Application of High-throughput Genomic Data in the Genetic Analysis of Pigs

by

Tianfu Yang

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

Doctor of Philosophy

in

Animal Science

Department of Agricultural, Food and Nutritional Science
University of Alberta

Abstract

The emergence of high-throughput genomic data provides new opportunities for genetic analysis. Especially, in terms of genetic improvement of animals, genomic tools provide new tools for animal selection, while still suffering from some statistical issues due to the high-dimension of the genomic data.

In my thesis work, I aimed to better interpret available genomic data by integrating information from various sources (e.g. different types of data) and knowledge in different areas (e.g. genomics, statistics and animal science), and attempted to make better use of high-throughput genomic data in the genetic analysis of pigs, especially in the context of genetic improvement. More specifically, the objective of the thesis work is twofold: 1) to improve the detection power and precision in two genome-wide association studies (GWAS), one for meat colour of pork and one for fetal response to PRRSV challenge in pigs; and 2) to explore a new strategy for constructing linkage maps. I worked toward this goal in four studies. For the studies about GWAS (Chapter 2-4), I discussed the issue about detection power and precision due to the high dimension of genomic data. In the study about the adaptive LASSO (Chapter 2), I assessed its performance and discussed how it may help to increase the detection power in GWAS. In the GWAS of fetal response to type 2 PRRSV challenge (Chapter 3), I made use of permutation to improve the precision of the results, and used transcriptomic data to further scan for candidate genes. In the TDT study (Chapter 4), I improved the precision of GWAS by integrating raw genotyping data (fluorescence intensity data) into the analysis. In the study about multiple sperm typing (Chapter 5), I proposed a new model for the allele dosage data of haploids, in order to improve the efficiency of linkage map construction.

In the thesis work. Twenty candidate genomic regions were found to be associated with meat colour of pork (Chapter 2), and 21 candidate regions were found for fetal response to PRRSV challenge in pigs (Chapter 3). These candidate regions may lead to new genetic markers for marker-assisted selection in the future. Some results implied better performance of the methods used in the GWAS, in terms of higher detection power and/or higher precision, which provides valuable experience in the analysis of high-dimensional genomic data. In addition, a new strategy was proposed to construct linkage maps using allele dosage data of sperm cells, which showed good accuracy in simulation studies (Chapter 5). The good performance in the simulation studies implied its application in animal genomics. Especially, its potential ability to genotype chromosome structural variations (CSVs) and estimate recombination rate on an individual level may provide additional materials for genetic improvement.

Preface

Chapter 2 of this thesis is a part of a research collaboration (Project name "Whole genome association analysis for pork quality in both cross- and pure-bred populations"). The project is a partnership between the University of Alberta and two pig breeding companies (Hypor Inc., Regina, SK, Canada, and Genesus Inc., Oakville, MB, Canada). The animals used in this study were raised as part of commercial pork production. The proposed work was reviewed by the University of Alberta Animal Care and Use Committee and considered as Category A (little or no animal manipulation) and no formal ethics approval was required. No other specific permissions were required for the work as the animals were produced as part of commercial pig breeding and pork operations and cared for according to the Canadian Quality Assurance Program, see http://www.cqa-aqc.com/index-e.php, which includes attention to animal health and well-being and is in line with the Canadian Council on Animal Care guidelines.

Section 2.2 in Chapter 2 has been accepted by the Canadian Journal of Animal Science, as Tianfu Yang, Zhiquan Wang, Younes Miar, Heather Bruce, Chunyan Zhang, and Graham Plastow, A Genome-wide Association Study of Meat Colour in Commercial Crossbred Pigs. I was responsible for designing and conducting the GWAS as well as the manuscript composition. G.P. conceived and designed the experiments. H.B., C.Z. and Y.M performed the experiments. C.Z. and Z.W. aided in the data analysis. All authors contributed to the manuscript.

Chapter 3 and Chapter 4 of this thesis are a part of a research collaboration (Project name "Application of genomics to improve swine health and welfare"). The project was led by Graham Plastow at the University of Alberta, with John Harding at the University of Saskatchewan and Bob Kemp at the PigGen Canada Inc. being the lead collaborators. The experiment described in

these chapters was approved by the University of Saskatchewan's Animal Research Ethics Board. It adhered to the Canadian Council on Animal Care guidelines for humane animal use (permit #20110102)

Chapter 3 of this thesis has been published as Tianfu Yang, James Wilkinson, Zhiquan Wang, Andrea Ladinig, John Harding and Graham Plastow, 2016. A genome-wide association study of fetal response to type 2 porcine reproductive and respiratory syndrome virus challenge, Scientific Reports, 6: 20305. I was responsible for designing and conducting the GWAS as well as the manuscript composition. J.H., G.P., A.L. and J.W. designed and performed the PRRS pregnant gilt challenge experiment. A.L. and J.H. analyzed the phenotypic responses in fetuses and dams including assessment of viral load. J.W. conducted the transcriptomic experiment. Z.W. and G.P. oversaw the data analysis. All authors contributed to the manuscript.

Preliminary results of the study described in Chapter 5 has been published as Tianfu Yang, Zhiquan Wang, Zhiqiu Hu and Graham Plastow, 2014. A New Method to Estimate Recombination Rate Based on SNP Allelic Dosage Data. in Proceedings of 10th World Congress of Genetics Applied to Livestock Production. ASAS.

Dedicated to my parents, my wife, and my late paternal grandmother.

Acknowledgements

First and foremost, I would like to thank my supervisors, Dr. Graham Plastow and Dr. Zhiquan Wang, for their tremendous guidance and support with every aspect of my PhD program. I have greatly benefited from their thorough knowledge and considerable experience. Dr. Plastow has always been supportive, kind and considerate, and I am very grateful for the insightful and creative ideas he has offered me. Dr. Wang has provided generous help with various aspects of my PhD program, and has shared his extensive expertise in statistics and genetics. I have been very fortunate to work under their supervision.

I am very grateful for the help of people who contributed to my thesis work. Dr. Paul Stothard is a member of my supervisory committee, and I would like to thank him for his valuable assistance with my projects. In my thesis work, I also learnt a lot from partner research groups. In particular, I would like to thank Dr. Heather Bruce, Dr. John Harding, Dr. Joan Lunney and their group for their help in different projects. I would also like to thank the people from Genesus Inc. and Hypor Inc. who contributed to the projects, especially Dr. Robert Kemp, Dr. Patrick Charagu and Dr. Abe Huisman.

It has also been a pleasure to work with colleagues at Livestock Gentec and Delta Genomics. I appreciate their help and support. In particular, I am very grateful for the assistance provided by Dr. Zhiqiu Hu, Dr. Chunyan Zhang, Yan Meng, Dr. Ghader Manafiazar, Dr. Younes Miar, Jackson Ma, Dr. Changxi Li, Dr. Chinyere Ekine-Dzivenu, Dr. Everestus Akanno, Dr. Arun Kommadath, Dr. Adriano Arantes, Dr. Elda Darvishi and Dr. Dagnachew Hailemariam.

I have also benefited from training opportunities provided by the University of Alberta. Special thanks to the classes provided by Dr. Ellen Macdonald, Dr. Dean Spaner and Dr. David Wishart, and the training sessions given by Dr. Robert Desjardins and Dr. Mebbie Bell.

I would like to sincerely thank my family. My parents, Xiao Yang and Fengqin Lin, and my wife, Fangshu Lin, have been encouraging and supportive throughout my PhD program. My late paternal grandmother, Yiyang Chen, was my first teacher who inspired me to learn and think. I would also like to thank my friends, especially Ethan Li, Cheng Li, Binnan Zhou, Tianbo Liu, Shuning Li, Tianxin Bao and Xunchen Liu, for their support. Special thanks to Sheng Tian, Keren Jiang and Yinan Wang for all the happy days we spent together in QiQiSiBa.

This thesis would not have been possible without funding from the following funding agencies and companies: NSERC, Hypor Inc., Genesus Inc., Genome Canada, Genome Prairie, Genome Alberta and Swine Innovation Porc.

Thanks to everyone involved in this thesis project on every level.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Scientists have been sending out warnings of a deluge of genomic data for more than 20 years (Aldhous 1993), while the volume of genomic data keeps increasing even faster than expected. For example, the whole genome shotgun (WGS) sequences processed at GenBank (in bases) increased more than 2,000 times in 14 years (http://www.ncbi.nlm.nih.gov/genbank/statistics/). The increase in genomic data was driven primarily by advances in high-throughput technologies, which result from developments in molecular biology and other related research fields, such as nanotechnology and chemistry.

Another significant change in genomic data is the fast increase of dimension, e.g. the density of genotype data. The advances in molecular biology and genetics have revealed a bigger and more complete picture of the genome, which implies a more complex structure of genomic data. One example of the challenge in high-dimensional data analysis is the issue of power analysis and significance testing in genome-wide association studies (GWAS) (Dupuis and O'Donnell 2007).

Fortunately, even though handling the data deluge is challenging, genomic research is still afloat and keeps moving forward. The rapid increase in genomic data significantly speeds up the development in applications of genomics, and impacts the experimental design and interpretation of results from genomic analysis. As researchers are facing more complicated scenarios, more tools in mathematics and computer sciences have been introduced for practical genomic data analysis (Xu 2013). The integration of methodologies from different disciplines provides opportunities to look at traditional questions in a new way, such as the application of directed acyclic graphs (DAGs) in pedigree comparison (Kirkpatrick et al. 2012).

In this Ph.D. program, I aimed to better interpret accessible genomic data by integrating different types of data and knowledge in different areas (e.g. genomics, statistics and animal science), and explore the possibility of better application of high-throughput genomic data in the genetic analysis of pigs. There are four studies in my thesis work, mainly concerning quantitative trait loci (QTL) mapping for pork quality and resistance to porcine reproductive and respiratory syndrome virus (PRRSV), and high-throughput genome mapping methodology. This chapter aims to provide a general introduction and literature review for the studies, including three sections: 1) genomic data, 2) application in livestock breeding, and 3) outline of the following chapters.

1.1 Genomic data

In this section, I first give a brief summary of the development of sequencing technology. Even though genomic data can also be produced by bio-technologies other than DNA sequencing (such as bead-based SNP genotyping microarrays), these technologies still depend on information provided by DNA sequencing to a degree. Then I focus on data generated from single nucleotide polymorphisms (SNPs), as this kind of data provided important information about genetic variation across individuals in my thesis work. Finally, I summarize existing DNA quantification methods, which have been increasingly used for genomic research recently and which play an important role in Chapter 5.

1.1.1 Genome sequencing

Genome sequencing does not have a very long history. Since the discovery of DNA structure in 1953 (Watson and Crick 1953) and the beginning of the interpretation of the genetic code in 1961 (Crick et al. 1961), people started to notice that the nucleotide sequence within the DNA molecule

carries specific genetic information. The first success in sequencing a complete gene was achieved in 1972, when Min Jou et al. established the entire nucleotide sequence of the coat gene of bacteriophage MS2 by characterizing RNA fragments (Min Jou et al. 1972). One of the most important sequencing methods, Sanger sequencing, was invented in 1975 (Sanger and Coulson 1975). The method was also called "plus and minus", which is based on a chain-terminating technique and radio-labelled nucleotides. With this powerful tool, Sanger et al. sequenced the full genome of bacteriophage PhiX174 (Sanger et al. 1982) in 1977. This method was then improved and ultimately automated in 1987 (Gocayne et al. 1987). Sanger sequencing is usually grouped as one of the "First generation sequencing methods", which also includes other sequencing technologies, such as Maxam–Gilbert sequencing (Maxam and Gilbert 1977) and fluorescent sequencing (Smith et al. 1986). However, these other methods were not used as widely as Sanger sequencing.

Even though Sanger sequencing had been the most favored sequencing technique for 20 years, its throughput did not increase as fast as needed. The advent of next-generation sequencing (NGS) methods then marked a new era. These methods include Pyrosequencing (Ronaghi 2001), Illumina sequencing, SOLiD sequencing, and semiconductor sequencing (e.g. Ion next generation sequencing). All of them outperform Sanger sequencing in one or more of four aspects: 1) sequencing cost, 2) sequencing speed, 3) degree of automation, and 4) sequencing accuracy (El-Metwally et al. 2014). Evaluations reported that each of these methods has certain advantages over others as well as its own limitations (Nakamura et al. 2011; Luo et al. 2012). Among these methods, Illumina sequencing is thought to be an "industry standard" for NGS and widely accepted, which may partially be explained by its low cost, high throughput, and the huge amount of available software for the data analysis.

However, the read length of these methods is usually no longer than 600 base pairs (bp) and often less (e.g. 2×150 bp for Illumina Hiseq 4000 System), which is low in some scenarios such as *de novo* sequencing. In *de novo* sequencing, there are no available genome maps, and genome assembly greatly depends on the information provided by only those reads (Chaisson et al. 2015). Overlaps between longer reads tend to be larger, which helps in ordering the reads and in the reconstruction of DNA sequence. On the other hand, longer reads from resequencing (i.e. sequencing with reference map available) can be mapped to reference maps more easily as they carry more information. As a result, some new sequencing methods mainly focus on obtaining extra-long reads (e.g. longer than 10,000 bp), such as PacBio sequencing (Zhu and Craighead 2012), Nanopore sequencing (Branton et al. 2008) and GemCode technology (Zheng et al. 2016). These sequencing technologies, however, still suffer from their high cost, and their error rate is also higher in part due to the greater read length.

The development of sequencing technologies keeps increasing the throughput, and plays an important role in many significant whole-genome sequencing projects. The first one of these projects was the first sequencing of the human genome, also known as the human genome project (HGP) (Lander et al. 2001; Venter et al. 2001). One outcome of the HGP was the discovery of a huge number of SNPs in the human genome (Sachidanandam et al. 2001), which directly enabled GWAS in human, especially those studies about human disease (Collins 1999; Collins and McKusick 2001). Whole-genome sequencing is also a powerful tool for livestock studies. For example, the bovine genome sequencing project (Elsik et al. 2009) contributed significantly in mapping bovine SNPs, which also enabled other genomic studies such as those about the genetic structure of cattle breeds (Bovine HapMap Consortium et al. 2009). The latest assembly of the swine genome (Sscrofa10.2) was constructed in 2011 through a hybrid approach combining BAC

derived sequencing and whole genome shotgun sequencing (Archibald et al. 2010; Groenen et al. 2012), and its output enables a part of the analysis in my thesis work. A new version of the swine genome (Sscrofa11) has been reported subsequently, which made use of the Pacbio sequencing to improve the sequence assembly (Warr et al. 2017).

