control (Hardy-Weinberg equilibrium and minor allele frequency (MAF)) need to be adjusted for the X chromosome, because the expected frequencies of genetic variants are sex dependent (Wise *et al.*, 2013); 4. Different imputation strategies compared with the imputation of autosomal markers. Imputation is crucial for combing data and is a cost-effective approach to augment genomic data from low-density or even non-genotyped individuals to higher marker density. The complexity of imputing the X chromosome lies with the fact that the X-chromosome inheritance is different from autosomes except in the PAR region where it is similar to the situation on autosomes. Imputation for the X chromosome in dairy cattle, which takes the PAR into account, has been investigated using existing imputation softwares (Mao *et al.*, 2016b). 5. The test statistics for autosome GWAS do not apply to the X chromosome, except for the situations that only females or the PAR are analyzed. A number of different test statistics have been proposed for the X-chromosome association studies (Nicolae, 2006; Clayton, 2008). In the future, more tailor-made guidelines should be applied for X-chromosome GWAS, such as tailor-made quality control and test statistics.

1.3 The A, D and I

1.3.1 From additive to non-additive inheritance

In quantitative genetics breeders have mainly focused on additive genetic variations. Also in applications such as GWAS and GS focus has almost exclusively been on additive genetic variance. Even though studies have reported that non-additive effects make a substantial contribution to

genetic variation in complex traits (Gengler *et al.*, 1997; Palucci *et al.*, 2007; Norris *et al.*, 2010), this source of genetic variation is often neglected. However, understanding the non-additive part of the genetic architecture of traits is helpful for planning breeding strategies and increasing genetic gains. For example, non-additive effects can be utilized by designing mating schemes that optimize favorable allele combinations, especially when family or clonal propagation are available in the breeding program (Muñoz *et al.*, 2014). Non-additive genetic variations include dominance which is the interaction between alleles at the same locus, and epistasis which is the interaction between alleles at different loci. Hill *et al.* (2008) examined a wide variety of theoretical models and showed high proportions of additive genetic variance even in the presence of non-additive gene action, basically because most alleles are likely to be at extreme frequencies. Due to the strong artificial selection in practical animal and plant breeding, gene frequencies are expected to be at extreme for the causal loci and therefore, a large proportion of non-additive gene action could contribute to the additive genetic variation. This influences the dissection of genetic architecture of complex traits and genetic evaluation.

Non-additive genetic variation in growth, carcass and fertility traits has been investigated in Australian beef cattle, using 729,068 SNPs (Bolormaa *et al.*, 2015). This study showed that the number of SNPs significantly associated with dominance effects was higher than expected by chance, and the authors presumed that most significant dominance effects were to increase fitness and in the opposite direction to inbreeding depression. Another study showed that dominance effects play a relevant role in the genetic architecture of number of teats in pigs (Lopes *et al.*, 2014). In this study, the dominance genetic variance of the four QTL detected explained 1.82% of the total phenotypic variance, corresponding to one-fourth of the additive genetic variance. Su *et al.* (2012) proposed a method to build a dominance relationship matrix and an epistatic interaction matrix using genome-wide SNPs and illustrated that models including non-additive genetic effects improved unbiasedness of genomic predictions for daily gain in pigs. Su *et al.* (2012) demonstrated that dominance genetic variance accounted for 5.6% while additive by additive epistatic genetic variance accounted for 9.5% of the total phenotypic variance. Munoz et al. (2014) analyzed height data from a multi-family population of the tree species, *Pinus taeda*. It was shown that realized genomic relationships built from markers yielded a more precise partition of additive and non-additive genetic variance, compared with pedigree based relationship. In addition, they showed that the additive and non-additive genetic variance were similar in magnitude in the analyzed population.

1.4 Association mapping Methods

During the past decade GWAS have become a popular tool for gene mapping, not least due to the availability of HD and WGS data. GWAS has obvious advantages compared to the traditional QTL mapping approaches such as candidate gene approach and linkage mapping analyses (Hirschhorn & Daly, 2005). For example, the linkage mapping analyses, which are often performed using microsatellite markers, require family information and have poor resolution. GWAS do not need related individuals because they utilize the LD information among markers in a population and normally have finer resolution. Compared with candidate gene approach, GWAS do not require prior knowledge of the analyzed genomic regions.

As in other statistical areas, methods of GWAS can be broadly classified into two types: frequentist and Bayesian approaches. The fundamental difference between frequentist and Bayesian approaches in the GWAS context is that Bayesian approaches quantify the probability of a SNP associated with the phenotype based on both the prior knowledge and current data (Stephens & Balding, 2009). Several models have been proposed for the frequentist approaches such as single-marker tests (Cleveland & Deeb, 2009), linear mixed model analysis (Yu *et al.*, 2006), genealogy based mixed-model, and haplotype models (Dashab *et al.*, 2012). In a frequentist approach, SNP is normally analyzed one at a time using a linear mixed model that includes the effect of a SNP, fixed effects such as batch effect, the year, the cohort or group to which the individual belongs, and the random polygenic effect of each individual. The polygenic value is fitted to consider all other genes affecting the trait besides the SNP under evaluation. The significance of association between a SNP and the phenotype of interest is tested by comparing with a null hypothesis of no association. Several Bayesian models were also proposed such as least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), and Bayesian variable selection models (BVS) (George & McCulloch, 1993). In a Bayesian approach such as BVS, all SNPs are fitted in the model simultaneously and Markov chain Monte Carlo (MCMC) algorithms are implemented (Satagopan *et al.*, 1996). Bayesian methods provide an alternative approach to test associations through so called Bayes factor, which is the ratio between the posterior odd ratio to prior odd ratio.

1.5 Current progress and applications of GWAS

GWAS were first applied in the analysis of human disease, and then were extended to the field of domestic animals due to the availability of the reference genomic sequences (Hillier *et al.*, 2004; Zimin *et al.*, 2009). By far, a large number of phenotype-genotype associations have been identified

in both humans and domestic animals. Table 2 lists some examples of reported GWAS in domestic animals.

In human, GWAS results have been or probably will be applied in the areas of disease prediction, biomarker identification, disease sub-classifications, personalized medicine like treatment selection and drug dosing. For disease prediction, some treatments will prevent disease entirely or are the most effective before clinical abnormalities or a firm clinical diagnosis. Thus, it is important to predict the risk of diseases earlier in life. For example, Type 1 diabetes mellitus results in serious morbidity and mortality, requiring life-long insulin treatment and is highly heritable. The β-cells will be almost destroyed, and there are no effective treatments, when Type 1 diabetes mellitus is detected clinically (Chatenoud *et al.*, 2012). Applying GWAS findings to clinical practice requires expertise from a wide range of disciplines, including molecular biology, clinical medicine, pharmacology, bioinformatics, implementation research, and clinician education (Manolio & Green, 2011).

In domestic animals, GWAS findings could be incorporated in a genetic evaluation model to increase accuracy of genetic prediction (Boichard *et al.*, 2012). In the genomic BLUP (GBLUP) model, it is expected that with different weights on different SNPs depending on their association results, the relationship between test and training animals will be estimated more accurately. Brøndum et al. (2015) investigated the effect on the accuracy of genomic prediction, of adding a small number of significant variants from single marker analysis based on WGS association results to the regular 50k SNP array data. In this study, 5 index traits from Nordic Holstein, French Holstein, and Nordic Red cattle were evaluated using GBLUP and Bayesian 2-distribution mixture model. Results showed that there were increases in accuracy of around four percentages in prediction reliability points for production traits and less than one percent for functional traits using GBLUP. The increases of in prediction reliabilities were less pronounced when using Bayesian models.

Table 2. Examples of reported GWAS in domestic animals.

Trait	Genotype	Phenotype	Method	Significant level	Finding	Reference
Milk yield	50K	62,343 Holstein Friesian cows sired by 798 bulls	Mixed linear models	P< 0.001	362 significant SNPs	(Hayes *et al.*, 2009)
Fertility	50K	2,531 Danish and Swedish Holstein bulls	Mixed model analysis	Chromosome-wise Bonferroni correction	74 significant SNPs mainly on BTA 3, 5, 10, 13, 19, 20, and 24	(Sahana *et al.*, 2010)
Tuberculosis	700k	1,223 female cattle (629 cases, 594 controls)	GRAMMAR,regional heritability mapping and haplotype-sharing analysis identified	Chromosome-wise Bonferroni correction	2 significant SNPs	(Bermingham *et al.*, 2014)
Fertility	Sequence	3,475 Nordic Holstein bulls from Denmark, Sweden and Finland	Mixed linear models	Chromosome-wise Bonferroni correction	9 genes	(Höglund *et al.*, 2014b)
Androstenone	60K	987 pigs divergent for androstenone concentration from a commercial Duroc- based sire line	QFAM test	FDR of q-value $\leqslant$ 0.05	37 significant SNPs	(Duijvesteijn *et al.*, 2010)
Canine atopicdermatitis	20K	48 Golden Retrievers including 25 with atopic dermatitis and 23 healthy controls	Chi-square test	P < 0.001	35 significant SNPs	(Wood *et al.*, 2009)
Racing distance	50K	118 elite Thoroughbred racehorses	Chi-square test	Bonferroni correction	a significant 690 kb region	(Hill *et al.*, 2010)

1.6 Objectives

The overall objectives of this research were to fine-map functional (calving and female fertility) and production (growth) traits in dairy cattle utilizing HD and imputed WGS data, and explore the genotype imputation of the X chromosome and mapping variants exhibiting dominance effects for female fertility.

1. **Chapter 2** performed fine-mapping of the QTL on BTA18 for calving traits in the Nordic Holstein cattle population using HD genotypes and WGS variants, and investigated the pleiotropic effects of the fine-mapped QTL on conformation traits using WGS variants.
2. **Chapter 3** carried out GWAS for growth traits in three dairy cattle breeds, Holsteins, Jerseys, and Red Dairy Cattle, and meta-analysis was performed to gain power by combining association results from these three breeds.
3. **Chapter 4** investigated the imputation accuracy of markers on the X chromosome in dairy cattle for non-genotyped animals and animals genotyped with low-density chips.
4. **Chapter 5** explored the genome-wide dominance effects for female fertility traits in Danish Holstein cattle using Illumina BovineSNP50 data, and evaluates the power, precision, and type 1 error of detecting dominance effects using simulated data.