1.1.2 Genetic markers and SNP

The discovery of a huge number of SNPs marked a new era in genomic studies, as it provides abundant genetic markers that are relatively easy to genotype. Before the wide use of SNPs, many other types of genetic markers were applied in genetic analysis, such as RFLP (Kan and Dozy 1978), VNTR (minisatellites) (Jeffreys et al. 1985; Nakamura et al. 1987), SSR (microsatellites) (Litt and Luty 1989; Weber and May 1989;), RAPD (Williams et al. 1990) and AFLP (Vos et al. 1995). The discovery and genotyping of these genetic markers did not require extensive DNA sequencing information, and therefore is easy to implement without high-throughput sequencing technologies. Even though the cost of whole genome sequencing keeps decreasing currently, these genetic markers are still being used in some species. Another kind of genetic marker, or variation, that was discovered extensively with whole genome sequencing is chromosome structural variations (CSVs) (Pinkel et al. 1998; Korbel et al. 2009; Alkan et al. 2011;). Generally, CSVs include copy number variations (CNVs), insertion/deletion (indel), translocations and inversions. It has been reported that CSVs underlie many heritable traits of importance. Efficient genotyping of CSVs is still challenging currently, and its analysis is one of the targets of my thesis work (Chapter 5). Some genetic variants may be grouped into multiple kinds of genetic markers mentioned above. For example, SSR is a specific type of CNV.

Among these genetic markers, SNPs show advantage over these other types of genetic marker in terms of density (most abundant in the genome), stability (stable inheritance over generations), and throughput of genotyping. The discovery of a huge number of SNPs provided a very important tool for genomic analysis. SNPs themselves are potential causative factors. For example, the HAL-1843 mutation is responsible for porcine stress syndrome (Fujii et al. 1991). More importantly, SNPs may serve as genetic markers of other common genetic variations among individuals (International HapMap Consortium et al. 2007), partially because of their high density. This property of SNPs directly enables GWAS as well as other advanced applications like genomic selection (Meuwissen et al. 2001). All the 4 studies in my thesis work (Chapter 2-5) involved SNP data in the analysis.

Currently, high-throughput SNP genotyping is mainly conducted with array-based methods and increasingly genotyping-by-sequencing (GBS) techniques. Among the array-based methods, two technologies are most popular. One of them was developed by Affymetrix (Santa Clara, USA) in 2003 (Kennedy et al. 2003). The latest version of the technology makes use of a semiconductor based photo-lithographic technology in order to increase the call rate. The other technology was first proposed in 2005 (Gunderson et al. 2005) and then improved in 2006 (Steemers et al. 2006) by Illumina (San Diego, USA). It is based on a two-color BeadArray and generates SNP allele dosage data to determine the genotypes of samples (Steemers et al. 2006; Hackett et al. 2013). As allele dosage data is one of the key concepts in Chapter 5, we also introduce it with more detail in the next sub-section (DNA quantification).

For GBS techniques, even though it was a whole-genome sequencing technology that helps discover a huge number of SNPs, the idea of whole-genome SNP genotyping by sequencing was not widely applied at first (Elshire et al. 2011). However, with the decrease of the cost for NGS

and the development of related statistical tools, GBS started to show advantages over array-based methods in some fields (Nielsen et al. 2011). By applying NGS to pooled DNA samples, GBS is able to achieve a high level of multiplexing, and significantly reduce the cost of genotyping per sample (De Donato et al. 2013). In addition, the GBS technique combines the discovery of new SNPs and the genotyping step, which is more cost-effective and efficient (De Donato et al. 2013). The advantage of no requirement for pre-discovered SNPs also makes GBS a better choice when reference sequence is not available (Elshire et al. 2011). On the other hand, currently, the analysis of NGS data may require additional investment such as bioinformatics training and computational infrastructure, which should also be taken into consideration for cost estimation.

1.1.3 Nucleic acid quantification

Another important type of genomic data used in my thesis work is nucleic acid quantification data. Compared with qualitative genotyping data, which are mainly about typing different sequences, quantification data carry information of the quantity/dosage of sequences. Common applications of nucleic acid quantification data include gene expression measurement (Schena et al. 1995), copy number variation detection (Pinkel et al. 1998), and pathogen detection (Strain et al. 2013). Currently, technologies for nucleic acid quantification include bead-based quantitative genotyping technology, sequencing methods, and PCR-based methods.

The aforementioned Illumina genotyping platform (e.g. Illumina PorcineSNP60v2+ Beadchip) is an example of quantitative genotyping. This method quantifies alleles for high-density whole-genome SNPs with a high throughput, which results in a low cost both financially and in experiment time. However, as the technology was originally designed for genotyping rather than DNA quantification, more bias may be introduced in the measurement compared with other

quantification techniques. For example, it has been reported that the result of DNA quantification may be biased because the use of different dyes in the Illumina Infinium II assay (Staaf et al. 2008).

Sequencing technology started to be widely used for gene expression measurement in 2008 as a quantification method (Marioni et al. 2008). The basic idea is that the number of sequencing reads containing a locus in a given sample can be a measurement of the abundance of the locus (Tarazona et al. 2011). Even though there are still some concerns about the accuracy of using RNA-seq in gene expression measurement, the results of many studies has provided support on its performance (Bullard et al. 2010; Griffith et al. 2010). For DNA quantification, there are few previous studies about applying sequencing in this way. One of the known biases during this process is introduced by amplification (a PCR step) incorporated in the sequencing protocol, as the PCR amplification rates for different loci may vary (Rhodes et al. 2014). A possible option may be the TruSeq PCR-free technology developed by Illumina, which does not incorporate PCR in the sequencing (Rhodes et al. 2014).

For nucleic acid quantification, the most commonly used methods are PCR-based technologies, such as real-time PCR (also known as quantitative PCR, qPCR) and digital PCR (dPCR). While these methods may have a relatively low throughput compared with the two methods mentioned above, they are able to achieve a higher accuracy in quantification. Indeed, it is still the case that reviewers request validation of RNA-seq results by this approach (qPCR). Among these methods, droplet digital PCR (ddPCR) provides the best accuracy currently. This technology was first proposed in 2011 (Pekin et al. 2011). By making use of the advances in nanofluidics, it performs PCR in a large number (e.g. millions) of droplets simultaneously. Previous evaluation showed that ddPCR provided absolute quantification of target sequences with high accuracy, dynamic range, and reproducibility (Hindson et al. 2013). Commonly used

platforms for ddPCR include the RainDrop system developed by Raindance (Billerica, USA), and the QX100/200 system developed by Bio-Rad (Hercules, USA). Some other dPCR platforms are also available in this field. Biomark HD provided by Fluidigm, for example, has been reported to have comparable quality in DNA quantification to the QX100 system in certain scenarios (Pavšič et al. 2016).

Other than these technologies, some other solutions are also available and may find a new (not necessarily better though) balance between throughput and accuracy, such as the nCounter system developed by nanoString (Seattle, USA). This method was reported to perform well in sensitivity, reproducibility and robustness for quantification, while keeping a relatively high throughput (Veldman-Jones et al. 2015). However, this technology is not able to distinguish sequences with minor differences, such as two different alleles for a SNP, which limits its application to a degree.

1.2 Application of genomic data in livestock breeding

The availability of high-throughput genomic data brings new opportunities for genomic studies especially in livestock. It significantly enhances our ability for detecting genetic mutations behind or underlying traits and improves our understanding of the biology of traits. In the context of livestock breeding, the genomic data also plays a special role. In this section, I give a brief introduction of two applications of genomic data in livestock, which are also the two main topics in this thesis. They are 1) QTL mapping and animal genetic improvement, and 2) genome mapping.

1.2.1 QTL mapping and animal genetic improvement

An important application of animal genomics in agriculture is to improve the genetic merit of animals. Depending on the underlying genomic structure of different traits, the strategy for genetic improvement can be different. My thesis work described in Chapter 2-4 is mostly involved in QTL mapping and GWAS, which may be applied in the genetic improvement of pigs. This section aims to give a review about some related topics discussed in the following chapters, while also including a brief introduction about other applications of genomic data in genetic improvement, such as genomic analysis of Mendelian traits and genomic selection.

1.2.1.1 Genetic improvement for Mendelian traits and complex traits

Genetic improvement in livestock refers to organised genetic changes in a desired direction in a population, resulting from artificial selection and breeding. Strategies for genetic improvement are slightly different between Mendelian traits and complex traits.

For Mendelian traits, their inheritance follows Mendel's laws and are controlled by a single genomic locus. Example includes a deleterious allele of gene *RYR1* (also known as the "Halothane gene") causing porcine stress syndrome (Fujii et al. 1991), which was the first genetic marker used in pig breeding. The genetic improvement for these traits can be done by reducing the frequency of the deleterious alleles. Mendelian traits usually showed clear pattern of inheritance, and it is relatively easy to find carriers of deleterious alleles even without genomic tests (e.g. through test mating). However, genomic tests have exclusive advantages in some cases, such as early stage tests for inherited disorders. Traditional genomic methods for mapping genes underlying Mendelian traits includes linkage analysis, which detects co-segregation of genetic markers and phenotypes in designed mapping population, such as the analysis of dwarfism in pigs

(Nielsen et al. 2000). The application of high-density SNP genotype data further enables the GWAS of Mendelian traits. Compared with linkage analysis, GWAS make use of the linkage disequilibrium (LD) between genetic markers and causative mutations, and does not require well-designed and specifically constructed mapping population to infer recombination fraction, which significantly reduced the cost of data collection (Ott et al. 2015). Successful examples of GWAS of Mendelian traits in animals includes the mapping of genes underlying congenital muscular dystonia in Belgian Blue cattle (Charlier et al. 2008).

Meanwhile, many traits of interest in livestock production are complex traits, such as milk yield or meat quality. They are commonly affected by multiple genetic variants with low-to moderate effects (Falconer and Mackay 1996). The joint average effect of these genetic variants that can be passed on to the next generation (also called breeding value) can serve as the measurement of genetic merit for animal selection. Genetic improvement can be achieved by selecting animals with high breeding value. Similar to Mendelian traits, this process can be done without genomic tests. Random effect models and BLUP (best linear unbiased prediction) have successfully been applied to estimate breeding value using phenotypic and pedigree information in the past decades (Robinson 1991). However, the genomic tools provided new opportunities in genetic improvement, including marker-assisted selection (MAS) (Smith 1967; Meuwissen and Goddard 1996) and genomic selection (GS) (Meuwissen et al. 2001). The genomic analysis leads to more possibilities in animal genetic improvement, such as reducing generation interval and improving selection accuracy. As a result, the genomic tools have advantages especially for some traits that are difficult to improve in traditional selection program, e.g. invasive recording or those requiring progeny test.

1.2.1.2 QTL mapping and marker-assisted selection (MAS)

Genomic loci that are associated with complex traits are also called quantitative trait loci (QTL). The development of QTL mapping has been largely driven by increasing accessibility of genomic data. Early efforts to map QTL suffered from the low density of available genetic markers, and therefore applied interval mapping to improve the detection power in QTL mapping (Lander and Botstein 1989). This method exploits available genetic linkage maps to infer the genotype of unobservable loci between two genotyped markers, and thereby overcomes the issue of low density to a degree. However, the resolution of mapping is usually low (e.g. 20 centimorgans), as the genotype information is very limited. Recent developments in molecular genetics have made it possible to efficiently genotype large amounts of genetic markers in the genome, and therefore enables genome-wide association studies (GWAS). By directly associating phenotypes with available high-density genotypes, GWAS has successfully identified QTL for many important complex traits, such as body composition and structural soundness traits in pigs (Fan et al. 2011) and milk fat yield in cattle (Jiang et al. 2010). More discussion about methods of GWAS that relate to my thesis work can be found in section 1.2.1.4.

Through GWAS, it becomes possible to identify genetic markers associated with traits of interest. These markers may be the functional genetic mutations underlying variation in those traits, or, in most cases they are more likely in LD with the functional mutations. The key idea of MAS is to use associated genetic markers (usually only those with statistically significant effects) to predict the genetic merit of animals, and then the predicted genetic merit can be used to make selection decisions. In MAS, not only validated causative mutations can be used, but all candidate SNPs showing great association with the trait can be integrated into the prediction. Even though part of the association may highly depend on the linkage disequilibrium (LD) between the SNPs

and real causative mutations, those SNPs still improve the predictability of genetic merit (Hayes and Goddard 2003).

This strategy has been applied in the selection of various livestock species (Goddard and Hayes 2009). For example, genetic markers have been used to improve traits of interest in the pork industry, such as pork quality, reproductive traits and animal health (van der Steen et al. 2005). For dairy cattle, MAS has been helping in the genetic improvement of reproductive traits, milk yield and milk quality (Dekkers 2004).

1.2.1.3 Genomic selection (GS)

As the density of available polymorphisms has increased, a different strategy, termed genomic selection (GS), was first introduced in 2001 (Meuwissen et al. 2001) and enabled in 2007 (Snelling et al. 2007). The idea behind this strategy is to use all the markers across the genome to predict breeding value, rather than selecting only significant markers. It is extremely important in those cases where the traits of interest are determined by many genetic variants with small effect (Meuwissen et al. 2016). In MAS, genetic variants with small effect are usually not statistically significant due to lack of statistical power (also see section 1.2.1.4.2). By using GS, all effects can be taken into consideration to achieve a better prediction. This strategy has dramatically increased the rate of genetic improvement in dairy cattle, which may result from the application of large reference populations (Hayes et al. 2009; Wiggans et al. 2011). Genomic selection is also being implemented in beef cattle and pigs (Lillehammer et al. 2011; Lourenco et al. 2015), however, its advantage over the traditional selection is not as great as that achieved in dairy cattle (Meuwissen et al. 2016).

Genomic selection does not require mapping specific QTL, or in other words, it captures all QTL along the genome without any selection (or previous identification) to predict the breeding value (Goddard and Hayes 2009). However, fine mapping of QTL may serve as prior information and help to improve the accuracy of GS, especially when huge amount of biological and genomic data are available (Hickey 2013). One current example to achieve this goal is BayesRC, a method to incorporate biological information into GS (MacLeod et al. 2016).

1.2.1.4 GWAS

This section aims to give a brief introduction of three more specific topics related to my thesis work. They are 1) contingency-table test for binary traits 2) multiple testing, and 3) linear regression for continuous traits.

1.2.1.4.1 Contingency table test

"Contingency-table" test is a family of tests in classical contingency table analysis, such as Pearson's Chi-square test and Fisher's exact test. These tests are usually used in the analysis of binary traits, i.e. traits with only two possible phenotypic values. The basic idea of these tests is to test the dependency between the phenotype and allelic frequency or genotypic frequency. If an allele or a genotype for a polymorphism has quite different frequencies for the two groups of animals, it is likely that the polymorphism is associated with the binary trait. Significant tests can then be used.

The transmission disequilibrium test (TDT) can be treated as a special type of contingency-table test. In TDT analysis, the genotype data are only available for animals with one certain phenotype, which is not uncommon in disease related studies. In this case, it is impossible to

directly test the dependency between phenotype and allelic / genotypic frequency. Instead, TDT analysis requires the parents' genotypes, and detects possible transmission disequilibrium (distortion from the Mendelian segregation) in the animals with genotype data, which can be used as a sign of association between the genetic marker and the trait (Spielman et al. 1993). More discussion about TDT can be found in Chapter 4.

The contingency-table test also suffers from some drawbacks. In most situations, the contingency-table test is conducted separately for each single SNP, i.e. single marker association. For complex traits that are affected by multiple genetic factors, single marker association is not able to provide optimal solutions, as the effects of all other SNPs are ignored during the calculation for a certain SNP. In other words, the simple (single marker) model is not able to describe the complex relationship between the phenotype and the underlying genetic factors. On the other hand, it is difficult to adjust for systematic effects, such as age, sex, population stratification, in most contingency-table tests. A common approach to deal with this problem is to obtain samples using a balanced design, where the individuals with the two phenotypes have identical (or at least similar) distribution of age, sex, genetic background and/or other known information. However, it is usually prohibitively difficult to match sampled individuals, and requires very good experimental design. Some methods have also been proposed to solve this problem (Reich and Goldstein 2001), but it is still very challenging when multiple factors need to be integrated in the analysis.

1.2.1.4.2 Multiple testing

In addition, when high-density SNP data are used, single-marker association is faced with the dilemma of choosing a significance level (α). The significance level determines how stringent the criteria we use to decide the significance of SNPs. With a higher significance value (i.e. higher

threshold of p-value), it is likely that more SNPs are detected to be significant in GWAS, at a cost of more false positives. On the other hand, a lower significance level means more false negatives (i.e. lower power). Generally, a lower significance level should be used when multiple tests are conducted simultaneously. Otherwise, the decision error (probability of false positives) tends to accumulate as the number of tests increase, and result in an unacceptably serious false positive error. The high density of SNPs introduced new statistical issues for GWAS. One issue for multiple marker association will be discussed in the next section. For single-marker association, denser SNPs suggests a larger number of tests, and the multiple testing problem is therefore more severe.

Existing methods of adjustment for multiple testing include a family of adjusting procedures, such as Bonferroni procedure and Dunn-Sidak procedure (Quinn and Keough 2002). The Bonferroni procedure is a widely used technique in GWAS, which provides a great control of false discovery. This method, however, has been criticized to be too conservative and reduce the power in the association analysis. The reason is not only its own statistical property, but also the fact that the number of independent tests in GWAS is usually overestimated due to the existence of LD between SNPs (Gao et al. 2009). It has been reported that false discovery rate (FDR), another statistic used in multiple testing, can successfully handle the LD between SNPs and is more powerful than the Bonferroni procedure (Benjamini and Hochberg 1995; Yekutieli and Benjamini 2001). This method was used in my thesis work (Chapter 4) for multiple testing in TDT. Another option to obtain higher detection power is resampling-based adjustment (e.g. permutation test) (Westfall and Young 1993; Westfall et al. 1993). While this method gives a good solution to the multiple testing problem, it can be extremely time-consuming in practice.

Another way to deal with the multiple testing issue is to reduce the number of tests. Some efforts have been made to correct the overestimation of the test number. The estimated number of

independent tests (also called as “effective number of independent tests”) highly depends on the distribution of LD along the genome, but is usually much lower than the number of SNPs used in the estimation (Duggal et al. 2008; Gao et al. 2009; Li et al. 2012). It is also reported that the negative impact of multiple testing on high-density SNP data analysis can be reduced by haplotype-based methods to some degree, as the haplotype phasing also takes account of the LD between SNPs. This method also depends on the distribution of LD along the genome (de Bakker et al. 2005).

Even though methods have been proposed to handle the multiple testing issue in GWAS, their effectiveness is still controversial. The dilemma of choosing significance level may be partially rooted in the huge number of potential genetic variations and the limited information available in the association analysis, and there may be no simple solution to this issue in the trade-off between increasing statistical power and reducing false positives. On the other hand, with the help of advanced statistical tools, biological information from other studies (e.g. transcriptomics, proteomics and metabolomics studies) may be another key in providing a satisfactory answer to our question. The aforementioned haplotype-based methods can serve as an example of successfully achieving better performance through integrating genomic information and knowledge (LD information) into the analysis. Another example is Multi-trait GWAS, which applies prior knowledge about the functional connection between traits in the analysis to improve power (Heid and Winkler 2016). Some advanced statistical tools may also help to integrate prior functional information into GWAS. For example, the aforementioned BayesRC has been reported to increase the accuracy of QTL mapping using prior biological information (MacLeod et al. 2016).

1.2.1.4.3 Linear regression

Compared with the contingency-table test, linear regression provides a more general solution to association analysis, as it is one of the best-studied statistical tools used in a wide range of fields. Especially, when the phenotypic value is continuous, modelling the phenotype in a discrete way and using contingency-table test may lose important information.

Linear regression aims to detect a possible linear relationship between a dependent variable and one or more independent variables. In the context of GWAS, it is usually, though not always (Li and Yin 2008), used to detect the linear relationship between phenotypic value (dependent variable) and SNP effects (independent variables). In a simple scenario, the analysis can reveal allelic substitution effects of SNPs, i.e. the average effect on the phenotype when one allele was replaced by the other allele at the same locus (Falconer and Mackay 1996). In addition, by appropriately constructing independent variables, linear regression techniques can also be used to model different types of genetic effects, such as dominance or epistatic effects (Gao et al. 2014).

By extending the common linear model into a generalized linear model (GLM), linear regression techniques can also apply to other traits, such as categorical traits (Xing et al. 2014). The most commonly used technique in the GLM family is logistic regression, where the logit (i.e. natural logarithm of the odds for a particular outcome) of a binary trait is treated as the dependent variable (Liu et al. 2009). This technique has been successfully used in a wide range of GWAS analyses (Sullivan et al. 2008; Wu et al. 2009; Ayers and Cordell 2010), and provides evidence for its flexibility in the application of linear regression.

Some drawbacks of single-marker association can also be solved by multiple regression, where more than one genetic effect can be estimated and tested simultaneously (Basu et al. 2011). This property is very desirable in GWAS of complex traits, as genetic factors with small effect on

complex traits they can be very hard to detect in single-marker association. Multiple regression can take into account the joint effect of the genetic factors, and increase the power of detection. In this case, multiple regression also provides better estimation of the size of genetic effects.

Even though multiple linear regression is a very powerful tool for GWAS, high-dimensional genomic data analysis is still a challenge. This problem is also known as "regression for high-dimensional data", which refers to the situation where the number of observed records for a dependent variable is lower than the number of independent variables in the regression (Bühlmann and van de Geer 2011). From a statistical point of view, this kind of data does not provide enough degrees of freedom for multiple regression techniques like ordinary least squares (OLS). Solutions for high-dimensional regression models is one of the hot topics in statistics at the moment. A set of techniques has been developed, such as variable selection techniques, shrinkage methods, and dimension reduction techniques (Hastie et al. 2009). However, these methods also bring some other concerns at some points, such as difficulty in providing a significance test. In my thesis work, we did some exploration in the application of LASSO (least absolute shrinkage and selection operator) and its extensions, which is an important member of the shrinkage methods. More introduction can be found in Chapter 2 and Chapter 3.

1.2.2 Genome maps

High-throughput genomic data can also speed up the construction of genome maps, which provides information about the relative positions of genetic markers, and has been widely used in genomic studies. In the rest of this section I introduce existing methods for the construction of linkage maps and physical maps, and their application in livestock genomics.

1.2.2.1 Linkage maps

The concept of genetic linkage was first introduced by Bateson, Saunders and Punnett in 1905 (Bateson et al. 1905). It was then further developed in Morgan's work in *Drosophila* (Morgan 1911). These works laid the foundations for the development of the first linkage map (Sturtevant 1913). A linkage map contains the relative positions of markers, as well as information about inter-marker distance. The construction of linkage maps greatly depends on the estimation of inter-marker genetic distance, and it is essentially based on detecting meiotic recombination.

The most widely used method for linkage mapping is based on pedigree information and the genotypes of individuals in a population. The pedigree-based method usually uses a multi-generation design, and applies to construct medium-density linkage maps (Rohrer et al. 1994; Barendse et al. 1994; Archibald et al. 1995; Groenen et al. 2000). However, these methods require high cost, long time scales (depending on generation interval), or have high technical difficulty associated with them (Hoh and Ott 2005). In addition, the maps are constructed based on population averages, so they are not able to reveal differences in recombination rate between individuals.

A better way for fine mapping using linkage is the single-sperm typing method (Li et al. 1988), which can also greatly reduce the cost and time associated with building and maintaining the mapping population. This method uses the genotypes of a large number of individual sperm cells to achieve a high resolution (Li et al. 1988; Cui et al. 1989; Lazzeroni et al. 1994). In 1988, Li et al. proposed a single-sperm typing method, which uses the genotypes of 80 single sperm cells to infer the recombination rate between 2 markers (Li et al. 1988). This method was then extended to whole-genome multipoint linkage mapping with the invention of the whole-genome single-cell amplification technique in 1992 (Zhang et al. 1992). The throughput of the single-sperm typing

strategy has kept increasing in the following two decades. Currently it is possible to genotype a single sperm using whole-genome single-cell sequencing. In 2012, Lu et al. create a linkage map with about 1.4 million heterozygous single nucleotide polymorphisms (SNPs) through sequencing 99 single sperm cells (Lu et al. 2012).

While the extremely high density of available markers reflects a significant increase in the throughput of linkage mapping, the resolution can be achieved in single sperm typing remains limited. It is essentially because the number of meioses that can be observed in practice is relatively small. As the occurrence of genetic recombination for a specific genomic locus is random (i.e. not guaranteed to appear in small samples), high-quality linkage maps require a fair number of sperm cells being genotyped, which can be labor-intensive and inefficient. In the aforementioned work, Lu et al. estimated that at least 1,000 sperm are required to achieve the power for detecting personal difference in recombination rate (Lu et al. 2012). In Chapter 5, I introduce our work on a multiple sperm typing method, which aimed to achieve an even higher throughput.

There are also some other methods using high-throughput genomic data in genetic mapping. In 2001 a new estimator of recombination rate was proposed, which is based on genotypic data from a population (Fearnhead and Donnelly 2001). The advantage of this method is that the mapping population does not need to be well designed, and pedigree information is not necessary. The method has been thought to be another option for fine linkage mapping. However, the estimation in this method greatly depends on some estimated parameters of the mapping population (e.g. effective population size), which may be difficult to estimate with a high accuracy.

1.2.2.2 Physical maps

Linkage maps provide important information about the recombination rate between genetic markers, however, the recombination rate is not evenly distributed along the genome. In some cases, it is important to assess the physical distance between genetic markers, and a physical genome map can be very useful.

One of the successful approaches for physical mapping is Radiation Hybrid (RH) mapping. This method makes use of irradiation technology and cell fusion to prepare samples with DNA fragments of chromosome, and measure the inter-marker physical distance with the association score (retention rate). The association score can then be used in physical mapping (Cox et al. 1990). This strategy then led to the development of HAPPY (HAPloid DNA samples using the PolYmerase chain reaction) mapping (Dear and Cook 1993). This *in vitro* method breaks genomic DNA through mechanical shear, irradiation or chemicals, and then using a limiting dilution strategy (i.e. most samples contain either one or zero target DNA molecules (Dong et al. 2014)) to assess the association score between markers (Dear and Cook 1993). This method has been applied in mapping the human genome (Dear et al. 1998) and detecting genome rearrangement (Pole et al. 2011). However, even though HAPPY mapping is less labor-intensive than RH mapping, it still suffers from its low throughput in practice.

Another physical mapping method, optical mapping, provided an example of how cutting-edge biotechnologies can make a traditional mapping method new again. Optical mapping was first proposed in 1993 (Schwartz et al. 1993), and has been used successfully to construct physical maps in both bacteria and plants since then (Lim et al. 2001; Zhou et al. 2009). However, its low throughput hinders its application in constructing high quality whole-genome maps. In 2012, with the help of technologies in molecular biology, automation, image and data analysis, a high-

throughput optical mapping method was proposed, which was used in the assembly of the goat genome (Dong et al. 2012).

Sequencing can also be treated as a physical mapping method, which provides relatively accurate positions in base pair of genetic markers. Current mainstream sequencing technologies, however, are still not able to construct whole-genome physical maps in a standalone way. Even though the reads produced by those technologies have been claimed to be "long enough" (e.g. 100k -300k base pairs), it is not ready for high-quality genome assembly. New developments in sequencing technology lead to the advent of the aforementioned long-read sequencing technologies (PacBio sequencing, Nanopore sequencing and GemCode technology). Their ability to generate long reads in sequencing may lead to high-throughput physical maps (Branton et al. 2008), as summarized in the previous section.

1.2.2.3 Application of genome maps in livestock genomics

In livestock genomics, an important application of genome maps is help in the process of genome assembly to produce a more accurate reference map. For example, a map of the bovine genome was constructed in 2007, which combined the information from pedigree-based linkage maps, bacterial artificial chromosome (BAC) maps, and radiation hybrid (RH) maps. The map was then used in the refinement of Btau3.1 draft assembly (Snelling et al. 2007). In the case of pigs, the use of BACs in the hybrid sequencing strategy aimed to help address this issue. Efficient construction of genome maps may play an even more important role in *de novo* sequencing, where a reference map is not available. A recent example is a linkage map constructed for Asian seabass, which was used to assist its *de novo* assembly (Wang et al. 2017). For genomic analysis (e.g. QTL mapping and genotyping imputation), these accurate reference maps are now an essential tool. On the other

hand, reference maps are usually constructed based on a representative individual or a small representative population, while the true map can be different across populations or even across individuals (Marques et al. 2007). If high-throughput genomic data can enable a more effective and cost-efficient genome mapping method, it may significantly help with the construction of more accurate genome maps for specific populations as needed.

For individualized genome maps, an additional important application is genome-wide genotyping of CSVs, including inversions, translocations, indels and CNVs. More and more evidence has been found in the past decade that CSVs may play a causative role in heritable traits, including some important animal production traits. For example, an indel close to the prolactin gene has been reported to be associated with milking traits in dairy cattle (Cowan et al. 1990), a translocation that encompassed the *KIT* gene has been reported to determine color sidedness in beef cattle (Durkin et al. 2012), and association has also been found between CNVs and fatness in pigs (Schiavo et al. 2014). An effective and low-cost solution for genome-wide detection and genotyping of CSVs will provide easy access to these previously "hidden" genetic variations, and thereby significantly enhance the study of such important traits. For animal breeding, better tools for CSV studies can help in the exploration of "missing heritability", and may lead to improved performance of breeding in livestock (Bickhart and Liu 2014).

Genome maps with high quality may also play other important roles in genomic analysis. For example, a linkage map can help in recombination hotspot analysis in livestock, which may lead to better results in haplotype phasing and genotype imputation (Weng et al. 2014). In terms of animal breeding, linkage maps at an individual level may help select animals with relatively high recombination rate, which may accelerate the process of breeding. Higher recombination rates lead to higher genetic variation with a population over generations, which is essential for artificial

selection. A previous simulation study has reported higher efficiency of breeding programs in a population with higher recombination rates (Battagin et al. 2016). On the other hand, the loss of genetic variation can be a side effect of animal selection, or due to natural processes such as genetic drift (Hill 2000). Identifying and keeping animals with high recombination rates may serve as a way to enhance or maintain genetic variation for sustainable application. The identification of these animals can be done through individualized linkage maps, which provide genome-wide recombination information.