2

General Discussion

The objectives of this thesis were to fine map QTL underlying functional (calving and fertility) and production (growth) traits by utilizing HD genotypes and imputed WGS variants, and explore genotype imputation for the X chromosome and mapping variants exhibiting dominance effects for female fertility. In chapter 2, *Bos taurus* autosome (BTA) 18, which had previously been reported to have a large effect on calving traits in Holstein cattle (Kühn *et al.*, 2003; Thomasen *et al.*, 2008; Cole *et al.*, 2011; Sahana *et al.*, 2011), was scanned using imputed HD genotypes followed by imputed WGS variants. *SIGLEC12*, *CD33* and *CEACAM18* were proposed as candidate genes. Pleiotropic effects of this QTL on direct calving and conformation traits were detected. In chapter 3, GWAS were performed on growth traits in Holstein, Jersey, and Red Dairy Cattle using WGS variants. *CYP19A1* is a strong candidate gene for a QTL on BTA10 which was highly significant in both Holstein and Red Dairy Cattle. It has previously been reported that the gene *PLAG1* has an effect on growth traits in humans and livestock (Gudbjartsson *et al.*, 2008; Lettre *et al.*, 2008; Pryce *et al.*, 2011) . In chapter 4, the imputation accuracy for markers on the X chromosome in Holstein cattle was assessed. We reported that the imputation accuracy can be improved if PAR is treated as autosomal, and also if the proportion of females in the reference group is increased. In chapter 5, genetic variants throughout the genome were scanned for additive and dominance effects on female fertility traits. Only a few QTL were significant for both additive and dominance effects in Nordic Holstein cattle. The simulations studies showed that the power to identify QTL was limited due to small sample size and when the QTL heritability was low.

In this general discussion part, I will discuss the benefits of using WGS data for mapping genes affecting economic traits and limiting factors in detecting causal mutations in cattle. Then, I will discuss the genotype imputation strategy for the markers on the X chromosome. Finally, CP associations and increasing cow data for enhancing GWAS power and exploring non-additive effects will be discussed.

2.1 Utilization of WGS for identifying causal mutations

Does WGS data lead us to identify the casual mutations affecting economically important traits in cattle? I studied this with a QTL having major effect on calving traits as an example (chapter 2). A QTL at approximately 57 Mb on BTA18 was reported to have a large effect on direct calving traits,

through linkage-based QTL-mapping studies (Kühn *et al.*, 2003; Thomasen *et al.*, 2008) and association studies (Cole *et al.*, 2009; Sahana *et al.*, 2011). SNP rs109478645 (allele C and A) at BTA18:57,589,121 bp was the most significant among markers on the Illumina 50K SNP array data (Cole *et al.*, 2009; Sahana *et al.*, 2011). This SNP is located in an intron of the sialic acid-binding Ig-like lectin-12 gene (SIGLEC12), which is expressed in the human placenta, and reported to function in the initiation of parturition (Brinkman-Van Der Linden *et al.*, 2007). However, the resolution of these studies was limited because of the low marker density for this QTL region. Based on the WGS data, variant rs454366488 (BTA18: 57 477 561) was found to have the strongest association. However, several other variants nearby also showed similar significance level, because the LD is quite high among these nearby variants (Figure 2). Another QTL targeted for fine-mapping was segregating in Holstein and Nordic Red cattle. In chapter 3, a QTL located on BTA10 at approximately 59.2 Mb was detected to be significantly associated with growth traits in both Holsteins and Red Dairy Cattle. The most significantly associated variants were located close to the gene *CYP19A1,* which is known to encode the aromatase that catalyzes the conversion of androgens to estrogens (Chwalisz & Fürbass, 2014), and was reported to affect both growth and reproduction in cattle (Wendorf *et al.*, 1983), mice (Heine *et al.*, 2000), and humans (Öz *et al.*, 2001). Thus, *CYP19A1* was proposed as a candidate gene for growth traits. However, even with WGS data, we did not succeed in pinpointing the exact causal variant due to strong long-distance LD. Haplotype-based analyses showed multiple haplotypes with a significant effect, indicating the underlying causal mutations might not follow simple bi-allelic QTL model or the causal mutation is an old mutant allele that is present in multiple haplotype background.

In principle, GWAS with WGS variants could directly locate most of the causal variants, because nearly all variants segregating in the population are included in the analysis. However, our study in chapter 2 and 3 showed that WGS data did not manage to pinpoint the causal mutations. The reasons could be: 1. not all variants are included in our WGS variant data set because some variants were removed during the quality-checking process, certain types of variants are not normally detectable using WGS and non-bi-allelic variants (structural variants) were removed due to limitations of the imputation software; 2. LD with causal mutations is a key factor for mapping but could hinder the precise identification of causal ones, especially for Holstein that maintain long range haplotypes (De Roos *et al.*, 2008); 3. The functional annotations for cattle genome were not detailed enough to prioritize the statistically significant variants. More detailed genomic annotation data could be used 1. to be directly incorporated in GWAS models, for example, weighting

sequence variants based on their annotation have been shown to improve the power of whole-genome association studies (Sveinbjornsson *et al.*, 2016), 2. to facilitate the explanations of GWAS findings in *post-hoc*, for example, the missense variants could be more likely to alter the gene translation than anonymous variants.

Even though WGS data did not help us to reveal the causal mutations, it should be noted that GWAS are essentially based on the correlation, but correlation does not imply causation (Aldrich, 1995). GWAS are the first step for identifying causal variant, because they are carried out without prior knowledge of the underlying biological mechanisms (Hirschhorn & Daly, 2005). Once a QTL has been detected in one population, follow-up studies should be carried out to confirm this QTL in different populations, and then functional studies should be carried out to identify causal mutations. GWAS results could provide a source for gene editing (Gaj *et al.*, 2013) targets to find out the molecular basis for certain traits. A candidate causal variant could be confirmed by looking at phenotypic change through precise and targeted changes. For example, the approach of combination of GWAS and CRISPR elucidates a mechanistic basis for the strongest genetic association (FTO locus) with obesity (Claussnitzer *et al.*, 2015). In human genetics, it is essential to know the causal mutations for the purpose of medical treatment. In animal breeding, knowing the causal mutations might not be that urgent because the significant variants could be implemented to weigh the genomic relationship to improve the selection accuracy (VanRaden, 2008; Brøndum *et al.*, 2015).

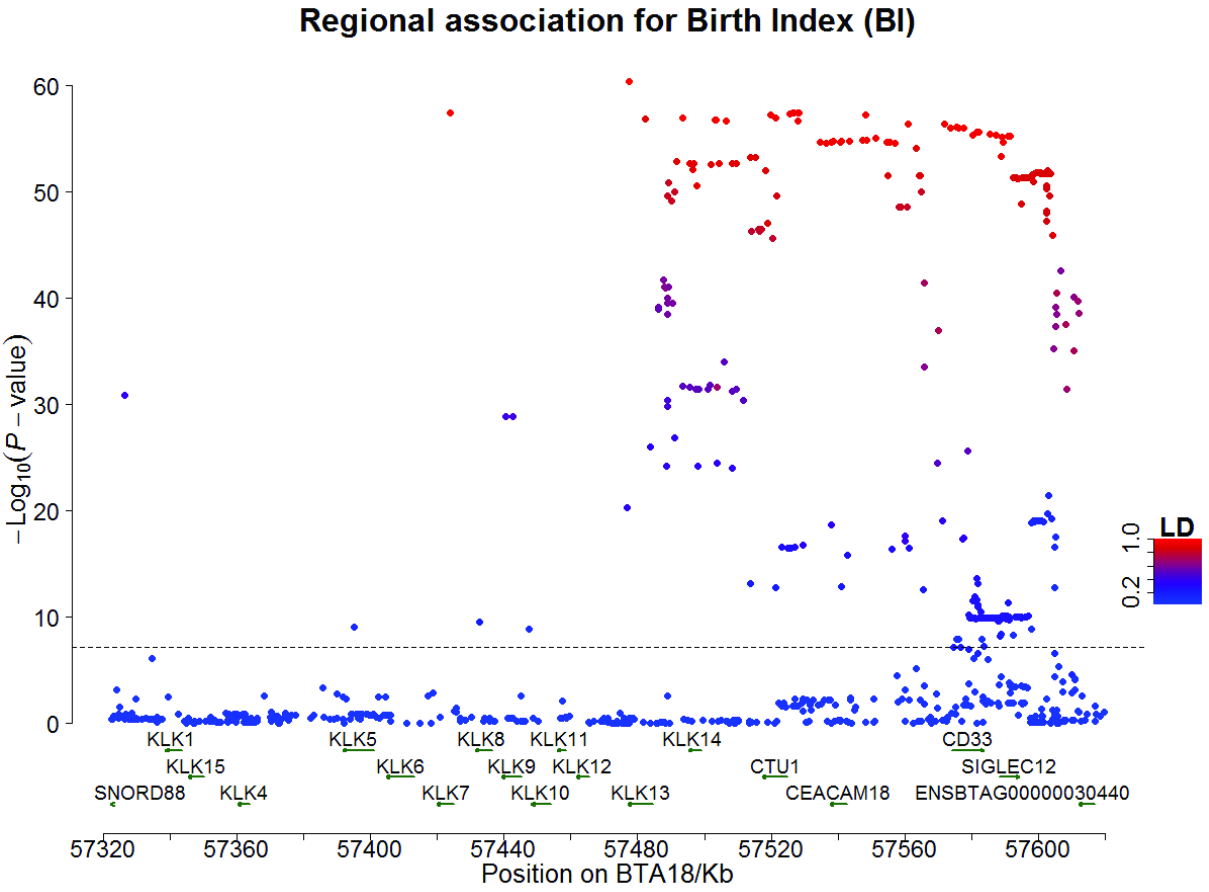

Figure 2. Regional single-marker analysis for the Birth Index (BTA18: 57,321,450–57,625,355) using imputed whole-genome sequence variants. All genes located in this region are represented with green arrows.

Computational costs have been an issue for GWAS utilizing WGS data. In chapter 2, the number of variants after quality control was 13,396,556 for Holsteins, 11,423,283 for Jerseys, and 14,002,305 for Red Dairy Cattle. Fitting these amounts of variants in a mixed model one by one will cost up to months in analysis, with a typical sample size such as 5,000 records. This expensive computation time could be shortened and different approaches have been proposed. Aulchenko et al. (2007) proposed a two-step method to approximate the maximum likelihood estimate for mixed models, which is currently implemented in *GRAMMAR*. In the first step, a reduced model without genotypes is fitted. In the second step, the residuals from the reduced model are fitted in a model with genotypes, assuming non-genetic effects and variance component are constant. The inference of the genotype effects is close to that obtained from the full model, when the inclusion of genotype does not change these parameters. However, the results could be inaccurate when genotype is correlated with non-genetic effects. Moreover, effect of the variant will be underestimated if the effect is large.