1.3 Outline

This thesis work consists of four independent but connected studies. Chapter 2 explores the application of the adaptive LASSO in GWAS, including a performance test and a GWAS of meat colour in pigs. Chapter 3 introduces a GWAS of fetal response to type 2 PRRSV challenge using a multi-marker generalized linear model. Chapter 4 introduces a genome-wide transmission disequilibrium test of fetal autolysis in the same population, which integrated pedigree data that was not used in the previous GWAS. The studies in these three chapters used SNP data. Chapter 5 introduces a study beyond SNPs, where a new genome mapping method is proposed to detect whole-genome structural variations. Chapter 6 gives a summary of the four studies, and discusses the outputs and future work.

1.4 References

Aldhous, P. 1993. Managing the genome data deluge. Science **262**: 502–3.

Alkan, C., Coe, B.P., and Eichler, E.E. 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. **12**: 363–376.

Archibald, A.L., Bolund, L., Churcher, C., Fredholm, M., Groenen, M. a M., Harlizius, B., Lee, K.-T., Milan, D., Rogers, J., Rothschild, M.F., Uenishi, H., Wang, J., and Schook, L.B. 2010. Pig genome sequence--analysis and publication strategy. BMC Genomics **11**: 438.

Archibald, A.L., Haley, C.S., Brown, J.F., Couperwhite, S., McQueen, H.A., Nicholson, D., Coppieters, W., Van de Weghe, A., Stratil, A., Winterø, A.K., Fredholm, M., Larsen, N.J., Nielsen, V.H., Milan, D., Woloszyn, N., Robic, A., Dalens, M., Riquet, J., Gellin, J., Caritez, J.C., Burgaud, G., Ollivier, L., Bidanel, J.P., Vaiman, M., Renard, C., Geldermann, H., Davoli, R., Ruyter, D., Verstege, E.J.M., Groenen, M.A.M., Davies, W., Høyheim, B., Keiserud, A., Andersson, L., Ellegren, H., Johansson, M., Marklund, L., Miller, J.R., Anderson Dear, D. V., Signer, E., Jeffreys, A.J., Moran, C., Le Tissier, P., Muladno, Rothschild, M.F., Tuggle, C.K., Vaske, D., Helm, J., Liu, H.C., Rahman, A., Yu, T.P., Larson, R.G., and Schmitz, C.B. 1995. The PiGMaP consortium linkage map of the pig (Sus scrofa). Mamm. Genome **6**: 157–175.

Ayers, K.L., and Cordell, H.J. 2010. SNP selection in genome-wide and candidate gene studies via penalized logistic regression. Genet. Epidemiol. **34**: 879–891.

de Bakker, P.I.W., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J., and Altshuler, D. 2005. Efficiency and power in genetic association studies. Nat. Genet. **37**: 1217–1223.

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

Basu, S., Pan, W., Shen, X., and Oetting, W.S. 2011. Multilocus association testing with penalized

regression. Genet. Epidemiol. **35**: 755–765.

Bateson, W., Saunders, E.R., and Punnett, R.C. 1905. Experimental studies in the physiology of heredity. Reports to Evol. Comm. R. Soc. **2**.

Battagin, M., Gorjanc, G., Faux, A.-M., Johnston, S.E., and Hickey, J.M. 2016. Effect of manipulating recombination rates on response to selection in livestock breeding programs. Genet. Sel. Evol. **48**: 44.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B. Stat. Methodol. **57**: 289–300.

Bickhart, D.M., and Liu, G.E. 2014. The challenges and importance of structural variation detection in livestock. Front. Genet. **5**: 1–14.

Bovine HapMap Consortium, Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse, W., Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M., Lien, S., Matukumalli, L.K., McEwan, J.C., Nazareth, L. V., Schnabel, R.D., Weinstock, G.M., Wheeler, D.A., Ajmone-Marsan, P., Boettcher, P.J., Caetano, A.R., Garcia, J.F., Hanotte, O., Mariani, P., Skow, L.C., Sonstegard, T.S., Williams, J.L., Diallo, B., Hailemariam, L., Martinez, M.L., Morris, C.A., Silva, L.O.C., Spelman, R.J., Mulatu, W., Zhao, K., Abbey, C.A., Agaba, M., Araujo, F.R., Bunch, R.J., Burton, J., Gorni, C., Olivier, H., Harrison, B.E., Luff, B., Machado, M.A., Mwakaya, J., Plastow, G., Sim, W., Smith, T., Thomas, M.B., Valentini, A., Williams, P., Womack, J., Woolliams, J.A., Liu, Y., Qin, X., Worley, K.C., Gao, C., Jiang, H., Moore, S.S., Ren, Y., Song, X.-Z., Bustamante, C.D., Hernandez, R.D., Muzny, D.M., Patil, S., San Lucas, A., Fu, Q., Kent, M.P., Vega, R., Matukumalli, A., McWilliam, S., Sclep, G., Bryc, K., Choi, J., Gao, H., Grefenstette, J.J., Murdoch, B., Stella, A., Villa-Angulo, R., Wright, M., Aerts, J., Jann, O., Negrini, R., Goddard, M.E., Hayes, B.J., Bradley, D.G., Barbosa da Silva, M.,

Lau, L.P.L., Liu, G.E., Lynn, D.J., Panzitta, F., and Dodds, K.G. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science **324**: 528–32.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S.B., Krstic, P.S., Lindsay, S., Ling, X.S., Mastrangelo, C.H., Meller, A., Oliver, J.S., Pershin, Y. V, Ramsey, J.M., Riehn, R., Soni, G. V, Tabard-Cossa, V., Wanunu, M., Wiggin, M., and Schloss, J.A. 2008. The potential and challenges of nanopore sequencing. Nat. Biotechnol. **26**: 1146–1153.

Bühlmann, P., and van de Geer, S. 2011. Statistics for High-Dimensional Data.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics **11**: 94.

Chaisson, M.J.P., Wilson, R.K., and Eichler, E.E. 2015. Genetic variation and the de novo assembly of human genomes. Nat. Rev. Genet. **16**: 627–640.

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J.S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.-C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B.R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R.J., and Georges, M. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. Nat. Genet. **40**: 449–54.

Collins, F.S. 1999. Shattuck lecture--medical and societal consequences of the Human Genome Project. N. Engl. J. Med. **341**: 28–37.

Collins, F.S., and McKusick, V.A. 2001. Implications of the Human Genome Project for medical science. JAMA **285**: 540–4.

Cowan, C.M., Dentine, M.R., Ax, R.L., and Schuler, L.A. 1990. Structural variation around prolactin gene linked to quantitative traits in an elite Holstein sire family. : 577–582.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S., and Myers, R.M. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science **250**: 245–50.

Crick, F.H.C., Barnett, L., Brenner, S., and Watts-Tobin, R.J. 1961. General nature of the genetic code for proteins. Nature **192**: 1227–1232.

Cui, X.F., Li, H.H., Goradia, T.M., Lange, K., Kazazian, H.H., Galas, D., and Arnheim, N. 1989. Single-sperm typing: determination of genetic distance between the G gamma-globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. Proc. Natl. Acad. Sci. U. S. A. **86**: 9389–93.

Dear, P.H., Bankier, A.T., and Piper, M.B. 1998. A high-resolution metric HAPPY map of human chromosome 14. Genomics **48**: 232–41.

Dear, P.H., and Cook, P.R. 1993. Happy mapping: linkage mapping using a physical analogue of meiosis. Nucleic Acids Res. **21**: 13–20.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. **82 E-Suppl**: E313-328.

De Donato, M., Peters, S.O., Mitchell, S.E., Hussain, T., and Imumorin, I.G. 2013. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PLoS One **8**: e62137.

Dong, L., Meng, Y., Wang, J., and Liu, Y. 2014. Evaluation of droplet digital PCR for characterizing plasmid reference material used for quantifying ammonia oxidizers and denitrifiers. Anal. Bioanal. Chem. **406**: 1701–1712.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J., and Wang, W. 2012. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat. Biotechnol. **31**: 135–141.

Duggal, P., Gillanders, E.M., Holmes, T.N., and Bailey-Wilson, J.E. 2008. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics **9**: 516.

Dupuis, J., and O'Donnell, C.J. 2007. Interpreting results of large-scale genetic association studies. JAMA **297**: 529.

Durkin, K., Coppieters, W., Drögemüller, C., Ahariz, N., Cambisano, N., Druet, T., Fasquelle, C., Haile, A., Horin, P., Huang, L., Kamatani, Y., Karim, L., Lathrop, M., Moser, S., Oldenbroek, K., Rieder, S., Sartelet, A., Sölkner, J., Stålhammar, H., Zelenika, D., Zhang, Z., Leeb, T., Georges, M., and Charlier, C. 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. Nature **482**: 81–4.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D.P., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C.S., Bowman, R., SEARCH collaborators, Meyer, K.B., Haiman, C.A., Kolonel, L.K., Henderson, B.E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C.-Y., Wu, P.-E., Wang, H.-C., Eccles, D., Evans, D.G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M.R., Rahman, N., Chenevix-Trench, G., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska,

B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K.-Y., Noh, D.-Y., Ahn, S.-H., Hunter, D.J., Hankinson, S.E., Cox, D.G., Hall, P., Wedren, S., Liu, J., Low, Y.-L., Bogdanova, N., Schürmann, P., Dörk, T., Tollenaar, R.A.E.M., Jacobi, C.E., Devilee, P., Klijn, J.G.M., Sigurdson, A.J., Doody, M.M., Alexander, B.H., Zhang, J., Cox, A., Brock, I.W., MacPherson, G., Reed, M.W.R., Couch, F.J., Goode, E.L., Olson, J.E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J.L., McCredie, M., Southey, M., Giles, G.G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y.-D., Spurdle, A.B., Beesley, J., Chen, X., kConFab, AOCS Management Group, Mannermaa, A., Kosma, V.-M., Kataja, V., Hartikainen, J., Day, N.E., Cox, D.R., and Ponder, B.A.J. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature **447**: 1087–93.

El-Metwally, S., Ouda, O.M., and Helmy, M. 2014. Next Generation Sequencing Technologies and Challenges in Sequence Assembly.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One **6**: e19379.

Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Muzny, D.M., Weinstock, G.M., Adelson, D.L., Eichler, E.E., Elnitski, L., Guigo, R., Hamernik, D.L., Kappes, S.M., Lewin, H.A., Lynn, D.J., Nicholas, F.W., Reymond, A., Rijnkels, M., Skow, L.C., Zdobnov, E.M., Schook, L., Womack, J., Alioto, T., Antonarakis, S.E., Astashyn, A., Chapple, C.E., Chen, H.-C., Chrast, J., Camara, F., Ermolaeva, O., Henrichsen, C.N., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Kokocinski, F., Landrum, M., Maglott, D., Pruitt, K., Sapojnikov, V., Searle, S.M., Solovyev, V., Souvorov, A., Ucla, C., Wyss, C., Anzola, J.M., Gerlach, D., Elhaik, E.,

Graur, D., Reese, J.T., Edgar, R.C., McEwan, J.C., Payne, G.M., Raison, J.M., Junier, T., Kriventseva, E. V., Eyras, E., Plass, M., Donthu, R., Larkin, D.M., Reecy, J., Yang, M.Q., Chen, L., Cheng, Z., Chitko-McKown, C.G., Liu, G.E., Matukumalli, L.K., Song, J., Zhu, B., Bradley, D.G., Brinkman, F.S.L., Lau, L.P.L., Whiteside, M.D., Walker, A., Wheeler, T.T., Casey, T., German, J.B., Lemay, D.G., Maqbool, N.J., Molenaar, A.J., Seo, S., Stothard, P., Baldwin, C.L., Baxter, R., Brinkmeyer-Langford, C.L., Brown, W.C., Childers, C.P., Connelley, T., Ellis, S.A., Fritz, K., Glass, E.J., Herzig, C.T.A., Iivanainen, A., Lahmers, K.K., Bennett, A.K., Dickens, C.M., Gilbert, J.G.R., Hagen, D.E., Salih, H., Aerts, J., Caetano, A.R., Dalrymple, B., Garcia, J.F., Gill, C.A., Hiendleder, S.G., Memili, E., Spurlock, D., Williams, J.L., Alexander, L., Brownstein, M.J., Guan, L., Holt, R.A., Jones, S.J.M., Marra, M.A., Moore, R., Moore, S.S., Roberts, A., Taniguchi, M., Waterman, R.C., Chacko, J., Chandrabose, M.M., Cree, A., Dao, M.D., Dinh, H.H., Gabisi, R.A., Hines, S., Hume, J., Jhangiani, S.N., Joshi, V., Kovar, C.L., Lewis, L.R., Liu, Y. -s., Lopez, J., Morgan, M.B., Nguyen, N.B., Okwuonu, G.O., Ruiz, S.J., Santibanez, J., Wright, R.A., Buhay, C., Ding, Y., Dugan-Rocha, S., Herdandez, J., Holder, M., Sabo, A., Egan, A., Goodell, J., Wilczek-Boney, K., Fowler, G.R., Hitchens, M.E., Lozado, R.J., Moen, C., Steffen, D., Warren, J.T., Zhang, J., Chiu, R., Schein, J.E., Durbin, K.J., Havlak, P., Jiang, H., Liu, Y., Qin, X., Ren, Y., Shen, Y., Song, H., Bell, S.N., Davis, C., Johnson, A.J., Lee, S., Nazareth, L. V., Patel, B.M., Pu, L.-L., Vattathil, S., Williams, R.L., Curry, S., Hamilton, C., Sodergren, E., Wheeler, D.A., Barris, W., Bennett, G.L., Eggen, A., Green, R.D., Harhay, G.P., Hobbs, M., Jann, O., Keele, J.W., Kent, M.P., Lien, S., McKay, S.D., McWilliam, S., Ratnakumar, A., Schnabel, R.D., Smith, T., Snelling, W.M., Sonstegard, T.S., Stone, R.T., Sugimoto, Y., Takasuga, A., Taylor, J.F., Van Tassell, C.P., MacNeil, M.D., Abatepaulo, A.R.R., Abbey,

C.A., Ahola, V., Almeida, I.G., Amadio, A.F., Anatriello, E., Bahadue, S.M., Biase, F.H., Boldt, C.R., Carroll, J.A., Carvalho, W.A., Cervelatti, E.P., Chacko, E., Chapin, J.E., Cheng, Y., Choi, J., Colley, A.J., de Campos, T.A., De Donato, M., Santos, I.K.F. d. M., de Oliveira, C.J.F., Deobald, H., Devinoy, E., Donohue, K.E., Dovc, P., Eberlein, A., Fitzsimmons, C.J., Franzin, A.M., Garcia, G.R., Genini, S., Gladney, C.J., Grant, J.R., Greaser, M.L., Green, J.A., Hadsell, D.L., Hakimov, H.A., Halgren, R., Harrow, J.L., Hart, E.A., Hastings, N., Hernandez, M., Hu, Z.-L., Ingham, A., Iso-Touru, T., Jamis, C., Jensen, K., Kapetis, D., Kerr, T., Khalil, S.S., Khatib, H., Kolbehdari, D., Kumar, C.G., Kumar, D., Leach, R., Lee, J.C.-M., Li, C., Logan, K.M., Malinverni, R., Marques, E., Martin, W.F., Martins, N.F., Maruyama, S.R., Mazza, R., McLean, K.L., Medrano, J.F., Moreno, B.T., More, D.D., Muntean, C.T., Nandakumar, H.P., Nogueira, M.F.G., Olsaker, I., Pant, S.D., Panzitta, F., Pastor, R.C.P., Poli, M.A., Poslusny, N., Rachagani, S., Ranganathan, S., Razpet, A., Riggs, P.K., Rincon, G., Rodriguez-Osorio, N., Rodriguez-Zas, S.L., Romero, N.E., Rosenwald, A., Sando, L., Schmutz, S.M., Shen, L., Sherman, L., Southey, B.R., Lutzow, Y.S., Sweedler, J. V., Tammen, I., Telugu, B.P.V.L., Urbanski, J.M., Utsunomiya, Y.T., Verschoor, C.P., Waardenberg, A.J., Wang, Z., Ward, R., Weikard, R., Welsh, T.H., White, S.N., Wilming, L.G., Wunderlich, K.R., Yang, J., and Zhao, F.-Q. 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science. **324**: 522–528.

Fan, B., Onteru, S.K., Du, Z.Q., Garrick, D.J., Stalder, K.J., and Rothschild, M.F. 2011. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. PLoS One 6: e14726.

Falconer, D.S., and Mackay, T.F.C. 1996. Introduction to quantitative genetics. Fourth Edi.

Fearnhead, P., and Donnelly, P. 2001. Estimating recombination rates from population genetic

data. Genetics **159**: 1299–318.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V., Weiler, J., O'Brien, P., and MacLennan, D. 1991. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science **253**: 448–451.

Gao, H., Wu, Y., Li, J., Li, H., Li, J., and Yang, R. 2014. Forward LASSO analysis for high-order interactions in genome-wide association study. Brief. Bioinform. **15**: 552–561.

Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D., and Province, M.A. 2009. Avoiding the high Bonferroni penalty in genome-wide association studies. Genet. Epidemiol. **34**: 100–105.

Gocayne, J., Robinson, D.A., FitzGerald, M.G., Chung, F.Z., Kerlavage, A.R., Lentes, K.U., Lai, J., Wang, C.D., Fraser, C.M., and Venter, J.C. 1987. Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic receptors obtained by automated DNA sequence analysis: further evidence for a multigene family. Proc. Natl. Acad. Sci. U. S. A. **84**: 8296–300.

Goddard, M.E., and Hayes, B.J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. **10**: 381–391.

Griffith, M., Griffith, O.L., Mwenifumbo, J., Goya, R., Morrissy, a S., Morin, R.D., Corbett, R., Tang, M.J., Hou, Y., Pugh, T.J., Robertson, G., Chittaranjan, S., Ally, A., Asano, J.K., Chan, S.Y., Li, H.I., McDonald, H., Teague, K., Zhao, Y., Zeng, T., Delaney, A., Hirst, M., Morin, G.B., Jones, S.J.M., Tai, I.T., and Marra, M.A. 2010. Alternative expression analysis by RNA sequencing. Nat. Methods **7**: 843–847.

Groenen, M.A., Cheng, H.H., Bumstead, N., Benkel, B.F., Briles, W.E., Burke, T., Burt, D.W., Crittenden, L.B., Dodgson, J., Hillel, J., Lamont, S., de Leon, A.P., Soller, M., Takahashi, H., and Vignal, A. 2000. A consensus linkage map of the chicken genome. Genome Res. **10**:

137–47.

Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A.F., Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R.C., Clee, C., Crooijmans, R.P.M.A., Dawson, H.D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G.J., Fowler, K.E., Fredholm, M., Fritz, E., Gilbert, J.G.R., Giuffra, E., Gorodkin, J., Griffin, D.K., Harrow, J.L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S.J., Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H.A., Li, Y., Liu, W., Loveland, J.E., Lu, Y., Lunney, J.K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M.P., Narayan, J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J.M., Rodriguez-Zas, S., Rohrer, G.A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J.G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B.R., Sperber, G., Stadler, P., Sweedler, J. V., Tafer, H., Thomsen, B., Wali, R., Wang, J., Wang, J., White, S., Xu, X., Yerle, M., Zhang, G., Zhang, J., Zhang, J., Zhao, S., Rogers, J., Churcher, C., and Schook, L.B. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature **491**: 393–398.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D., Helgason, A.,

Rafnar, T., Bergthorsson, J.T., Agnarsson, B. a, Baker, A., Sigurdsson, A., Benediktsdottir, K.R., Jakobsdottir, M., Xu, J., Blondal, T., Kostic, J., Sun, J., Ghosh, S., Stacey, S.N., Mouy, M., Saemundsdottir, J., Backman, V.M., Kristjansson, K., Tres, A., Partin, A.W., Albers-Akkers, M.T., Godino-Ivan Marcos, J., Walsh, P.C., Swinkels, D.W., Navarrete, S., Isaacs, S.D., Aben, K.K., Graif, T., Cashy, J., Ruiz-Echarri, M., Wiley, K.E., Suarez, B.K., Witjes, J.A., Frigge, M., Ober, C., Jonsson, E., Einarsson, G. V, Mayordomo, J.I., Kiemeney, L. a, Isaacs, W.B., Catalona, W.J., Barkardottir, R.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., and Stefansson, K. 2007. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat. Genet. **39**: 631–637.

Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G., and Chee, M.S. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. Nat. Genet. **37**: 549–554.

Hackett, C.A., McLean, K., and Bryan, G.J. 2013. Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PLoS One **8**: e63939.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning.

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

Hayes, B., and Goddard, M.E. 2003. Evaluation of marker assisted selection in pig enterprises. Livest. Prod. Sci. **81**: 197–211.

Heid, I.M., and Winkler, T.W. 2016. A multitrait GWAS sheds light on insulin resistance. Nat. Genet. **49**: 7–8.

Hickey, J.M. 2013. Sequencing millions of animals for genomic selection 2.0. J. Anim. Breed. Genet. **130**: 331–332.

Hill, W.G. 2000. Maintenance of quantitative genetic variation in animal breeding programmes.

Livest. Prod. Sci. **63**: 99–109.

Hindson, C.M., Chevillet, J.R., Briggs, H. a, Gallichotte, E.N., Ruf, I.K., Hindson, B.J., Vessella, R.L., and Tewari, M. 2013. Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat. Methods **10**: 1003–1005.

Hoh, J., and Ott, J. 2005. Genetic linkage mapping. *in* Encyclopedia of Life Sciences.

International HapMap Consortium, Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.-F., Galver, L.M., Fan, J.-B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.-F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Kwok, P.-Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Song, Y.Q., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I.W., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C.,

Varilly, P., Altshuler, D., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Tsunoda, T., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Gibbs, R.A., Belmont, J.W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Wheeler, D.A., Yakub, I., Gabriel, S.B., Onofrio, R.C., Richter, D.J., Ziaugra, L., Birren, B.W., Daly, M.J., Altshuler, D., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., and Stewart, J. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature **449**: 851–61.

Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985. Hypervariable "minisatellite" regions in human DNA. Nature 314: 67–73.

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y., and Zhang, Q. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One **5**: e13661.

Kan, Y.W., and Dozy, A.M. 1978. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. Proc. Natl. Acad. Sci. U. S. A. **75**: 5631–5.

Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., Liu, W., Yang, G., Di, X., Ryder, T., He, Z., Surti, U., Phillips, M.S., Boyce-Jacino, M.T., Fodor, S.P., and Jones, K.W. 2003. Large-scale genotyping of complex DNA. Nat. Biotechnol. **21**: 1233–1237.

Kerem, B.S., Buchanan, J.A., Durie, P., Corey, M.L., Levison, H., Rommens, J.M., Buchwald, M., and Tsui, L.C. 1989. DNA marker haplotype association with pancreatic sufficiency in cystic fibrosis. Am. J. Hum. Genet. **44**: 827–34.

Kirkpatrick, B., Reshef, Y., Finucane, H., Jiang, H., Zhu, B., and Karp, R.M. 2012. Comparing pedigree graphs. J. Comput. Biol. **19**: 998–1014.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z., Tanzer, A., Saunders, a C.E., Chi, J., Yang, F., Carter, N.P., Hurles, M.E., Weissman, S.M., Harkins, T.T., Gerstein, M.B., Egholm, M., and Snyder, M. 2009. Paired-end mapping reveals extensive structural variation in the human genome. Science **318**: 420–6.

Lander, E.S., and Botstein, S. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**: 185.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J.,

Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M. V, Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C.,

Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E. V, Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J. V, Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., and International Human Genome Sequencing Consortium 2001. Initial sequencing and analysis of the human genome. Nature **409**: 860–921.

Lazzeroni, L.C., Arnheim, N., Schmitt, K., and Lange, K. 1994. Multipoint mapping calculations for sperm-typing data. Am. J. Hum. Genet. **55**: 431–6.

Li, H.H., Gyllensten, U.B., Cui, X.F., Saiki, R.K., Erlich, H.A., and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature **335**: 414–7.

Li, L., and Yin, X. 2008. Sliced inverse regression with regularizations. Biometrics **64**: 124–131.

Li, M.-X., Yeung, J.M.Y., Cherny, S.S., and Sham, P.C. 2012. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum. Genet. **131**: 747–56.

Lillehammer, M., Meuwissen, T.H.E., and Sonesson, A.K. 2011. Genomic selection for maternal traits in pigs. J. Anim. Sci. **89**: 3908–3916.

Lim, A., Dimalanta, E.T., Potamousis, K.D., Yen, G., Apodoca, J., Tao, C., Lin, J., Qi, R., Skiadas,

J., Ramanathan, A., Perna, N.T., Plunkett, G., Burland, V., Mau, B., Hackett, J., Blattner, F.R., Anantharaman, T.S., Mishra, B., and Schwartz, D.C. 2001. Shotgun optical maps of the whole Escherichia coli O157:H7 genome. Genome Res. **11**: 1584–93.

Litt, M., and Luty, J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44: 397–401.

Liu, J., Chen, J., and Ye, J. 2009. Large-scale sparse logistic regression. *in* Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09.

Lourenco, D.A.L., Tsuruta, S., Fragomeni, B.O., Masuda, Y., Aguilar, I., Legarra, A., Bertrand, J.K., Amen, T.S., Wang, L., Moser, D.W., and Misztal, I. 2015. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. J. Anim. Sci. **93**: 2653.

Lu, S., Zong, C., Fan, W., Yang, M., Li, J., Chapman, A.R., Zhu, P., Hu, X., Xu, L., Yan, L., Bai, F., Qiao, J., Tang, F., Li, R., and Xie, X.S. 2012. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. Science **338**: 1627–30.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K.T. 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One **7**: e30087.

MacLeod, I.M., Bowman, P.J., Vander Jagt, C.J., Haile-Mariam, M., Kemper, K.E., Chamberlain, A.J., Schrooten, C., Hayes, B.J., and Goddard, M.E. 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics **17**: 144.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. 2008. RNA-seq: An

assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. **18**: 1509–1517.

Marques, E., de Givry, S., Stothard, P., Murdoch, B., Wang, Z., Womack, J., and Moore, S.S. 2007. A high resolution radiation hybrid map of bovine chromosome 14 identifies scaffold rearrangement in the latest bovine assembly. BMC Genomics **8**: 254.

Maxam, A.M., and Gilbert, W. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U. S. A. **74**: 560–4.

Meuwissen, T., Hayes, B., and Goddard, M. 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front. **6**: 6.

Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics **157**: 1819–29.

Meuwissen, T.H.E., and Goddard, M.E. 1996. The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. 28: 161.

Min Jou, W., Haegeman, G., Ysebaert, M., and Fiers, W. 1972. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature **237**: 82–8.

Morgan, T.H. 1911. Random segregation versus coupling in Mendelian inheritance. Science **34**: 384.

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., and Kanaya, S. 2011. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. **39**: e90–e90.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., and Kumlin, E. 1987. Variable number of tandem repeat (VNTR) markers for

human gene mapping. Science **235**: 1616–22.

Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. 2011. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. **12**: 443–451.

Nielsen, V.H., Bendixen, C., Arnbjerg, J., Sørensen, C.M., Jensen, H.E., Shukri, N.M., and Thomsen, B. 2000. Abnormal growth plate function in pigs carrying a dominant mutation in type X collagen. Mamm. Genome **11**: 1087–1092.

Ott, J., Wang, J., and Leal, S.M. 2015. Genetic linkage analysis in the age of whole-genome sequencing. Nat. Rev. Genet. **16**: 275–284.

Pavšič, J., Žel, J., and Milavec, M. 2016. Assessment of the real-time PCR and different digital PCR platforms for DNA quantification. Anal. Bioanal. Chem. **408**: 107–121.

Pekin, D., Skhiri, Y., Baret, J.-C., Le Corre, D., Mazutis, L., Ben Salem, C., Millot, F., El Harrak, A., Hutchison, J.B., Larson, J.W., Link, D.R., Laurent-Puig, P., Griffiths, A.D., and Taly, V. 2011. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab Chip **11**: 2156.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., and Albertson, D.G. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. **20**: 207–11.

Pole, J.C.M., McCaughan, F., Newman, S., Howarth, K.D., Dear, P.H., and Edwards, P.A.W. 2011. Single-molecule analysis of genome rearrangements in cancer. Nucleic Acids Res. **39**: e85–e85.

Quinn, G.P., and Keough, M.J. 2002. Experimental Design and Data Analysis for Biologists.

Reich, D.E., and Goldstein, D.B. 2001. Detecting association in a case-control study while

correcting for population stratification. Genet. Epidemiol. **20**: 4–16.

Rhodes, J., Beale, M.A., and Fisher, M.C. 2014. Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of Cryptococcus neoformans using Illumina HiSeq. PLoS One **9**: e113501.

Robinson, G.K.G. 1991. That BLUP is a good thing: The estimation of random effects. Stat. Sci. **6**: 15–32.

Rohrer, G.A., Alexander, L.J., Keele, J.W., Smith, T.P., and Beattie, C.W. 1994. A microsatellite linkage map of the porcine genome. Genetics **136**: 231–45.

Ronaghi, M. 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res. **11**: 3–11.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., Altshuler, D., and International SNP Map Working Group 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409**: 928–33.

Sanger, F., and Coulson, A.R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. **94**: 441–448.

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., and Petersen, G.B. 1982. Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. **162**: 729–773.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**: 467–70.