Rather than fixing both the non-genetic effects and variance components, a mixed model that only fixes the variance components over all variants was proposed (Kang *et al.*, 2010). This model assumes that each variant explains only a small fraction of the total genetic variance of a trait, avoiding estimating the time-consuming variance component repeatedly for each variant. Thus, the computational time is decreased but slower than *GRAMMAR*. This method is implemented in software *EMMAX*. The drawback of this method is that it still does not solve the problem of underestimating the variant with a large effect. In chapter 2, we have applied *EMMAX* in our analyses of WGS data and the computation time was reduced from weeks to only a few hours. The greatly reduced computation cost enables us to carry out GWAS on three breeds using imputed WGS data. The drawback of *EMMAX*, which underestimates variants having large effects, did not seem to impact our results as the major QTL were detected due to large significances. And we could re-analyze the region of interest by running the complete model to obtain proper effect estimates.

2.2 The X chromosome

In chapters 2 and 3, genome-wide analyses were carried out using imputed HD and WGS genotypes. However, the X chromosome was not included in those analyses. A similar situation is also common in human GWAS. It was reported that only 242 out of 743 human GWAS conducted from January 2010 to December 2011 included the X chromosome in the analyses (Wise *et al.*, 2013). However, the X chromosome should not be ignored because the X chromosome in bovine genome is relatively long (148,823,899 bp), constituting approximately 6% of the total physical genome (Zimin et al. 2009). Ensembl (release 82) reports 833 (4.2%) of 19,981 protein coding genes on the bovine X chromosome. For example, non-synonymous mutations on the X chromosome were reported to be associated with andrological (traits that related to male fertility) and growth traits in beef cattle (de Camargo *et al.*, 2015). In our case, the X chromosome was not include because the markers on the X chromosome were not imputed to higher density, due to limitations of imputation tools at the moment of analyses. The imputation strategy needs to be different for the X chromosome from that for autosomes. The X chromosome for mammals has a special genetic mode of inheritance: females have two copies and males have only one, leading to different inheritance pattern for the X chromosome from that for the autosomes.

In chapter 4, new strategies for imputing the markers on the X chromosome in Holstein cattle for both non-genotyped animals and animals genotyped with low-density (Illumina BovineLD,

Illumina Inc., San Diego, CA) chips were investigated. The existing software *FImpute V2.2* (Sargolzaei *et al.*, 2014) was used, and large number of animals (26,884) genotyped with medium-density (Illumina BovineSNP50) chips were used as the reference group. The X chromosome was successfully imputed for both non-genotyped and low-density genotyped animals. We also gained imputation accuracy of markers on the X chromosome when treating the PAR as autosomal. This improved imputation accuracy suggests that the PAR should be imputed separately and treated as autosomal when imputing genotypes, because the PAR is inherited differently from the X-linked region. Besides the treatment of the PAR region, other factors affecting imputation accuracy for the X chromosome were also assessed in chapter 4, such as different male/female proportions in the reference group, and cumulated degree of relationship between the reference group and target group. Chapter 4 provides a strategy and more insights for imputing the X chromosome, which could be used as reference for routine imputation of the X chromosome for GWAS.

In chapter 5, a genome scan including all the autosomes plus the X chromosomes was carried out for variants having additive and/or dominance effects on female fertility traits in Holstein cattle, using raw female fertility phenotypes. This study utilized 3,040 heifers and 4,483 cows genotyped with Illumina BovineSNP50 BeadChip version 1 and 2 (Illumina Inc., San Diego, CA). The genotypes had between 2,448 (BTA1) and 707 (BTX) markers after quality control. No SNP with significant effects (additive or dominance) was identified on the X chromosome. Moreover, the SNPs on the X chromosome in general had lower significance compared with SNPs on the other autosomes. Sex chromosomes have important roles in germ cell development in mammals, being enriched in genes expressed in the testis and ovary (Heard & Turner, 2011). For example, in humans, some cases of male infertility arise because of mutations in testis-expressed 11 gene (TEX11) in the maternal X chromosome that prevent development of viable sperm (Yatsenko *et al.*, 2015). Therefore, it is logical to assume that some fertility genes could be located on the X chromosome. However, this is not the case in our study. The reasons might be: 1. the genotyping platform for our study has low coverage for the X chromosome. The Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) (Matukumalli *et al.*, 2009) contains only 1.6% (894 out of 55,298) of total SNPs on the X chromosome; 2. The statistical power of our study design is too low to detect the additive and dominance effects. The chapter 5 did not consider specific test statistics for the analyses of the X chromosome (Nicolae, 2006; Clayton, 2008), because only females were used in our study. In the future, better coverage of the X chromosome and larger sample size should be used to map QTL on the X chromosome.

2.3 Cross-phenotype associations

In chapter 2, the WGS variants in the QTL region (BTA18: 57,321,450–57,625,355 bp) for calving traits were also tested for the associations with conformation traits, including body depth, bone structure and stature. These WGS variants (BTA18: 57 321 450– 57 625 355) exhibit moderate associations with body depth and bone structure, but weak association with stature. Besides, the variants that were highly significant were common for both conformation traits and calving traits. Cole et al. (2009) performed a genome scan for 27 traits in American Holstein bulls. They also reported this calving trait QTL exhibits pleiotropic effects on conformation, economic merit and longevity. This pleiotropic effect might be categorized as mediated pleiotropy (Solovieff *et al.*, 2013), because calving and conformation traits could be phenotypically related. However, we can not here distinguish mediated pleitropic effect from biological pleitropy, because we do not know if one gene or a genetic locus underlying this QTL is directly affecting the biology of both calving and conformation traits.

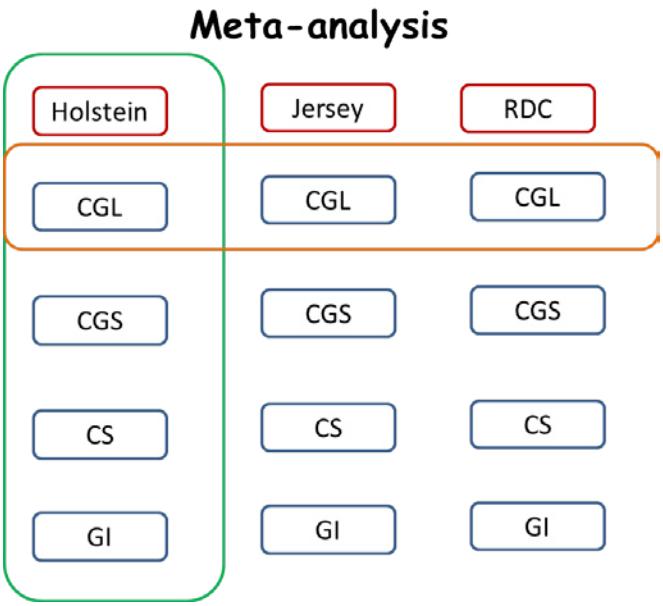

Figure 3. Demonstration of multi-breed meta-analysis and cross-phenotype (CP) meta-analysis. Orange rectangle highlights the multi-breed meta-analysis and green rectangle highlights the CP meta-analysis. CGL = carcass gain with a long finishing period; CGS = carcass gain with a short finishing period; CS = carcass conformation score; GI = growth index.

In chapter 3, a multi-breed meta-analysis was carried out to gain power by combining information from Holstein, Jersey, and Red Dairy Cattle. Additional CP meta-anaysis was carried out for growth traits. The difference between multi-breed and CP meta-analysis is shown in Figure 3. In CP meta-analysis, we utilize an approach that can integrate association evidence from summary statistics of CGL, CGS, CS, and GI from Holstein, Jersey, and Red Dairy Cattle (Zhu *et al.*, 2014). For comparison, the results of CP meta-analysis for CGL, CGS, CS, and GI are shown in Figure 4, while the manhattan plot of GI in multi-breed meta-analysis is shown in Figure 5. In CP meta-analysis, more significant variants were detected compared to the multi-breed meta-analysis of a single trait. In addition, the large QTL effects which was detected in the multi-breed meta-analsyis were still remain high significance in the CP meta-analysis. This CP meta-analysis suggests that analyzing multiple traits could improve the statistical power for the QTL that were missed in a single trait analysis. As more traits were recoded in one animal and single-trait GWAS has been performed on most of the traits, CP meta-analysis could be easily carried out. The summary statistics of these single-trait GWAS could be combined in this way to integrate the phenome-wide data available for genetic association analysis. However, we need to pay attention that double counting might occur when combing summary statistics from multiple traits, because multi-trait models could have been applied in the producre of estimating EBV in routein genetic evaluation in cattle. In chapter 5, declines in genetic level for many functional traits has been observed in dairy cattle, due to the negative genetic correlations between production traits such as milk yield and functional traits such as fertility (Lucy, 2001). Thus, the CP association analysis might be worth to be carried out to dissect this negative genetic correlation in future studies.

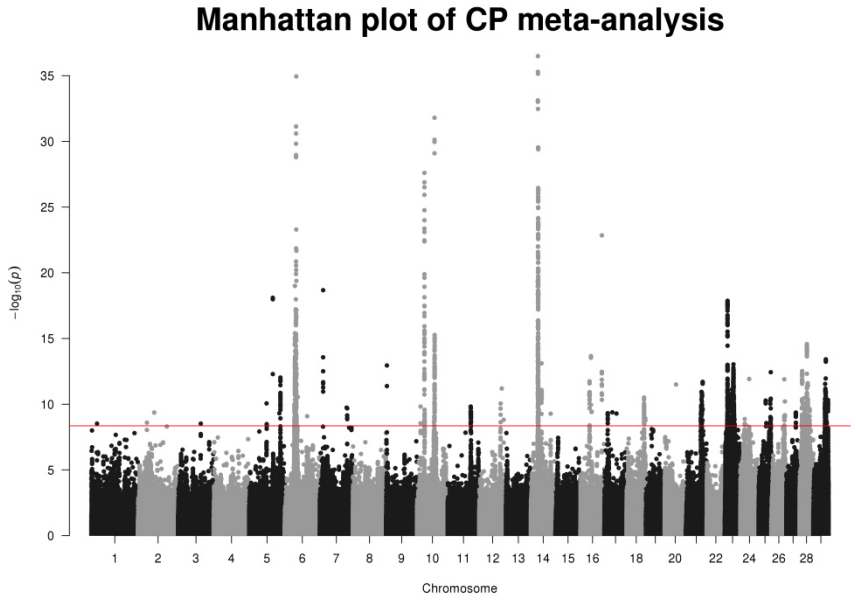

Figure 4. Manhattan plot of cross-phenotype (CP) meta-analysis for CGL, CGS, CS, and GI in Holstein, Jersey, and Red Dairy cattle. CGL = carcass gain with a long finishing period; CGS = carcass gain with a short finishing period; CS = carcass conformation score; GI = growth index.