Schiavo, G., Dolezal, M.A., Scotti, E., Bertolini, F., Calò, D.G., Galimberti, G., Russo, V., and Fontanesi, L. 2014. Copy number variants in Italian Large White pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. Anim. Genet. **45**: 745–749.

Schwartz, D.C., Li, X., Hernandez, L.I., Ramnarain, S.P., Huff, E.J., and Wang, Y.K. 1993. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science **262**: 110–4.

Smith, C. 1967. Improvement of metric traits through specific genetic loci. Anim. Prod. **9**: 349–358.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., and Hood, L.E. 1986. Fluorescence detection in automated DNA sequence analysis. Nature **321**: 674–9.

Snelling, W.M., Chiu, R., Schein, J.E., Hobbs, M., Abbey, C.A., Adelson, D.L., Aerts, J., Bennett, G.L., Bosdet, I.E., Boussaha, M., Brauning, R., Caetano, A.R., Costa, M.M., Crawford, A.M., Dalrymple, B.P., Eggen, A., Everts-van der Wind, A., Floriot, S., Gautier, M., Gill, C.A., Green, R.D., Holt, R., Jann, O., Jones, S.J., Kappes, S.M., Keele, J.W., de Jong, P.J., Larkin, D.M., Lewin, H.A., McEwan, J.C., McKay, S., Marra, M.A., Mathewson, C.A., Matukumalli, L.K., Moore, S.S., Murdoch, B., Nicholas, F.W., Osoegawa, K., Roy, A., Salih, H., Schibler, L., Schnabel, R.D., Silveri, L., Skow, L.C., Smith, T.P., Sonstegard, T.S., Taylor, J.F., Tellam, R., Van Tassell, C.P., Williams, J.L., Womack, J.E., Wye, N.H., Yang, G., and Zhao, S. 2007. A physical map of the bovine genome. Genome Biol. **8**: R165.

Spielman, R.S., McGinnis, R.E., and Ewens, W.J. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am.

J. Hum. Genet. **52**: 506–16.

Staaf, J., Vallon-Christersson, J., Lindgren, D., Juliusson, G., Rosenquist, R., Höglund, M., Borg, A., and Ringnér, M. 2008. Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. BMC Bioinformatics **9**: 409.

Steemers, F.J., Chang, W., Lee, G., Barker, D.L., Shen, R., and Gunderson, K.L. 2006. Whole-genome genotyping with the single-base extension assay. Nat. Methods **3**: 31–33.

van der Steen, H.A.M., Prall, G.F.W., and Plastow, G.S. 2005. Application of genomics to the pork industry. J. Anim. Sci. **83**: E1–E8.

Strain, M.C., Lada, S.M., Luong, T., Rought, S.E., Gianella, S., Terry, V.H., Spina, C.A., Woelk, C.H., and Richman, D.D. 2013. Highly precise measurement of HIV DNA by droplet digital PCR. PLoS One **8**: e55943.

Sturtevant, A.H. 1913. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. **14**: 43–59.

Sullivan, P.F., Lin, D., Tzeng, J.-Y., van den Oord, E., Perkins, D., Stroup, T.S., Wagner, M., Lee, S., Wright, F.A., Zou, F., Liu, W., Downing, A.M., Lieberman, J., and Close, S.L. 2008. Genomewide association for schizophrenia in the CATIE study: results of stage 1. Mol. Psychiatry **13**: 570–584.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. 2011. Differential expression in RNA-seq: A matter of depth. Genome Res. **21**: 2213–2223.

Veldman-Jones, M.H., Brant, R., Rooney, C., Geh, C., Emery, H., Harbron, C.G., Wappett, M., Sharpe, A., Dymond, M., Barrett, J.C., Harrington, E.A., and Marshall, G. 2015. Evaluating robustness and sensitivity of the NanoString technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. Cancer Res. **75**: 2587–2593.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V, Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D.,

Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K. V, Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. 2001. The sequence of the human genome. Science **291**: 1304–51.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. Van De, Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23**: 4407–4414.

Wang, L., Bai, B., Liu, P., Huang, S.Q., Wan, Z.Y., Chua, E., Ye, B., and Yue, G.H. 2017. Construction of high-resolution recombination maps in Asian seabass. BMC Genomics **18**: 63.

Warr, A., Hall, R., Kim, K.S., Elizabeth, T., Koren, S., Phillippy, A.M., Bickhart, D.M., Rosen, B.D., Schroeder, S.G., Hume, D.A., Talbot, R., Rund, L., Schook, L.B., Chow, W., Howe,

K., Nonneman, D., Rohrer, G.A., Putnam, N., Green, E., Watson, M., Smith, T.P.L., and Archibald, A.L. 2017. Exploiting Long Read Sequencing Technologies to Establish High Quality Highly Contiguous Pig Reference Genome Assemblies. *in* Plant and Animal Genome Conference XXIV.

Watson, J.D., and Crick, F.H.C. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature **171**: 737–738.

Weber, J.L., and May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388–96.

Weng, Z.-Q., Saatchi, M., Schnabel, R.D., Taylor, J.F., and Garrick, D.J. 2014. Recombination locations and rates in beef cattle assessed from parent-offspring pairs. Genet. Sel. Evol. **46**: 34.

Westfall, P.H., and Young, S.S. 1993. Resampling-based multiple testing: Examples and methods for p-value adjustment.

Westfall, P.H., Young, S.S., and Wright, S.P. 1993. On adjusting P-values for multiplicity. Biometrics **49**: 941.

Wiggans, G.R., VanRaden, P.M., and Cooper, T.A. 2011. The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. **94**: 3202–3211.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S. V 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18**: 6531–6535.

Wu, T.T., Chen, Y.F., Hastie, T., Sobel, E., and Lange, K. 2009. Genome-wide association analysis by lasso penalized logistic regression. Bioinformatics **25**: 714–721.

Xing, J., Gao, H., Wu, Y., Wu, Y., Li, H., and Yang, R. 2014. Generalized linear model for

mapping discrete trait loci implemented with LASSO algorithm. PLoS One **9**: e106985.

Xu, S. 2013. Principles of Statistical Genomics.

Yekutieli, D., and Benjamini, Y. 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. **29**: 1165–1188.

Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T., Bannai, M., Tokunaga, K., Sonoda, S., Komaki, R., Ihara, M., Saenko, V.A., Alipov, G.K., Sekine, I., Komatsu, K., Takahashi, H., Nakashima, M., Sosonkina, N., Mapendano, C.K., Ghadami, M., Nomura, M., Liang, D.-S., Miwa, N., Kim, D.-K., Garidkhuu, A., Natsume, N., Ohta, T., Tomita, H., Kaneko, A., Kikuchi, M., Russomando, G., Hirayama, K., Ishibashi, M., Takahashi, A., Saitou, N., Murray, J.C., Saito, S., Nakamura, Y., and Niikawa, N. 2006. A SNP in the ABCC11 gene is the determinant of human earwax type. Nat. Genet. 38: 324–330.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. 1992. Whole genome amplification from a single cell: implications for genetic analysis. Proc. Natl. Acad. Sci. U. S. A. **89**: 5847–51.

Zheng, G.X.Y., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M., Mudivarti, P.A., Wyatt, P.W., Bharadwaj, R., Makarewicz, A.J., Li, Y., Belgrader, P., Price, A.D., Lowe, A.J., Marks, P., Vurens, G.M., Hardenbol, P., Montesclaros, L., Luo, M., Greenfield, L., Wong, A., Birch, D.E., Short, S.W., Bjornson, K.P., Patel, P., Hopmans, E.S., Wood, C., Kaur, S., Lockwood, G.K., Stafford, D., Delaney, J.P., Wu, I., Ordonez, H.S., Grimes, S.M., Greer, S., Lee, J.Y., Belhocine, K., Giorda, K.M., Heaton, W.H., McDermott, G.P., Bent, Z.W., Meschi, F., Kondov, N.O., Wilson, R., Bernate, J.A., Gauby, S., Kindwall, A.,

Bermejo, C., Fehr, A.N., Chan, A., Saxonov, S., Ness, K.D., Hindson, B.J., and Ji, H.P. 2016. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat. Biotechnol. **34**: 303–311.

Zhou, S., Wei, F., Nguyen, J., Bechner, M., Potamousis, K., Goldstein, S., Pape, L., Mehan, M.R., Churas, C., Pasternak, S., Forrest, D.K., Wise, R., Ware, D., Wing, R.A., Waterman, M.S., Livny, M., and Schwartz, D.C. 2009. A single molecule scaffold for the maize genome. PLoS Genet. **5**: e1000711.

Zhu, P., and Craighead, H.G. 2012. Zero-mode waveguides for single-molecule analysis. Annu. Rev. Biophys. **41**: 269–93.

Chapter 2 Performance of the adaptive LASSO and its application in GWAS of meat colour in pigs

This chapter mainly discusses the application of the adaptive least absolute shrinkage and selection operator ("the adaptive LASSO") in genome-wide association studies (GWAS). The results and discussions of this chapter provides useful information for the selection of statistical tools in the context of GWAS.

As reviewed in Chapter 1, multiple linear regression is an important method for genome-wide association studies (GWAS) because of its ability to model the polygenic nature of quantitative traits. However, as the density of available genomic variants is getting higher and higher, GWAS using the high-dimensional linear model seems to be an example of the "curse of dimensionality" in genomic analysis (Jiang and Neapolitan 2012), where the high dimension of genomic data significantly hinders the application of multiple linear regression. While the problem can partially be solved through progress in biology and genomics (e.g. better functional annotation of genome can narrow down our list of candidate genomic variants), advanced statistical tools also provide a toolkit for GWAS with high-dimensional genomic data. The adaptive LASSO has been widely used in the analysis of high-dimensional linear regression (Bühlmann and van de Geer 2011), and may serve as an good option in GWAS.

We conducted two studies to explore the application of the adaptive LASSO in GWAS. In study one, we tested the performance of the adaptive LASSO in a public simulated dataset (QTLMAS2010). The adaptive LASSO showed higher detection power (i.e. higher true positive rate) than other methods (e.g. BayesC/CPi), while its precision is relatively low. In study two, we

applied adaptive LASSO in the GWAS of meat colour in pigs. A total of 20 genomic regions were detected to be associated with multiple meat colour traits. These candidate regions were supported by previous GWAS results, known QTL of meat quality and lipid metabolism, and genes functionally related to meat colour traits. The study has been accepted by the Canadian Journal of Animal Science for publication[1]. I also contributed to the analysis and helped co-author two publications related to study two (Appendix 2).

2.1 STUDY 1 – Evaluation of the performance of the adaptive LASSO

2.1.1 Introduction

Currently statistical techniques for high-dimensional regression include traditional model selection methods (e.g. stepwise selection), Bayesian regression techniques, and regulation techniques (Berk 2008). The adaptive LASSO is an important member of regularization techniques. It is an extension of the original LASSO (least absolute shrinkage and selection operator), which has been widely used as a simple and time-efficient solution for high-dimensional regression (James et al. 2013).

The adaptive LASSO inherits the good statistical properties of the original LASSO, especially the in-built variable selection mechanism (Zou 2006). It can be important in the context of GWAS. The variable selection mechanism tends to shrink most of the estimated SNP effects into zero during the regression, and assumes that only a small proportion of SNPs have relatively large effect. As a result, the regression can serve as a selector to filter SNPs that are less likely to be associated with the traits under investigation. The original LASSO, however, has been reported as failing in the control of false positive errors (Bühlmann and van de Geer 2011). The adaptive

[1] Tianfu Yang, Zhiquan Wang, Younes Miar, Heather Bruce, Chunyan Zhang, and Graham Plastow, A Genome-wide Association Study of Meat Colour in Commercial Crossbred Pigs. Accepted by Canadian Journal of Animal Science.

LASSO can be treated as a way to correct for the "overestimate" behaviour of the original LASSO, and has been shown to be successful in controlling the false positive issue (Zou 2006).

Many regulation regression techniques have previously been successfully applied in genomic analysis, especially in GWAS. For example, ridge regression has been a powerful tool in regression for decades and it also finds a new role in detecting rare genomic mutations (Zhan and Xu 2012). In terms of genomic prediction, ridge regression is equivalent to RR-BLUP (Endelman 2011), which is one version of the widely used genomic BLUP (gBLUP). The first application of the LASSO in GWAS was no later than 2009 (Wu et al. 2009), and it also shows great performance in genomic prediction (Xu et al. 2014). Elastic net (EN) can be treated as a combination of ridge regression and the LASSO, which tries to find a balance between these two methods. An evaluation of the LASSO and EN in GWAS was conducted in a previous study, which confirmed their advantages in such scenarios (Waldmann et al. 2013). While application of LASSO (and other "family members") in similar analyses has been reported before (such as QTL mapping with experimental crosses (Sun et al. 2010)), performance of the adaptive LASSO in GWAS remained undetermined. The performance of the adaptive LASSO in GWAS remained vague as well as other members in this family, while its application in closely related cases has been reported before, such as QTL mapping with experimental crosses

The current study aimed to test the performance, mainly statistical power and precision, of the adaptive LASSO. We applied the adaptive LASSO to a public simulated dataset from QTLMAS 2010, and compared its performance with those of other methods in previous studies, such as Bayes C/CPi (Mucha et al. 2011).

2.1.2 Materials and methods

2.1.2.1 Simulated Dataset

Data used in this study came from the QTLMAS 2010 Workshop. More details for this dataset were described in the proceedings of QTLMAS 2010 Workshop (Szydlowski and Paczyńska 2011). In brief, this dataset includes genomic records of 3,226 animals in 5 generations. Each of the genomic records consists of genotype of 10,031 SNPs on 5 chromosomes. One continuous trait (QT) and one binary trait (BT) were simulated based on 30 additive QTLs, 2 QTL pairs with epistatic effect, and 3 imprinted loci. In the current study, we are mainly interested in the continuous phenotype (QT) in this dataset.

In the previous QTLMAS 2010 Workshop, seven QTL mapping methods have been evaluated for QTL analysis by different groups using the simulated data (Table 1). The criteria and results of the comparison were summarized by Mucha et al. (2011). Basically, a simulated QTL was treated as "detected" if it was within 1 Mb from a reported position. For each method in the comparison, its detection power (also called true positive rate) was calculated as the ratio of the number of detected QTL to the total number of true QTL simulated.

$$Power = \frac{number\ of\ detected\ QTL}{total\ number\ of\ simulated\ QTL}$$

The precision for the different methods was also evaluated, which is the ratio of the number of true positives to the total number of detected regions, where the true positives referred to the reported regions within 1 Mb from any simulated QTL.

$$Precision = \frac{number\ of\ true\ positives}{total\ number\ of\ detected\ regions}$$

2.1.2.2 The adaptive LASSO

As the performance of the adaptive LASSO in GWAS was not evaluated in the QTLMAS 2010 Workshop, it was evaluated here using the simulated dataset with the following multi-marker association model.

$$y = \mu + Zs + \varepsilon,$$

where μ is the intercept, Z is the design matrix associated with SNP effects (s), and ε is the residual in this model. All SNP effects were treated as fixed effects. The adaptive LASSO was conducted using package R/parcor (Krämer et al. 2009). All SNPs with non-zero effects were treated as detected regions.

2.1.3 Results and discussion

The result of the analysis was summarized in Figure 2.1 and Table 2.1. Compared with previous studies analysing the QTLMAS 2010 data using other methods (Mucha et al. 2011), the adaptive LASSO outperformed other methods in detection power, though had a lower precision.

For the 37 QTLs for the simulated dataset, the adaptive LASSO successfully detected 28 of them. For those methods reported previously (Mucha et al. 2011), the BayesC (15 QTL detected) and BayesCPi (16 QTL detected) obtained a similar detection power, while the other 5 methods mapped no more than 11 QTL. The power of the adaptive LASSO is 33 percentage points higher than the most powerful method previously reported (BayesCPi). On the other hand, the adaptive LASSO also showed a lower precision than most of the methods previously reported, which suggested that the adaptive LASSO might generate more false positive results.

There is usually a trade-off between detection power and precision in GWAS, and the optimal balance may be case specific. Historically, the high dimension of genomic data has caused more concern about the false positives (i.e. type I errors) in GWAS (Johnson et al. 2010), which might somehow lead to improved methodology in terms of different aspects (e.g. population stratification, multiple testing), in order to remove systematic false discoveries. These opinions emphasized more the importance of high precision. On the other hand, concerns also arose about possible false negatives (i.e. type II errors) in GWAS (Shi et al. 2011), such as those false negatives that result from too stringent criteria in the significance tests (Gao et al. 2009). Especially, higher detection power is very helpful for detecting genetic variants with small-to-moderate effect and/or low-to moderate frequency, which may be important for some traits such as complex diseases (Manolio et al. 2009).

Based on the results in this study, the adaptive LASSO seems to be a good option for GWAS when detection power is of importance. In terms of its precision, it may be feasible to combine the adaptive LASSO with additional screening of detected genomic region, in order to reduce false discoveries.

2.2 STUDY 2 - A genome-wide association study of meat colour in commercial crossbred pigs

The increased detection power in Study 1 implied possible application of the adaptive LASSO in scenarios where detection power has priority over other concerns (e.g. false discovery). Here we apply the adaptive LASSO to a GWAS of meat colour in pigs. Using the same dataset, three genomic regions had been reported to be associated with three meat colour traits using Bayesian

regression (Zhang et al. 2015)[2]. One aim of the study was to explore the possibility of detecting more candidate genomic regions through the application of the adaptive LASSO, while we also aimed to investigate if any common genetic basis are underlying multiple meat colour traits in pigs.

2.2.1 Introduction

Meat colour is the most important guide in buying pork, as it has a key effect on the visual attractiveness of pork to consumers (Lu et al. 2000; Mancini and Hunt 2005). Meanwhile, meat colour measurements can serve as an easy-to-measure indicator of other meat quality traits (e.g. drip loss), as they are usually correlated (Warriss et al. 2006; Li et al. 2012; Miar et al. 2014). Improving meat colour would be economically important for the pork industry.

For pigs, meat colour can be affected by a set of environmental factors, with approximately 10-75% explained by genetics depending on the trait, breed, and population (Sonesson et al. 1998; Miar et al. 2014). It has been reported that pork colour can be adjusted through diet supplementation of vitamin D3 (Wilborn et al. 2004), altering available space (Gentry et al. 2004), or exercise (Rosenvold et al. 2002). Well-known examples of genetic factors include the genes *PRKAG3* and *RYR1* (or *CRC1*), both of which interact with environmental factors to a degree. The gene *PRKAG3* (protein kinase AMP-activated non-catalytic subunit gamma 3) was first identified as the RN gene (Milan 2000; Ciobanu et al. 2001; Škrlep et al. 2009). This gene has been reported to affect the lightness, redness and yellowness of pork (Lindahl et al. 2004), while its effect on

[2] I also contributed to the statistical analysis in that study and was a co-author of the publication: Zhang, C., Wang, Z., Bruce, H., Kemp, R.A., Charagu, P., Miar, Y., Yang, T., and Plastow, G. 2015. Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs. BMC Genet. 16: 33.

pork quality may vary according to different meat processing ingredients, such as phosphate (de Vries et al. 2000). The gene *RYR1* (ryanodine receptor 1, known widely as the "Halothane gene") gives another example of gene by environment interaction, where the colour of pork was reported to be affected jointly by *RYR1* and pre-harvest handling and stunning (Fujii et al. 1991; Chang et al. 2003; Guàrdia et al. 2009).

In pig breeding, discovering new genomic regions associated with meat colour can be very useful for its genetic improvement as measurement occurs post mortem. Candidate regions provide important information for identifying causative mutations, which may ultimately lead to a more accurate description of the genetic basis underlying the traits. Even if the causative mutations remain unknown, the associated SNPs in the candidate regions are useful for marker-assisted selection (Goddard and Hayes 2009). This strategy has been successfully applied in the selection of various livestock species. Genetic markers have been used in the genetic improvement of different meat quality traits, reproductive traits and animal health (van der Steen et al. 2005; Otto et al. 2007).

The meat colour dataset used in the current study came from a population of crossbred commercial pigs. Two studies, a genetic parameters estimation (Miar et al. 2014) and a GWAS using BayesB (Zhang et al. 2015), have been published for the meat colour traits in the same population. In the present study, we report the results of a GWAS of meat colour using the adaptive LASSO with the same dataset used by Zhang et al. (2015) plus additional 34 animals. One objective of the current study is to detect genomic regions associated with meat colour in addition to those reported by Zhang et al (2015) through the adaptive LASSO, as the method has shown a relatively high detection power in Section 2.1. Meanwhile, medium to high genetic correlations had been observed between meat colour traits in the population (Miar et al. 2014), which implied

a common genetic basis underlying those traits. In order to investigate this possibility, the current study aimed to detect genomic regions that were associated with multiple meat colour traits.

2.2.2 Materials and methods

2.2.2.1 Meat colour data

A total of 1,977 crossbred commercial pigs from two breeding companies (Hypor Inc., Regina, SK, Canada, and Genesus Inc., Oakville, MB, Canada) were used for this study. These commercial pigs were offspring of Duroc boars and F1 hybrid sows (Landrace × Large White). The feeding and raising environment for these animals were described previously (Miar et al. 2014; Zhang et al. 2015). The animals used in this study were raised as part of commercial pork production. The proposed work was approved by the University of Alberta Animal Care and Use Committee.

Minolta colours L*, a* and b* were used to measure the meat colour. The fresh colours were measured at 24 h after exsanguination on the longissimus dorsi muscle (FMCOL) and subcutaneous fat above the longissimus dorsi muscle (FCOL) on the fresh loin, and on the muscle of gluteus medius (GMCOL), quadriceps femoris (QFCOL) and iliopsoas (ICOL) on the fresh ham surface. All the packaged loins in the same batch were simultaneously frozen (−20°C) within 24 hours after exsanguination and maintained frozen. Prior to thawed muscle measurement, the pork loins were thawed for 72 hours at 4 °C. Thawed muscle colours (TMCOL) were measured on the thawed loin muscle. In total, 18 meat colour measurements were collected and used for this study (Table 2.2). All colour measurements were repeated from three different sites at the same position and the average value was made for the final analysis. The colour assessment was made using a Konica-Minolta Chroma-meter CR-400 (Konica Minolta Sensing Inc., Japan).

The DNA samples of the 1,977 animals were genotyped using Illumina PorcineSNP60 V2 Genotyping Beadchip (Illumina, Inc., San Diego, CA, USA). SNPs were filtered out if they met any one of the following criteria: 1) genotyping call rate < 95%, 2) Chi-square value of Hardy-Weinberg equilibrium test > 600, 3) minor allele frequency < 5%, and 4) no valid genomic position information available. After quality control, 42,721 SNPs remained in the datasets for subsequent analysis. And all missing values in the genotypes were imputed using FImpute V2.2 (Sargolzaei et al. 2014).

2.2.2.2 Population structure

We tested the population structure of the 1,977 animals in the combined population. Genetic difference among animals was measured as pairwise identity-by-state Hamming distances in Plink 1.07 (Purcell et al. 2007), and a multidimensional scaling (MDS) analysis was conducted to show the top two dimensions of the population structure. A principal component analysis (PCA) was also performed by eigenvalue decomposition of a marker-based relationship matrix. The PCA was conducted using R/rrBLUP (Endelman 2011).

2.2.2.3 Multiple linear regression

All phenotypes were pre-adjusted before association analysis as Eq. 2.1,

$$y^* = y - (s + cg + r + p), \qquad \text{(Equation 2.1)}$$

where y^* is the adjusted phenotype, y is the raw phenotype, s is the effect of sex, cg is the effect of contemporary group (combination of slaughter batch, year and group during the test), r is the

effect of room, and p is the effect of pen. A multiple linear model was then used for association analysis. The model used is described in Eq. 2.2,

$$y^* = c + Zs + \varepsilon, \qquad \text{(Equation 2.2)}$$

where c is the effect of the source of animals (Hypor Inc. or Genesus Inc.), Z is the design matrix associated with SNP effects (s), and ε is the residual in this model. All SNP effects were treated as fixed effects in the model.

The linear model (Eq. 2.2) was then implemented using the adaptive LASSO. As the results in Study 1 suggested, the adaptive LASSO might have a higher detection power in GWAS than other methods. It may therefore be a good choice as the study aimed to extend the number of potential candidate genomic regions associated with meat colour. Here we apply the same strategy as Study 1. The analysis was conducted using R/parcor (Krämer et al. 2009).

2.2.2.4 Candidate genomic regions detection

The estimated SNP effects were used to infer candidate genomic regions. We divided the genome into windows with a width of 2 mega base pairs (Mbp). Any windows where one or more SNPs have non-zero effects were considered as candidate windows. In the case that association was found in two adjacent genomic windows, they were combined as one candidate window, as the detected association may essentially reflect the effect of the same gene or causative mutation. Among the detected genomic regions, we were more interested in the windows that associated with more than one meat colour trait. A previous study revealed that the meat colour traits have relatively high genetic correlations with each other in the same population (Miar et al. 2014). In

this case, genomic regions associated with multiple colours may provide evidence for a common genetic basis underlying these traits.

2.2.2.5 Comparison with known QTL

For those genomic regions associated with more than one meat colour trait, we searched pigQTLdb (Hu et al. 2013) for known QTL that may support the association detected in the analysis. Those QTL that met these three criteria were selected: 1) overlapping with at least one of the candidate regions; 2) spanning no more than 30 Mbp (as those QTLs spanning a larger region provide limited support); and 3) reported for functionally related traits, especially meat colour traits.

2.2.2.6 Significance tests

For the SNPs in those genomic regions associated with more than one meat colour trait, single marker association analysis was conducted to test the significance of the association. The analysis was conducted with R/rrBLUP (Endelman 2011). In the analysis, a mixed linear model was used for the testing, as Eq. 2.3.

$$y^* = c + Z_i s_i + g + \varepsilon \qquad \text{(Equation 2.3)}$$

For the ith SNP, its effect was modelled as a fixed effect (s_i) and Z_i is its design matrix. A polygenic term (g) was also included to model the rest of genetic effect, and was treated as a random effect. Genomic kinship matrix was calculated based on the genotype data, and then used in solving the model. The calculated P-values were adjusted using false discovery rate (FDR) for every trait. The significance of the associations was determined based on a commonly used

threshold of FDR < 0.1 (Li and Ma 2013), i.e. all SNPs with FDR lower than 0.1 were treated as significant.

2.2.3 Results and discussion

2.2.3.1 Population structure

Figure 1 shows the top two dimensions of the population structure. Circles and Triangles represent the animals that came from the two different breeding companies (Hypor or Genesus Inc.). The figure shows that the difference in source had obvious contribution to the variation along dimension 1, and this factor was already considered in the model described above (Eq. 2 and Eq. 3). In the PCA, the three highest eigenvalues were 5.18%, 1.53% and 0.97% of the summation of all eigenvalues. The latter two were lower than 2%, which implies low population stratification within each source so that no additional factors were included in the model.

2.2.3.2 Association results

In the 1,291 genomic windows constructed in the analysis, we detected 132 genomic windows that are associated with at least one of the meat colour traits, and 23 of them showed association with more than one meat colour trait. After merging adjacent windows, we found 20 candidate genomic regions in total (Table 2 and 3). Eight of them were associated with at least three traits. Two candidate regions were associated with one of the fat colour measurements (FCOL b*). They were located on SSC1 (268-270 MP) and SSC7 (38-40 Mbp), respectively. The 20 candidate genomic regions harboured 1,034 SNPs in the current dataset. Significant associations (FDR < 0.1) were found in the single marker association for all the 20 candidate regions.

2.2.3.3 Comparison with previous results

The results were compared with a previous report of the genetic parameters for the same combined population (Miar et al. 2014). Candidate regions were found associated with meat colour traits that were genetically correlated, even when the phenotypic correlations were relatively low (Table 4). For example, the estimated genetic correlation between GMCOL L* and GMCOL a* was -0.42, whilst their phenotypic correlation was 0.04. We found two genomic regions (SSC2:154-156 Mbp and SSC15:130-134 Mbp) that had significant SNPs for both traits. The detected associations were more likely to be due to a common genetic basis underlying the two traits, rather than being caused by phenotypic correlation due to environmental factors, as the phenotypic correlation is very low. These results supported our hypothesis that there is a common genetic basis underlying these traits.

We also compared our GWAS results with that reported by Zhang et al. (2015), where six genomic regions were found associated with three meat colour traits. Three of them were also detected in the current analysis. The major QTL detected in that study, the candidate region on SSC15 (harbouring *PRKAG3*), had large effects on three meat colour traits (QFCOL a*, QFCOL b* and TMCOL b). In the current study, we also detected the same association, while there were also significant associations (FDR < 0.1) between that genomic region on SSC15 and 7 other meat colour traits. The candidate region located around 35 Mbp on SSC1 was associated with three meat colour traits in the current analysis. The candidate region located around 1 Mbp on SSCX was associated with only one meat colour trait here and was not selected for further analysis and not included in Table 2 and 3. The other three candidate regions were not detected in the current analysis. On the other hand, 17 new genomic regions showed significance in the single-marker test.

The difference in the results may reflect the different behaviour and performance of the two regression methods used in the analysis (the adaptive LASSO and BayesB), as the two studies used almost the same dataset (All 1943 animals used by Zhang et al. were included in the current study, with records from 34 additional animals added since then) and the same multiple linear model. Similar results were observed in a previous assessment of the adaptive LASSO and its comparison with BayesC/Cpi using a public simulation dataset (Section 2.1). In that case, the adaptive LASSO detected more QTL but suffered from lower precision (more false positives). The candidate regions detected by the adaptive LASSO is overlapping but was not able to detect all of the regions detected by BayesC/Cpi tested in the same dataset.