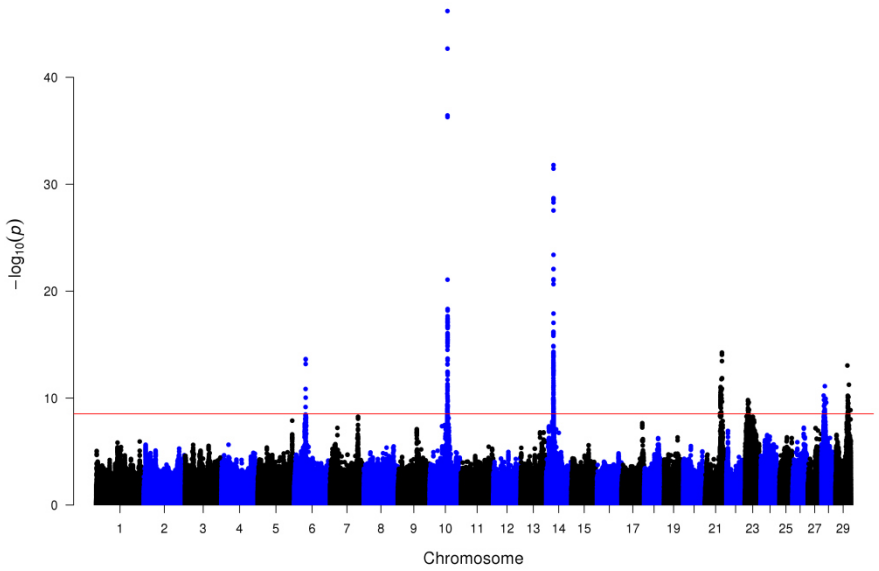

Figure 5. Manhattan plot of genome-wide -$\log_{10}$(*P*-values) for the growth index in multi-breed meta-analysis including Holstein, Jersey, and Red Dairy Cattle.

2.4 Increasing of cow data

Cow data could be utilized in both gene mapping and genomic prediction. In chapter 2 and 3, EBVs obtained from routine genetic evaluations or DRP were used as response variables. Thus, the non-additive part of the genetic architectures of calving traits and growth traits remain unexplored. In recent years, more and more cows are genotyped. This accumulation of cow genotype data enables us to perform GWAS for both the additive and dominance effects by using recordings on individual rather than EBVs. In chapter 5, female fertility data from 3,040 Holstein heifers and 4,483 Holstein cows was utilized to investigate on the additive and dominance effects. In the results, four QTL were detected for IFL in heifers and only one QTL was detected for IFL in cows. The simulations showed that the current sample size provides limited power to detect dominance effects for female fertility traits. With the accumulation of cow data, it is expected that the power to detect non-additive effects will increase. In addition, with the available imputation tools, genotypes of non-genotyped cows could also be imputed in future studies.

Increased cow data impacts not only the power of GWAS, but also GS. In most GS for dairy cattle populations, progeny tested bulls constitute the reference population. A sufficient size of the reference population is required to obtain accurate estimation of marker effects and thus reasonable genomic prediction accuracy (Lund *et al.*, 2011; Wiggans *et al.*, 2011). There are several solutions proposed to enlarge the reference population size: 1. genotypes are shared within and across countries (Lund *et al.*, 2011); 2. Information from different breeds are combined (Hayes *et al.*, 2009); 3. More genotyped cows are added in the reference population, especially for small population (Thomasen *et al.*, 2014). Increasing number of genotyped cows is both attractive and achievable. Some traits such as milk production, female fertility, and calving traits are only expressed in females and are already under routine recording. Thus, investigation on these cows could directly lead us to the underlying biology. In addition, the genotyping cost has decreased to be able to afford a large-scale genotyping on cows. Apart from the direct genotyping, more cows without genotypes or genotyped with low-density SNP chip could be imputed utilizing a rich resource of imputation tools (Marchini & Howie, 2010; Browning & Browning, 2013; Hickey & Kranis, 2013; Sargolzaei *et al.*, 2014). However, attention needs to be paid on the possible biased selection of cows to genotype, if only superior and elite dams are genotyped. Thus, more young cows where no selection has been carried on should also be genotyped.

2.5 Conclusions

GWAS have been widely implemented for dissecting the genetic architecture underlying complex traits in livestock and human, and numerous associations have been reported for several species. Our study has carried out GWAS with sequence-level variants, investigated on the 'missing' X chromosome from association studies and explored non-additive effects. The power of GWAS is expected to enhance in the future as more animals are accumulated with phenotype and genotype data. Imputation to WGS variants has become a standard practice due to the availability of reference sequences. This will greatly improve the resolution of gene mapping, especially for species with low LD in the genome, because in these species the causal mutations may be statistically separated from the nearby variants. More detailed genomic annotation data are also needed either to be directly incorporated in GWAS models, or to facilitate the explanations of GWAS findings in *post-hoc*. Inspired by human GWAS studies, GWAS consortiums could be established to combine data or summary statistics across countries or breeding organizations to increase the size of data for dairy cattle. In addition, as more and more traits are recorded in one individual from routine genetic evaluation, CP meta-analysis could be carried out to increase the power of GWAS. Follow-up functional studies are also necessary to confirm the association and finally to locate the causal mutations.

References

Aldrich, J. 1995. Correlations Genuine and Spurious in Pearson and Yule. Stat. Sci. 10:364–376. doi:10.2307/2246135.

Aliloo, H., J.E. Pryce, O. González-Recio, B.G. Cocks, and B.J. Hayes. 2015. Validation of markers with non-additive effects on milk yield and fertility in Holstein and Jersey cows. BMC Genet. 16:89. doi:10.1186/s12863-015-0241-9.

Andersson, L. 2016. Domestic animals as models for biomedical research. Ups. J. Med. Sci. 121:1–11. doi:10.3109/03009734.2015.1091522.

Aulchenko, Y.S., D.-J. de Koning, and C. Haley. 2007. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics. 177:577–85. doi:10.1534/genetics.107.075614.

Barendse, W., S.M. Armitage, L.M. Kossarek, a. Shalom, B.W. Kirkpatrick, a. M. Ryan, D. Clayton, L. Li, H.L. Neibergs, N. Zhang, W.M. Grosse, J. Weiss, P. Creighton, F. McCarthy, M. Ron, a. J. Teale, R. Fries, R. a. McGraw, S.S. Moore, M. Georges, M. Soller, J.E. Womack, and D.J.S. Hetzel. 1994. A genetic linkage map of the bovine genome. Nat. Genet. 6:227–235. doi:10.1038/ng0394-227.

Bellmann, O., J. Wegner, C. Rehfeldt, F. Teuscher, F. Schneider, J. Voigt, M. Derno, H. Sauerwein, J. Weingärtner, and K. Ender. 2004. Beef versus dairy cattle: a comparison of metabolically relevant hormones, enzymes, and metabolites. Livest. Prod. Sci. 89:41–54. doi:10.1016/j.livprodsci.2003.12.007.

Van den Berg, S.M., M.H.M. de Moor, K.J.H. Verweij, R.F. Krueger, M. Luciano, A. Arias Vasquez, L.K. Matteson, J. Derringer, T. Esko, N. Amin, S.D. Gordon, N.K. Hansell, A.B. Hart, I. Seppälä, J.E. Huffman, B. Konte, J. Lahti, M. Lee, M. Miller, T. Nutile, T. Tanaka, A. Teumer, A. Viktorin, J. Wedenoja, A. Abdellaoui, G.R. Abecasis, D.E. Adkins, A. Agrawal, J. Allik, K. Appel, T.B. Bigdeli, F. Busonero, H. Campbell, P.T. Costa, G.D. Smith, G. Davies, H. de Wit, J. Ding, B.E. Engelhardt, J.G. Eriksson, I.O. Fedko, L. Ferrucci, B. Franke, I. Giegling, R. Grucza, A.M. Hartmann, A.C. Heath, K. Heinonen, A.K. Henders, G. Homuth, J.-J. Hottenga, W.G. Iacono, J. Janzing, M. Jokela, R. Karlsson, J.P. Kemp, M.G. Kirkpatrick, A. Latvala, T. Lehtimäki, D.C. Liewald, P. a F. Madden, C. Magri, P.K.E. Magnusson, J. Marten, A. Maschio, H. Mbarek, S.E. Medland, E. Mihailov, Y. Milaneschi, G.W. Montgomery, M. Nauck, M.G. Nivard, K.G. Ouwens, A. Palotie, E. Pettersson, O. Polasek, Y. Qian, L. Pulkki-Råback, O.T. Raitakari, A. Realo, R.J. Rose, D. Ruggiero, C.O. Schmidt, W.S. Slutske, R. Sorice, J.M. Starr, B. St Pourcain, A.R. Sutin, N.J. Timpson, H. Trochet, S. Vermeulen, E. Vuoksimaa, E. Widen, J. Wouda, M.J. Wright, L. Zgaga, D. Porteous, A. Minelli, et al. 2015. Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. Behav. Genet. doi:10.1007/s10519-015-9735-5.

Bergelson, J., and F. Roux. 2010. Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. Nat. Rev. Genet. 11:867–879. doi:10.1038/nrg2896.

Bermingham, M.L., S.C. Bishop, J.A. Woolliams, R. Pong-Wong, A.R. Allen, S.H. McBride, J.J. Ryder, D.M. Wright, R.A. Skuce, S.W. McDowell, and E.J. Glass. 2014. Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. Heredity (Edinb). 112:543–551. doi:10.1038/hdy.2013.137.

Bhattacharjee, S., P. Rajaraman, K.B. Jacobs, W.A. Wheeler, B.S. Melin, P. Hartge, M. Yeager, C.C. Chung, S.J. Chanock, and N. Chatterjee. 2012. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am. J. Hum. Genet. 90:821–835. doi:10.1016/j.ajhg.2012.03.015.

Van Binsbergen, R., M.C. Bink, M.P. Calus, F. a van Eeuwijk, B.J. Hayes, I. Hulsegge, and R.F. Veerkamp. 2014. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. Genet. Sel. Evol. 46:41. doi:10.1186/1297-9686-46-41.

Boelling, D., U.S. Nielsen, J. Pösö, J.-å Eriksson, and G.P. Aamand. 2007. Genetic Evaluation of Calving Traits in Denmark , Finland and Sweden. INTERBULL Bull. no. 37, 2007. 179–185.

Boichard, D., F. Guillaume, A. Baur, P. Croiseau, M.N. Rossignol, M.Y. Boscher, T. Druet, L. Genestout, J.J. Colleau, L. Journaux, V. Ducrocq, and S. Fritz. 2012. Genomic selection in French dairy cattle. Anim. Prod. Sci. 52:115–120. doi:10.1071/AN11119.

Bolormaa, S., B.J. Hayes, J.H.J. van der Werf, D. Pethick, M.E. Goddard, and H.D. Daetwyler. 2016. Detailed phenotyping identifies genes with pleiotropic effects on body composition. BMC Genomics. 17:224. doi:10.1186/s12864-016-2538-0.

Bolormaa, S., J.E. Pryce, A. Reverter, Y. Zhang, W. Barendse, K. Kemper, B. Tier, K. Savin, B.J. Hayes, and M.E. Goddard. 2014. A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. PLoS Genet. 10:e1004198. doi:10.1371/journal.pgen.1004198.

Bolormaa, S., J.E. Pryce, Y. Zhang, A. Reverter, W. Barendse, B.J. Hayes, and M.E. Goddard. 2015. Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. Genet. Sel. Evol. 47:26. doi:10.1186/s12711-015-0114-8.

Bouwman, A.C., J.M. Hickey, M.P. Calus, and R.F. Veerkamp. 2014. Imputation of non-genotyped individuals based on genotyped relatives: assessing the imputation accuracy of a real case scenario in dairy cattle. Genet. Sel. Evol. 46:6. doi:10.1186/1297-9686-46-6.

Brinkman-Van Der Linden, E.C.M., 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. 17:922–931. doi:10.1093/glycob/cwm065.

Brøndum, R.F., B. Guldbrandtsen, G. Sahana, M.S. Lund, and G. Su. 2014. Strategies for imputation to whole genome sequence using a single or multi-breed reference population in cattle. BMC Genomics. 15:728. doi:10.1186/1471-2164-15-728.

Brøndum, R.F., G. Su, L. Janss, G. Sahana, B. Guldbrandtsen, D. Boichard, and M.S. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. J. Dairy Sci. doi:10.3168/jds.2014-9005.

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. Genet. 84:210–223. doi:10.1016/j.ajhg.2009.01.005.

Browning, B.L., and S.R. Browning. 2013. Improving the Accuracy and Efficiency of Identity by Descent Detection in Population Data. Genetics. 1–36. doi:10.1534/genetics.113.150029.

Browning, B.L., and S.R. Browning. 2016. Genotype Imputation with Millions of Reference Samples. Am. J. Hum. Genet. 98:116–126. doi:10.1016/j.ajhg.2015.11.020.

De Camargo, G.M.F., L.R. Porto-Neto, M.J. Kelly, R.J. Bunch, S.M. McWilliam, H. Tonhati, S. a Lehnert, M.R.S. Fortes, and S.S. Moore. 2015. Non-synonymous mutations mapped to chromosome X associated with andrological and growth traits in beef cattle. BMC Genomics. 16:384. doi:10.1186/s12864-015-1595-0.

Chatenoud, L., K. Warncke, and A.-G. Ziegler. 2012. Clinical immunologic interventions for the treatment of type 1 diabetes. Cold Spring Harb. Perspect. Med. 2:a007716. doi:10.1101/cshperspect.a007716.

Chwalisz, M., and R. Fürbass. 2014. Evaluation of coding-independent functions of the transcribed bovine aromatase pseudogene CYP19P1. BMC Res. Notes. 7:378. doi:10.1186/1756-0500-7-378.

Claussnitzer, M., S.N. Dankel, K.-H. Kim, G. Quon, W. Meuleman, C. Haugen, V. Glunk, I.S. Sousa, J.L. Beaudry, V. Puviindran, N. a Abdennur, J. Liu, P.-A. Svensson, Y.-H. Hsu, D.J. Drucker, G. Mellgren, C.-C. Hui, H. Hauner, and M. Kellis. 2015. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N. Engl. J. Med. 373:895–907. doi:10.1056/NEJMoa1502214.

Clayton, D. 2008. Testing for association on the X chromosome. Biostatistics. 9:593–600. doi:10.1093/biostatistics/kxn007.

Cleveland, M.A., and N. Deeb. 2009. Evaluation of a genome-wide approach to multiple marker association considering different marker densities. BMC Proc. 3:S5–S5.

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. 92:2931–46. doi:10.3168/jds.2008-1762.

Cole, J.B., B. Waurich, M. Wensch-Dorendorf, D.M. Bickhart, and H.H. Swalve. 2014. A genome-wide association study of calf birth weight in Holstein cattle using single nucleotide polymorphisms and phenotypes predicted from auxiliary traits. J. Dairy Sci. 97:3156–72. doi:10.3168/jds.2013-7409.

Cole, J.B., G.R. Wiggans, L. Ma, T.S. Sonstegard, T.J. Lawlor, B. a Crooker, C.P. Van Tassell, J. Yang, S. Wang, L.K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 12:408. doi:10.1186/1471-2164-12-408.

Conneally, P.M., and W.H. Stone. 1965. Association between a Blood Group and Butterfat Production in Dairy Cattle. Nature. 206:115.

Consortium, 1000 Genomes Project, others, and W. Africa. 2010. A map of human genome variation from population-scale sequencing. Nature. 467:1061–73. doi:10.1038/nature09534.

Cooper, H., and L. V Hedges. 2009. Research synthesis as a scientific process. Handb. Res. Synth. meta-analysis. 3–17.

Cotsapas, C., B.F. Voight, E. Rossin, K. Lage, B.M. Neale, C. Wallace, G.R. AbecasisGon??, J.C. Barrett, T. Behrens, J. Cho, P.L. De Jager, J.T. Elder, R.R. Graham, P. Gregersen, L. Klareskog, K.A. Siminovitch, D.A. van Heel, C. Wijmenga, J. Worthington, J.A. Todd, D.A. Hafler, S.S. Rich, and M.J. Daly. 2011. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 7. doi:10.1371/journal.pgen.1002254.

Daetwyler, H.D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R.F. Brøndum, X. Liao, A. Djari, S.C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P.J. Bowman, D. Coote, A.J. Chamberlain, C. Anderson, C.P. VanTassell, I. Hulsegge, M.E. Goddard, B. Guldbrandtsen, M.S. Lund, R.F. Veerkamp, D. a Boichard, R. Fries, and B.J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46:858–865. doi:10.1038/ng.3034.

Daetwyler, H.D., G.R. Wiggans, B.J. Hayes, J. a Woolliams, and M.E. Goddard. 2011. Imputation of missing genotypes from sparse to high density using long-range phasing. Genetics. 189:317–27. doi:10.1534/genetics.111.128082.

Das, P.J., B.P. Chowdhary, and T. Raudsepp. 2009. Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. Cytogenet. Genome Res. 126:139–147. doi:10.1159/000245913.

Dashab, G.R., N.K. Kadri, M.M. Shariati, and G. Sahana. 2012. Comparison of linear mixed model analysis and genealogy-based haplotype clustering with a Bayesian approach for association mapping in a pedigreed population. BMC Proc. 6:S4. doi:10.1186/1753-6561-6-S2-S4.

DerSimonian, R., and N. Laird. 1986. Meta-analysis in clinical trials. Control. Clin. Trials. 7:177–88. doi:10.1016/0197-2456(86)90046-2.

Duijvesteijn, N., E.F. Knol, J.W.M. Merks, R.P.M. a Crooijmans, M. a M. Groenen, H. Bovenhuis, and B. Harlizius. 2010. A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC Genet. 11:42. doi:10.1186/1471-2156-11-42.

Ekine, C.C., S.J. Rowe, S.C. Bishop, and D.-J. de Koning. 2014. Why breeding values estimated using familial data should not be used for genome-wide association studies. G3 (Bethesda). 4:341–7. doi:10.1534/g3.113.008706.

Elo, K.T., J. Vilkki, D.J. De Koning, R.J. Velmala, and A. V. Mäki-Tanila. 1999. A quantitative trait locus for live weight maps to bovine chromosome 23. Mamm. Genome. 10:831–835. doi:10.1007/s003359901098.

Evangelou, E., and J.P. a Ioannidis. 2013. Meta-analysis methods for genome-wide association studies and beyond. Nat. Rev. Genet. 14:379–89. doi:10.1038/nrg3472.

Fisher, R. 1925. Statistical methods for research workers. xv, 356 p. pp.

Flicek, P., M.R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C.G. Girón, L. Gordon, T. Hourlier, S. Hunt, N. Johnson, T. Juettemann, A.K. Kähäri, S. Keenan, E. Kulesha, F.J. Martin, T. Maurel, W.M. McLaren, D.N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H.S. Riat, M. Ruffier, D. Sheppard, K. Taylor, A. Thormann, S.J. Trevanion, A. Vullo, S.P. Wilder, M. Wilson, A. Zadissa, B.L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T.J.P. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D.R. Zerbino, and S.M.J. Searle. 2014. Ensembl 2014. Nucleic Acids Res. 42. doi:10.1093/nar/gkt1196.

Fries, R., A. Eggen, and G. Stranzinger. 1990. The bovine genome contains polymorphic microsatellites. Genomics. 8:403–406. doi:10.1016/0888-7543(90)90301-A.

Fries, R., A. Eggen, and J.E. Womack. 1993. The bovine genome map. Mamm. Genome. 4:405–428. doi:10.1007/BF00296815.

Gaj, T., C. a. Gersbach, and C.F. Barbas. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31:397–405. doi:10.1016/j.tibtech.2013.04.004.

Gao, H., M.S. Lund, Y. Zhang, and G. Su. 2013. Accuracy of genomic prediction using different models and response variables in the Nordic Red cattle population. J. Anim. Breed. Genet. 130:333–340. doi:10.1111/jbg.12039.

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. doi:10.1186/1297-9686-41-55.

Gengler, N., L.D. Van Vleck, M.D. MacNeil, I. Misztal, and F.A. Pariacote. 1997. Influence of Dominance Relationships on the Estimation of Dominance Variance with Sire-Dam Subclass Effects. J. Anim. Sci. 75:2885–2891.

George, E.I., and R.E. McCulloch. 1993. Variable Selection via Gibbs Sampling. J. Am. Stat. Assoc. 88:881–889. doi:10.1080/01621459.1993.10476353.

Goodwin, S., J.D. McPherson, and W.R. McCombie. 2016. Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17:333–351. doi:10.1038/nrg.2016.49.

Grobet, L., L.J. Martin, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, a Schoeberlein, S. Dunner, F. Ménissier, J. Massabanda, R. Fries, R. Hanset, and M. Georges. 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat. Genet. 17:71–74. doi:10.1038/ng0997-71.

Gudbjartsson, D.F., G.B. Walters, G. Thorleifsson, H. Stefansson, B. V Halldorsson, P. Zusmanovich, P. Sulem, S. Thorlacius, A. Gylfason, S. Steinberg, A. Helgadottir, A. Ingason, V. Steinthorsdottir, E.J. Olafsdottir, G.H. Olafsdottir, T. Jonsson, K. Borch-Johnsen, T. Hansen, G. Andersen, T. Jorgensen, O. Pedersen, K.K. Aben, J.A. Witjes, D.W. Swinkels, M. den Heijer, B. Franke, A.L.M. Verbeek, D.M. Becker, L.R. Yanek, L.C. Becker, L. Tryggvadottir, T. Rafnar, J. Gulcher, L.A. Kiemeney, A. Kong, U. Thorsteinsdottir, and K. Stefansson. 2008. Many sequence variants affecting diversity of adult human height. Nat. Genet. 40:609–615. doi:10.1038/ng.122.

Guo, G., M.S. Lund, Y. Zhang, and G. Su. 2010. Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. J. Anim. Breed. Genet. 127:423–432. doi:10.1111/j.1439-0388.2010.00878.x.

Hayes, B.J., P.J. Bowman, A.C. Chamberlain, K. Verbyla, and M.E. Goddard. 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Sel. Evol. 41:51. doi:10.1186/1297-9686-41-51.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, K. Savin, C.P. van Tassell, T.S. Sonstegard, and M.E. Goddard. 2009b. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. PLoS One. 4. doi:10.1371/journal.pone.0006676.

Heard, E., and J. Turner. 2011. Function of the Sex Chromosomes in Mammalian Fertility. Cold Spring Harb. Perspect. Biol. 1–18.

Heine, P.A., J.A. Taylor, G.A. Iwamoto, D.B. Lubahn, and P.S. Cooke. 2000. Increased adipose tissue in male and female estrogen receptor-alpha knockout mice. Proc. Natl. Acad. Sci. U. S. A. 97:12729–12734. doi:10.1073/pnas.97.23.12729.

Helbig, I., H.C. Mefford, A.J. Sharp, M. Guipponi, M. Fichera, A. Franke, H. Muhle, C. de Kovel, C. Baker, S. von Spiczak, K.L. Kron, I. Steinich, A. a Kleefuss-Lie, C. Leu, V. Gaus, B. Schmitz, K.M. Klein, P.S. Reif, F. Rosenow, Y. Weber, H. Lerche, F. Zimprich, L. Urak, K.

Fuchs, M. Feucht, P. Genton, P. Thomas, F. Visscher, G.-J. de Haan, R.S. Møller, H. Hjalgrim, D. Luciano, M. Wittig, M. Nothnagel, C.E. Elger, P. Nürnberg, C. Romano, A. Malafosse, B.P.C. Koeleman, D. Lindhout, U. Stephani, S. Schreiber, E.E. Eichler, T. Sander, S. Girirajan, J. a Rosenfeld, B.P. Coe, S. Parikh, N. Friedman, A. Goldstein, R. a Filipink, J.S. McConnell, B. Angle, W.S. Meschino, M.M. Nezarati, A. Asamoah, K.E. Jackson, G.C. Gowans, J. a Martin, E.P. Carmany, D.W. Stockton, R.E. Schnur, L.S. Penney, D.M. Martin, S. Raskin, K. Leppig, H. Thiese, R. Smith, E. Aberg, D.M. Niyazov, L.F. Escobar, D. El-Khechen, K.K.D. Johnson, R.R. Lebel, K. Siefkas, S. Ball, N. Shur, M. McGuire, C.K. Brasington, J.E. Spence, L.S. Martin, C. Clericuzio, B.C. Ballif, L.G. Shaffer, E.E. Eichler, D. Malhotra, J. Sebat, M. Tropeano, J.W. Ahn, R.J.B. Dobson, G. Breen, J. Rucker, A. Dixit, D.K. Pal, P. McGuffin, A. Farmer, P.S. White, J. Andrieux, et al. 2013. The penetrance of copy number variations for schizophrenia and developmental delay. Nat. Genet. 43:749–64. doi:10.1016/j.ajhg.2010.04.006.

Hickey, J.M., M.G. Keane, D. a Kenny, a R. Cromie, and R.F. Veerkamp. 2007. Genetic parameters for EUROP carcass traits within different groups of cattle in Ireland. J. Anim. Sci. 85:314–21. doi:10.2527/jas.2006-263.

Hickey, J.M., B.P. Kinghorn, B. Tier, J.F. Wilson, N. Dunstan, and J.H.J. van der Werf. 2011. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. Genet. Sel. Evol. 43:12. doi:10.1186/1297-9686-43-12.

Hickey, J.M., and a Kranis. 2013. Extending long-range phasing and haplotype library imputation methods to impute genotypes on sex chromosomes. Genet Sel Evol. 45:10. doi:10.1186/1297-9686-45-10.

Hill, E.W., B. a McGivney, J. Gu, R. Whiston, and D.E. Machugh. 2010. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. BMC Genomics. 11:552. doi:10.1186/1471-2164-11-552.

Hill, W.G., M.E. Goddard, and P.M. Visscher. 2008. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4. doi:10.1371/journal.pgen.1000008.

Hillier, L.W., W. Miller, E. Birney, W. Warren, R.C. Hardison, C.P. Ponting, P. Bork, D.W. Burt, M.A.M. Groenen, and M.E. Delany. 2004. Sequencing and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 432:695 – 716. doi:10.1038/nature03394.

Hines, H.C., C.A. Kiddy, E.W. Brum, and C.W. Arave. 1969. Linkage among Cattle Blood and Milk Polymorphisms. Genetics. 62:401–412.

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

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. doi:10.1186/1471-2156-13-71.

Höglund, J.K., G. Sahana, R.F. Brøndum, B. Guldbrandtsen, B. Buitenhuis, and M.S. Lund. 2014a. Fine mapping QTL for female fertility on BTA04 and BTA13 in dairy cattle using HD SNP and sequence data. BMC Genomics. 15:790. doi:10.1186/1471-2164-15-790.

Höglund, J.K., G. Sahana, R.F. Brøndum, B. Guldbrandtsen, B. Buitenhuis, and M.S. Lund. 2014b. Fine mapping QTL for female fertility on BTA04 and BTA13 in dairy cattle using HD SNP and sequence data. BMC Genomics. 15:790. doi:10.1186/1471-2164-15-790.

Hou, Y., P. Madsen, R. Labouriau, Y. Zhang, M.S. Lund, and G. Su. 2009. Genetic analysis of days from calving to first insemination and days open in Danish Holsteins using different models and censoring scenarios. J. Dairy Sci. 92:1229–1239. doi:10.3168/jds.2008-1556.

Ioannidis, J.P.A., T.A. Trikalinos, and M.J. Khoury. 2006. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. Am. J. Epidemiol. 164:609–614. doi:10.1093/aje/kwj259.

Israel, C., and J.I. Wellert. 2002. Estimation of quantitative trait loci effects in dairy cattle populations. J. Dairy Sci. 85:1285–97. doi:10.3168/jds.S0022-0302(02)74193-3.

Johansson, K., U.S. Nielsen, G.P. Aamand, D. Agricultural, A. Service, N. Cattle, and G. Evaluation. 2008. Joint Nordic Genetic Evaluation of Growth and Carcass traits in Dairy Breeds. 49–54.

Kang, H.M., J.H. Sul, S.K. Service, N.A. Zaitlen, S.-Y. Kong, N.B. Freimer, C. Sabatti, and E. Eskin. 2010. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42:348–54. doi:10.1038/ng.548.

Kappes, S.M., J.W. Keele, R.T. Stone, R.A. McGraw, T.S. Sonstegard, T.P.L. Smith, N.L. Lopez-Corrales, and C.W. Beattie. 1997. A second-generation linkage map of the bovine genome. Genome Res. 7:235–249. doi:10.1101/gr.7.3.235.

Kavvoura, F.K., and J.P.A. Ioannidis. 2008. Methods for meta-analysis in genetic association studies: A review of their potential and pitfalls. Hum. Genet. 123:1–14. doi:10.1007/s00439-007-0445-9.

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. 86:360–8.

Lee, S.H., B.H. Choi, D. Lim, C. Gondro, Y.M. Cho, C.G. Dang, A. Sharma, G.W. Jang, K.T. Lee, D. Yoon, H.K. Lee, S.H. Yeon, B.S. Yang, H.S. Kang, and S.K. Hong. 2013. Genome-wide

association study identifies major loci for carcass weight on BTA14 in Hanwoo (Korean cattle). PLoS One. 8:e74677. doi:10.1371/journal.pone.0074677.

Lettre, G., A.U. Jackson, C. Gieger, F.R. Schumacher, S.I. Berndt, S. Sanna, S. Eyheramendy, B.F. Voight, J.L. Butler, C. Guiducci, T. Illig, R. Hackett, I.M. Heid, K.B. Jacobs, V. Lyssenko, M. Uda, M. Boehnke, S.J. Chanock, L.C. Groop, F.B. Hu, B. Isomaa, P. Kraft, L. Peltonen, V. Salomaa, D. Schlessinger, D.J. Hunter, R.B. Hayes, G.R. Abecasis, H.-E. Wichmann, K.L. Mohlke, and J.N. Hirschhorn. 2008. Identification of ten loci associated with height highlights new biological pathways in human growth. Nat. Genet. 40:584–591. doi:10.1038/ng.125.

Lin, H.K., P.A. Oltenacu, L.D. Van Vleck, H.N. Erb, and R.D. Smith. 1989. Heritabilities of and genetic correlations among six health problems in Holstein cows. J. Dairy Sci. 72:180–186.

Lopes, M.S., J.W.M. Bastiaansen, B. Harlizius, E.F. Knol, and H. Bovenhuis. 2014. A genome-wide association study reveals dominance effects on number of teats in pigs. PLoS One. 9:1–8. doi:10.1371/journal.pone.0105867.

Lucy, M.C. 2001. Reproductive Loss in High-Producing Dairy Cattle: Where Will It End? J. Dairy Sci. 84:1277–1293. doi:10.3168/jds.S0022-0302(01)70158-0.

Lund, M.S., S.P. de Ross, A.G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43:43. doi:10.1186/1297-9686-43-43.

Lyons, R.E., N.T. Loan, L. Dierens, M.R.S. Fortes, M. Kelly, S.S. McWilliam, Y. Li, R.J. Bunch, B.E. Harrison, W. Barendse, S. a Lehnert, and S.S. Moore. 2014. Evidence for positive selection of taurine genes within a QTL region on chromosome X associated with testicular size in Australian Brahman cattle. BMC Genet. 15:6. doi:10.1186/1471-2156-15-6.

Mahajan, A., M.J. Go, W. Zhang, J.E. Below, K.J. Gaulton, T. Ferreira, M. Horikoshi, A.D. Johnson, M.C.Y. Ng, I. Prokopenko, D. Saleheen, X. Wang, E. Zeggini, G.R. Abecasis, L.S. Adair, P. Almgren, M. Atalay, T. Aung, D. Baldassarre, B. Balkau, Y. Bao, A.H. Barnett, I. Barroso, A. Basit, L.F. Been, J. Beilby, G.I. Bell, R. Benediktsson, R.N. Bergman, B.O. Boehm, E. Boerwinkle, L.L. Bonnycastle, N. Burtt, Q. Cai, H. Campbell, J. Carey, S. Cauchi, M. Caulfield, J.C.N. Chan, L.-C. Chang, T.-J. Chang, Y.-C. Chang, G. Charpentier, C.-H. Chen, H. Chen, Y.-T. Chen, K.-S. Chia, M. Chidambaram, P.S. Chines, N.H. Cho, Y.M. Cho, L.-M. Chuang, F.S. Collins, M.C. Cornelis, D.J. Couper, A.T. Crenshaw, R.M. van Dam, J. Danesh, D. Das, U. de Faire, G. Dedoussis, P. Deloukas, A.S. Dimas, C. Dina, A.S. Doney, P.J. Donnelly, M. Dorkhan, C. van Duijn, J. Dupuis, S. Edkins, P. Elliott, V. Emilsson, R. Erbel, J.G. Eriksson, J. Escobedo, T. Esko, E. Eury, J.C. Florez, P. Fontanillas, N.G. Forouhi, T. Forsen, C. Fox, R.M. Fraser, T.M. Frayling, P. Froguel, P. Frossard, Y. Gao, K. Gertow, C. Gieger, B. Gigante, H. Grallert, G.B. Grant, L.C. Grrop, C.J. Groves, E. Grundberg, C. Guiducci, A. Hamsten, B.-G. Han, et al. 2014. Genome-wide trans-ancestry meta-analysis

provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat. Genet. 46:234–44. doi:10.1038/ng.2897.

Manolio, T.A., and E.D. Green. 2011. Genomics reaches the clinic: From basic discoveries to clinical impact. Cell. 147:14–16. doi:10.1016/j.cell.2011.09.012.

Mao, X., A.M. Johansson, G. Sahana, B. Guldbrandtsen, and D.-J. de Koning. 2016a. Imputation of markers on the Bovine X chromosome. J. Dairy Sci. 833:1–6. doi:10.3168/jds.2016-11160.

Mao, X., N.K. Kadri, J.R. Thomasen, D.J. De Koning, G. Sahana, and B. Guldbrandtsen. 2015. Fine mapping of a calving QTL on Bos taurus autosome 18 in Holstein cattle. J. Anim. Breed. Genet. 133:207–218. doi:10.1111/jbg.12187.

Mao, X., G. Sahana, D.-J. De Koning, and B. Guldbrandtsen. 2016b. Genome-wide association studies of growth traits in three dairy cattle breeds using whole-genome sequence data. J. Anim. Sci. 94:1426. doi:10.2527/jas.2015-9838.

Marchini, J., and B. Howie. 2010. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11:499–511. doi:10.1038/nrg2796.

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.L. 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:e5350. doi:10.1371/journal.pone.0005350.

McCarthy, M.I., G.R. Abecasis, L.R. Cardon, D.B. Goldstein, J. Little, J.P.A. Ioannidis, and J.N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. 9:356–69. doi:10.1038/nrg2344.

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

Muers, M. 2009. Genomics: Milking the cow genome. Nat. Rev. Genet. 10:346–347. doi:10.1038/nrg2598.

Muñoz, P.R., M.F.R. Resende, S.A. Gezan, M.D.V. Resende, G. de los Campos, M. Kirst, D. Huber, and G.F. Peter. 2014. Unraveling additive from nonadditive effects using genomic relationship matrices. Genetics. 198:1759–1768. doi:10.1534/genetics.114.171322.

Nicolae, D.L. 2006. Testing untyped alleles (TUNA)-applications to genome-wide association studies. Genet. Epidemiol. 30:718–727. doi:10.1002/gepi.

Nicolazzi, E.L., S. Biffani, and G. Jansen. 2013. Short communication: imputing genotypes using PedImpute fast algorithm combining pedigree and population information. J. Dairy Sci. 96:2649–53. doi:10.3168/jds.2012-6062.

Norris, D., L. Varona, J.W. Ngambi, D.P. Visser, C.A. Mbajiorgu, and S.F. Voordewind. 2010. Estimation of the additive and dominance variances in SA Duroc pigs. Livest. Sci. 131:144–147. doi:10.1016/j.livsci.2010.03.005.

Öz, O.K., R. Millsaps, R. Welch, J. Birch, and J.E. Zerwekh. 2001. Expression of aromatase in the human growth plate. J. Mol. Endocrinol. 27:249–253. doi:10.1677/jme.0.0270249.

Palucci, V., L.R. Schaeffer, F. Miglior, and V. Osborne. 2007. Non-additive genetic effects for fertility traits in Canadian Holstein cattle (Open Access publication). Genet. Sel. Evol. 39:181–193. doi:10.1186/1297-9686-39-2-181.

Pereira, T. V., N.A. Patsopoulos, G. Salanti, and J.P.A. Ioannidis. 2009. Discovery properties of genome-wide association signals from cumulatively combined data sets. Am. J. Epidemiol. 170:1197–1206. doi:10.1093/aje/kwp262.

Pfeiffer, R.M., M.H. Gail, and D. Pee. 2009. On Combining Data From Genome-Wide Association Studies to Discover Disease-Associated SNPs. Stat. Sci. 24:547–560. doi:10.1214/09-STS286.

Pryce, J.E., B.J. Hayes, S. Bolormaa, and M.E. Goddard. 2011. Polymorphic regions affecting human height also control stature in cattle. Genetics. 187:981–4. doi:10.1534/genetics.110.123943.

Pryce, J.E., J. Johnston, B.J. Hayes, G. Sahana, K. a Weigel, S. McParland, D. Spurlock, N. Krattenmacher, R.J. Spelman, E. Wall, and M.P.L. Calus. 2014. Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. J. Dairy Sci. 97:1799–811. doi:10.3168/jds.2013-7368.

Replication, D.G., M.D. Consortium, and A.G. Epidemiology. 2014. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat. Genet. 46:234–44. doi:10.1038/ng.2897.

De Roos, A.P.W., B.J. Hayes, R.J. Spelman, and M.E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics. 179:1503–1512. doi:10.1534/genetics.107.084301.

Roxström, A., E. Strandberg, B. Berglund, U. Emanuelson, and J. Philipsson. 2001. Genetic and Environmental Correlations Among Female Fertility Traits, and Between the Ability to Show Oestrus and Milk Production in Dairy Cattle. Acta Agric. Scand. 51:192–199. doi:10.1080/090647001750367285.

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

Sahana, G., B. Guldbrandtsen, and M.S. Lund. 2011. Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. J. Dairy Sci. 94:479–86. doi:10.3168/jds.2010-3381.

Sahana, G., J.K. Höglund, B. Guldbrandtsen, and M.S. Lund. 2015. Loci associated with adult stature also affect calf birth survival in cattle. BMC Genet. 16:47. doi:10.1186/s12863-015-0202-3.

Sandor, C., F. Farnir, S. Hansoul, W. Coppieters, T. Meuwissen, and M. Georges. 2006. Linkage disequilibrium on the bovine X chromosome: Characterization and use in quantitative trait locus mapping. Genetics. 173:1777–1786. doi:10.1534/genetics.106.059329.

Sargolzaei, M., J.P. Chesnais, and F.S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 15:478. doi:10.1186/1471-2164-15-478.

Satagopan, J.M., B.S. Yandell, M.A. Newton, and T.C. Osborn. 1996. A bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. Genetics. 144:805–816. doi:10.1186/1753-6561-5-S3-S4.

Saxena, R., B.F. Voight, V. Lyssenko, N.P. Burtt, P.I.W. de Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly, T.E. Hughes, L. Groop, D. Altshuler, P. Almgren, J.C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Boström, B. Isomaa, G. Lettre, U. Lindblad, H.N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Råstam, E.K. Speliotes, M.-R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjögren, M. Sterner, A. Surti, M. Svensson, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S.B. Gabriel, G.-W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke, and S. Purcell. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 316:1331–6. doi:10.1126/science.1142358.

SEGES. 2015. Årsstatistik Avl - 2014/15 (Annual Statistics Breeding - 2014/15), 30th edition.

Setoguchi, K., M. Furuta, T. Hirano, T. Nagao, T. Watanabe, Y. Sugimoto, and A. Takasuga. 2009. Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate. BMC Genet. 10:43. doi:10.1186/1471-2156-10-43.

Solovieff, N., C. Cotsapas, P.H. Lee, S.M. Purcell, and J.W. Smoller. 2013. Pleiotropy in complex traits: challenges and strategies. Nat. Rev. Genet. 14:483–95. doi:10.1038/nrg3461.

Stearns, F.W. 2010. One hundred years of pleiotropy: A retrospective. Genetics. 186:767–773. doi:10.1534/genetics.110.122549.

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. 86:2228–2235.

Stephens, M., and D.J. Balding. 2009. Bayesian statistical methods for genetic association studies. Nat Rev Genet. 10:681–690. doi:10.1038/nrg2615.

Su, G., O.F. Christensen, T. Ostersen, M. Henryon, and M.S. Lund. 2012. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. PLoS One. 7:1–7. doi:10.1371/journal.pone.0045293.

Su, G., B. Guldbrandtsen, G.P. Aamand, I. Strandén, and M.S. Lund. 2014. Genomic relationships based on X chromosome markers and accuracy of genomic predictions with and without X chromosome markers. Genet. Sel. Evol. 46:47. doi:10.1186/1297-9686-46-47.

Sun, C., P. Madsen, U.S. Nielsen, Y. Zhang, M.S. Lund, and G. Su. 2009. Comparison between a sire model and an animal model for genetic evaluation of fertility traits in Danish Holstein population. J. Dairy Sci. 92:4063–71. doi:10.3168/jds.2008-1918.

Sveinbjornsson, G., A. Albrechtsen, F. Zink, S. a. Gudjonsson, A. Oddson, G. Másson, H. Holm, A. Kong, U. Thorsteinsdottir, P. Sulem, D.F. Gudbjartsson, K. Stefansson, a. Oddsson, G. Masson, A. Kong, U. Thorsteinsdottir, P. Sulem, D.F. Gudbjartsson, and K. Stefansson. 2016. Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nat. Genet. Submitted:314–317. doi:10.1038/ng.3507.

The Bovine Genome Sequencing and Analysis Consortium;, and K.C. Elsik, C. G.; Tellam, R. L.; Worley. 2009. The Genome Sequence of Taurine. Science (80-.). 324:522–8. doi:10.1126/science.1169588.

Thomas, G., G. Thomas, K.B. Jacobs, K.B. Jacobs, M. Yeager, M. Yeager, P. Kraft, P. Kraft, S. Wacholder, S. Wacholder, N. Orr, N. Orr, K. Yu, K. Yu, N. Chatterjee, N. Chatterjee, R. Welch, R. Welch, A. Hutchinson, A. Hutchinson, A. Crenshaw, A. Crenshaw, G. Cancel-Tassin, G. Cancel-Tassin, B.J. Staats, B.J. Staats, Z. Wang, Z. Wang, J. Gonzalez-Bosquet, J. Gonzalez-Bosquet, J. Fang, J. Fang, X. Deng, X. Deng, S.I. Berndt, S.I. Berndt, E.E. Calle, E.E. Calle, H.S. Feigelson, H.S. Feigelson, M.J. Thun, M.J. Thun, C. Rodriguez, C. Rodriguez, D. Albanes, D. Albanes, J. Virtamo, J. Virtamo, S. Weinstein, S. Weinstein, F.R. Schumacher, F.R. Schumacher, E. Giovannucci, E. Giovannucci, W.C. Willett, W.C. Willett, O. Cussenot, O. Cussenot, A. Valeri, A. Valeri, G.L. Andriole, G.L. Andriole, E.D. Crawford, E.D. Crawford, M. Tucker, M. Tucker, D.S. Gerhard, D.S. Gerhard, J.F. Fraumeni, J.F. Fraumeni, R. Hoover, R. Hoover, R.B. Hayes, R.B. Hayes, D.J. Hunter, D.J. Hunter, S.J. Chanock, and S.J. Chanock. 2008. Multiple loci identified in a genome-wide association study of prostate cancer. Nat. Genet. 40:310–5. doi:10.1038/ng.91.

Thomasen, J.R., B. Guldbrandtsen, P. Sørensen, B. Thomsen, and M.S. Lund. 2008. Quantitative trait loci affecting calving traits in Danish Holstein cattle. J. Dairy Sci. 91:2098–105. doi:10.3168/jds.2007-0602.

Thomasen, J.R., a C. Sørensen, M.S. Lund, and B. Guldbrandtsen. 2014. Adding cows to the reference population makes a small dairy population competitive. J. Dairy Sci. 97:5822–32. doi:10.3168/jds.2014-7906.

Tibshirani, R. 1996. Regression Selection and Shrinkage via the Lasso. J. R. Stat. Soc. B. 58:267–288. doi:10.2307/2346178.

Utsunomiya, Y.T., A.S. Do Carmo, R. Carvalheiro, H.H. Neves, M.C. Matos, L.B. Zavarez, A.M. Pérez O'Brien, J. Sölkner, J.C. McEwan, J.B. Cole, C.P. Van Tassell, F.S. Schenkel, M.V. Da Silva, L.R. Porto Neto, T.S. Sonstegard, and J.F. Garcia. 2013. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. BMC Genet. 14:52. doi:10.1186/1471-2156-14-52.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–23. doi:10.3168/jds.2007-0980.

VanRaden, P.M., D.J. Null, M. Sargolzaei, G.R. Wiggans, M.E. Tooker, J.B. Cole, T.S. Sonstegard, E.E. Connor, M. Winters, J.B.C.H.M. van Kaam, a Valentini, B.J. Van Doormaal, M. a Faust, and G. a Doak. 2013. Genomic imputation and evaluation using high-density Holstein genotypes. J. Dairy Sci. 96:668–78. doi:10.3168/jds.2012-5702.

VanRaden, P.M., and G.R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737–2746. doi:10.3168/jds.S0022-0302(91)78453-1.

Wendorf, G.L., M.S. Lawyer, and N.L. First. 1983. Role of the adrenals in the maintenance of pregnancy in cows. J. Reprod. Fertil. 68:281–287. doi:10.1530/jrf.0.0680281.

Wiggans, G.R., P.M. Vanraden, and T.A. Cooper. 2011. The genomic evaluation system in the United States: past, present, future. J. Dairy Sci. 94:3202–11. doi:10.3168/jds.2010-3866.

Williams, J.L., A. Eggen, L. Ferretti, C.J. Farr, M. Gautier, G. Amati, G. Ball, T. Caramorr, R. Critcher, S. Costa, P. Hextall, D. Hills, A. Jeulin, S.L. Kiguwa, O. Ross, A.L. Smith, K. Saunier, B. Urquhart, and D. Waddington. 2002. A bovine whole-genome radiation hybrid panel and outline map. Mamm. Genome. 13:469–474. doi:10.1007/s00335-002-3001-x.

Wise, A.L., L. Gyi, and T. a. Manolio. 2013. EXclusion: Toward integrating the X chromosome in genome-wide association analyses. Am. J. Hum. Genet. 92:643–647. doi:10.1016/j.ajhg.2013.03.017.

Womack, J.E., and Y.D. Moll. 1986. Gene map of the cow: conservation of linkage with mouse and man. J. Hered. 77:2–7.

Wood, S.H., X. Ke, T. Nuttall, N. McEwan, W.E. Ollier, and S.D. Carter. 2009. Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs. Immunogenetics. 61:765–772. doi:10.1007/s00251-009-0402-y.

WTCCC. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 447:661–78. doi:10.1038/nature05911.

Yatsenko, A.N., A.P. Georgiadis, A. Röpke, A.J. Berman, T. Jaffe, M. Olszewska, B. Westernströer, J. Sanfilippo, M. Kurpisz, A. Rajkovic, S.A. Yatsenko, S. Kliesch, S. Schlatt, and F. Tüttelmann. 2015. X-Linked TEX11 Mutations, Meiotic Arrest, and Azoospermia in Infertile Men. N. Engl. J. Med. 372:2097–107. doi:10.1056/NEJMoa1406192.

Yu, J., G. Pressoir, W. Briggs, I. Bi, and M. Yamasaki. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of. Nat. Genet. 38:203–208.

Zhu, X., T. Feng, B.O. Tayo, J. Liang, J.H. Young, N. Franceschini, J.A. Smith, L.R. Yanek, Y.V. Sun, T.L. Edwards, W. Chen, M. Nalls, E. Fox, M. Sale, E. Bottinger, C. Rotimi, Y. Liu, B. McKnight, K. Liu, D.K. Arnett, A. Chakravati, R.S. Cooper, and S. Redline. 2014. Meta-analysis of Correlated Traits via Summary Statistics from GWASs with an Application in Hypertension. Am. J. Hum. Genet. 96:21–36. doi:10.1016/j.ajhg.2014.11.011.

Zimin, A. V, A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.P. Van Tassell, T.S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. a Yorke, and S.L. Salzberg. 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10:R42. doi:10.1186/gb-2009-10-4-r42.

Individual Training Plan

TRAINING (30 ECTS minimum)		
Mandatory courses	**Where/when**	**ECTS**
Welcome to EGS-ABG	AU/Sep 2012	2
Scientific writing and presentation	AU/Nov 2012	2
How to write and publish a scientific paper	SLU/Jan-Mar 2015	3
EGS-ABG Fall Research School	Ethiopia/Oct 2013	2
Research Ethics for Science and Technology	SLU/Nov 2015	2
Science in Practice	SLU/Oct 2015	2
Leading organizations, projects and processes	SLU/Mar-Apr 2015	3
Advanced scientific courses (≥18 ECTS)		
Linear Models in Animal Breeding	Norway/June 2015	3
Selection and Response on Quantitative Trait	Germany/July 2013	3
Sequence data analysis training school	Netherlands/Dec 2012	1.5
Linkage and association mapping	AU/Jan-Mar 2013	5
Introduction to Perl Programming	AU/May 2013	1.5
Breeding Plans for Sustainable Animal Breeding	AU/Aug 2013	5
EGS-ABG fall course "The Sustainability Concept in Animal Breeding"	SLU/Oct 2015	2
Population Genomics in Crops and Farm Animals	Austria/Nov 2012	2
Total credits (≥30 ECTS)		**39**

DISSEMINATION OF KNOWLEDGE	
Teaching/MSc supervision	**Where/When**
Teaching assistant for the course "Genetics"	AU/ Mar 2013 - June 2013
Teaching assistant for the course "Gene mapping"	AU/Dec 2013 – Jan2014
Teaching assistant for the course "Quantitative genetics"	AU/Dec 2013 – Jan2014
International conferences (minimum of 3)	
Poster in Conference of European Association for Animal Production(EAAP-France)	France/Aug 2013
Oral presentation in Conference of 10th World Congress on Genetics Applied to Livestock Production (WCGALP)	Canada/Aug 2014
Oral presentation in Conference of "Breeding for Bacon, Beer and Biofuels"	United Kingdom/Apr 2015
Oral presentation in Conference of European Association for Animal Production	Poland/Sep 2015
Poster in Conference of 5th International Conference on Quantitative Genetics	United States/June 2016
Seminars and workshop (minimum 1)	
Agricultural Research Connections Workshop	Kenya/July 2013
Seminar in Department of Animal Breeding and Genetics	SLU/May 2015
Other Activity	
Organizer for weekly "Gene mapping" team meeting	AU/ Jan 2013 – Sep 2014