The different behaviours of the two methods are likely to result from the joint effect of a set of underlying and confounding factors during the analysis, such as different basic assumptions and/or implementations, with some of the factors not comparable in the current analysis. Here we only propose one noticeable possibility. From a Bayesian point of view, one possible reason for the difference in results is the different prior assumptions used in the methods. The LASSO family gives equivalent estimation of coefficients in linear regression as Bayesian regression but with different prior distributions. The adaptive LASSO can be interpreted as Bayesian regression with an independent double exponential distribution as the prior distribution (Leng et al. 2014). Meanwhile, for the method used in the previous analysis by Zhang et al, BayesB, the prior assumption is an independent mixture distribution with a point mass at zero with a given probability Pi and a univariate-t distribution with probability 1-Pi. As the prior distributions reflects the prior assumptions about the regression coefficients (i.e. marker effects), how close the prior distributions are to the true distribution may have a large impact on the result of estimation.

This impact may vary between SNPs, which may then result in overlapping but distinct candidate regions being detected by the two methods and the two studies.

2.2.3.4 Connection with known QTL

Seventeen out of the 20 candidate regions overlapped with at least one of known QTL for meat colour or related meat quality traits. We found 110 QTL for pork quality from pigQTLdb that overlap with our candidate regions. These QTL had been reported for traits including meat colour, related pathways (e.g. glycolytic potential, fibre type proportion, pH), and other pork quality measurements (e.g. shear force, drip loss). In addition, the two candidate regions associated with fat colour overlapped with QTL for fatness and fat composition traits. Twelve QTL for fatness overlapped with the candidate region on SSC1 (268-270 Mbp), and 35 QTL for fatness and fat composition overlapped with the candidate region on SSC7 (38-40 Mbp). Some of the QTL span no more than 5 Mbp or even 1 Mbp, which may provide support to our detected association to a degree.

2.2.3.5 Potential candidate genes

We also searched for genes that are potentially involved in pathways related to meat colour in these candidate regions. Genome annotation information was obtained from UCSC Genome Browser (Kent et al. 2002) and RefSeq (Pruitt et al. 2007). The distances between the genes and those significant SNPs (FDR < 0.1) were taken into consideration, and the maximum distance was set as 1 Mbp for the search. The candidate region on SSC7 (26 – 28 Mbp) overlaps with the pig major histocompatibility complex (SLA) Class I region, and harboured 56 genes. Other than that region, most candidate regions harboured no more than 11 genes in the search. For 11 of the 20

candidate regions, we found genes that may be functionally related to meat colour. In this section, we use the standard notation of italics for gene names and plain text for the corresponding proteins.

Muscle colour is related to biological factors such as the redox forms of myoglobin, the morphology of muscle structure and the distribution of myoglobin, which may be further affected by a set of genetic and environmental factors (Seideman et al. 1984). Among these factors, muscle colour is primarily determined by the concentration and chemical forms of myoglobin (Seideman et al. 1984; Mancini and Hunt 2005), which has four major chemical forms that results in different colours. In total, we found 12 genes within the 20 candidate regions, we found 12 genes functionally related to 1) muscle fibre type composition, and 2) glucose metabolism. For the two candidate regions associated with fat colour, two genes were found to be related to lipid metabolism, and may thereby contribute to the phenotypic variation in fat colour. These genes may potentially explain the associations we found between these genomic regions and the meat colour traits.

2.2.3.5.1 Muscle fibre type composition

It has been reported that porcine muscle fibre type composition is functionally related to muscle colour. Through histochemical and biochemical methods, muscle fibre can be grouped into different types. A commonly used fibre typing system classifies muscle fibres into four types, type I, IIA, IIX, and IIB (Klont et al. 1998). As the concentration and chemical forms of myoglobin differ between types of muscle fibre, they tend to be associated with different colours. For example, type I muscle fibres have a higher concentration of myoglobin than type IIB muscle fibres, and therefore muscle with more type I muscle fibre tends to be redder (Seideman et al. 1984). This relationship has been supported by previous studies. For example, it has been reported that the

amount of type I or IIA muscle fibres is negatively correlated with muscle yellowness in pigs (Eggert et al. 2002). The abundance of type IIA and IIX muscle fibres in psoas has also been reported to be correlated to high pork quality, while the proportion of type IIB muscle fibres in longissimus dorsi was associated with low quality (Chang et al. 2003). The effect of muscle fibre type on muscle colour can also be partially explained by the difference in glucose metabolism, which will be discussed separately. Ten genes related to muscle fibre type composition have been found in our candidate regions. They can be further grouped into three subgroups: 1) genes involved in the regulation of MYOD1 (myogenic differentiation 1) in myogenesis; 2) genes involved in the GH-IGF (growth hormone 1 and insulin-like growth factor) system; and 3) other genes reported to be associated with muscle fibre type composition.

Myoblast determination protein 1 has been reported to be a critical regulatory factor for myogenesis. Especially, MYOD1 has been found to play a key role in determining muscle fibre type during muscle differentiation. In rodents, both the gene expression level and protein level of MYOD1 showed differences between type I and type II muscle fibres, and MYOD1 affected the proportions of muscle fibre types (Hughes et al. 1997; Hennebry et al. 2008). It has been reported that two polymorphisms in *MYOD1* were associated with muscle fibre composition (Lee et al. 2012) in pigs. In the present study, we found *MYOD1* in one of our candidate regions on SSC2. In addition, the candidate regions also harbour two other genes that may be involved in this process. One of them encodes myocardin (MYOCD), which has been reported to repress the function of MYOD1 during skeletal muscle development (Long et al. 2007). The other one encodes tumour necrosis factor (TNF), which may suppress the expression of *MYOD1* gene and destabilize MYOD1 in the process of skeletal muscle differentiation (Guttridge 2000; Langen et al. 2004).

The GH-IGF system also plays a regulatory role in myogenesis. Growth hormone 1 (GH) has been reported to be associated with muscle fibre type composition. In hypophysectomized rats, growth hormone treatment affected the gene expression level of different forms of myosin heavy chain (Loughna and Bates 1994), and increased the proportion of type I muscle fibres (Ayling et al. 1989). Similar association was also found in transgenic mouse lines (Schuenke et al. 2008). Growth hormone also indirectly affects myogenesis especially through induction of hepatic insulin-like growth factor (IGF) gene expression (Florini et al. 1996). The IGFs (IGF-1 and IGF-2) and their binding proteins (IGFBPs) are critical in myogenesis and associated with pork quality (Oksbjerg et al. 2004). For example, the IGF-2 gene has been reported as a candidate gene of the percentage of type I and type IIB muscle fibre (Estelle et al. 2008) and meat colour (Rohrer et al. 2012). Within our identified candidate regions, we found the genes encoding GH and two IGFBPs (IGFBP1/3). Both of the IGFBPs serve as carrier proteins for the IGFs in the circulation, prolong the half-life of the IGFs, and have effects on IGF-stimulated actions (Florini et al. 1996). Notably polymorphisms in *IGFBP3* have been reported to be associated with pork colour in previous studies (Wang et al. 2009).

The candidate genomic regions also harbour three other genes that have been reported to be associated with muscle fibre type composition. They encode actin (alpha 1) in skeletal muscle (ACTA1), TEA domain transcription factor 1 (TEAD1), and nuclear receptor subfamily 4 group A member 3 (NR4A3), respectively. ACTA1 is critical for the basic functions of skeletal muscle through actin-myosin interaction. It has been found that mutations in *ACTA1* were associated with congenital fibre type disproportion in human. Individuals suffering from this disorder have irregular composition of muscle fibre types (e.g. relatively low proportion of type I muscle fibre) due to selective atrophy (Laing et al. 2004; Clarke et al. 2007). TEAD1 is a key regulator of the

gene expression in slow muscle. It has been reported to be directly involved in the regulation in slow-to-fast muscle fibre type conversion (Zhang et al. 2014). In addition, over-expression of *TEAD1* in mouse skeletal muscle induces fast-to-slow fibre type conversion (Tsika et al. 2008). These observations provide support for the association between *TEAD1* and muscle fibre type composition. NR4A3 is a transcription factor and may be involved in various biological processes in skeletal muscle (Pearen et al. 2006). It has been reported that the activation of *NR4A3* in mice was associated with muscle fibre remodelling from type IIB muscle fibre to type IIX and type IIA muscle fibre (Pearen et al. 2012). *NR4A3* may also play other roles in the determination of muscle colour and fat colour, which will be discussed in the following sections.

2.2.3.5.2 Glucose metabolism

Another important biological pathway related to muscle colour is glucose metabolism. It has been reported that the glycolytic potential of skeletal muscle, which is a measurement of resting glycogen level, is highly associated with post-mortem pH and pork colour (Mancini and Hunt 2005). The association can be explained by the post-mortem changes in lactate level and pH in muscle (Choe et al. 2008). Extremely low pH may trigger a set of changes in muscle tissue (e.g. protein denaturation, myoglobin oxidation, and the increasing of light scattering properties) (Klont et al. 1998; Mancini and Hunt 2005), and consequently result in a paler pork colour and a lower water holding capacity. Conversely, low glycolytic potential may ultimately contribute to a darker colour of pork (Seideman et al. 1984). This process has been reported to underlie the association between pork colour and known candidate genes. For example, mutations in *PRKAG3* may affect the muscle glycogen storage in pigs, and thereby lead to changes in pork colour (Milan 2000; Ciobanu et al. 2001).

In addition to *PRKAG3* that had been found in one of our candidate regions on SSC15, we also found three genes related to glucose metabolism in detected candidate regions. These genes encode glucagon-like peptide 1 receptor (GLP1R), reticulon 4 (RTN4), and NR4A3, respectively. Glucagon-like peptide 1 receptor is a protein involved in the signalling pathways related to insulin secretion. It has been reported that the accumulation of muscle glycogen significantly increased in *GLP1R* knockout mice when compared with wild-type mice (Ayala et al. 2009), which suggested the role of GLP1R in the regulation of muscle glycogen storage. GLP1R is also involved in the regulation of plasma glucose level (Doyle and Egan 2007), and therefore functionally related to glycolytic potential and muscle colour. Reticulon 4 plays a role in the development and metabolism of muscle (Chen et al. 2010), possibly through the regulation of cytoplasmic Ca^{2+} level (Jozsef et al. 2014) and glycolysis (Sutendra et al. 2011). *RTN4* has been reported to express differentially in prenatal skeletal muscle between Tongcheng and Landrace pigs, possibly due to its role in skeletal muscle contraction (Tang et al. 2007). In previous proteomic study, RTN4 was found to be associated with ultimate pH in pork (te Pas et al. 2013). Interestingly, it has been reported that the level of *RTN4* was correlated with the muscle fibre types in pigs under certain circumstances (Yang et al. 2014), which implies another connection between *RTN4* and pork colour traits. The aforementioned gene *NR4A3* may also affect the glycolytic potential in skeletal muscle. In a transgenic mouse model, evidence was found that NR4A3 enhanced glucose transport and utilization in skeletal muscle (Close et al. 2013), and the activation of *NR4A3* resulted in an increased glycogen content in skeletal muscle (Pearen et al. 2013). The potential connection between *NR4A3* and meat colour was also supported by another previous study where silencing *NR4A3* leaded to the accumulation of lactate (Pearen et al. 2008).

2.2.3.5.3 Lipid Metabolism

Fatty acid metabolism and fat deposition process are functionally connected to the colour of fat. The correlation between the colour and fatty acid composition in fat tissue has been reported in different studies (Wood et al. 2004; Gjerlaug-Enger et al. 2011). The unsaturation of fatty acids is associated with their melting point, therefore affecting the firmness of fat tissue, and may eventually lead to changes in the colour (Wood et al. 2004). Additionally, the different ability of fatty acids to oxidise may also play a role in the determination of fat colour, as peroxidation of body fat results in the accumulation of lipofuscin, a yellow-brown pigment (Ruiter et al. 1978). The colour of fat can also be affected by the fat deposition process. Reduced fat deposition may lead to higher concentration of connective tissues and pigments (e.g. lipofuscin and carotenoid) in the fat tissue, which may ultimately lead to grey or yellow colour, respectively (Wood 1984; Dunne et al. 2009).

The association between our candidate regions and the fat colour can be explained by these mechanisms, as the two regions associated with fat colour overlapped with many QTL for fatness (e.g. average backfat thickness and fat percentage in carcass) and fatty acids composition. The aforementioned genes *GLP1R* and *NR4A3* were located in these two candidate regions, and both of them were related to lipid metabolism. *GLP1R* was previously reported as a candidate gene for backfat thickness in pigs (Huang et al. 2011), which may imply its role in lipid deposition. Evidence from mice also support the possible relationship between GLP1R signaling and fat mass (Nogueiras et al. 2009). It is possible that *GLP1R* also affected fatty acid composition, as it was reported to be involved in the metabolism of fatty acids (e.g. synthesis, oxidation and composition) (Doyle and Egan 2007; Svegliati-Baroni et al. 2011; Kawaguchi et al. 2014). *NR4A3* has also been reported to be involved in lipid metabolism. The expression of *NR4A3* was associated with

intramuscular fatty acid composition in previous transcriptomic study in pigs (Ramayo-Caldas et al. 2012), and altered expression of *NR4A3* was also observed in human obesity (Veum et al. 2012).

2.2.4 Conclusions

Twenty genomic regions were found to be associated with multiple meat colour traits in the association analysis using the adaptive LASSO. All the 20 candidate regions were confirmed in single marker association, where significant SNPs (FDR < 0.1) were found in all of the candidate regions. The results were compared with previous analysis conducted using the same population. Three of the six genomic regions that were reported for the population were confirmed, while 17 new candidate regions showed significant association in the current analysis. For the 20 regions detected in the current analysis, 17 of them overlapped with pork quality QTL that had been previously discovered, including QTL of meat colour and functionally related pathways. Within the candidate regions, 12 genes were found close to at least one of the significant SNPs (less than 1 Mbp), and functionally related to meat colour. These genes may affect muscle fibre type composition and/or glucose metabolism, and ultimately explain part of the phenotypic variation observed in meat colour. The results can be used in discovering new genetic causative mutations for meat colour traits, and will provide new information for pig breeding. Being consistent to previous report, the results also showed additional evidence that adaptive LASSO may help to detect extra candidate genomic regions and achieve higher detection power compared with other approaches (e.g. BayesB).

2.3 Overall conclusions

Avoiding both false positive and false negative results has been an important topic in GWAS, with the dimension of genomic data increases dramatically due to the emergence of high-throughput biotechnologies. The two studies in this chapter investigated if adaptive LASSO may help to achieve a higher detection power in GWAS. In study 1, we showed that the adaptive LASSO outperformed several other methods, including BayesC/CPi, in detection power, while it suffered from a low precision. In study 2, we applied the adaptive LASSO in a GWAS of meat colour in pigs. Twenty genomic regions showed association with multiple meat colour traits. These regions were supported by evidence from single-marker significance tests, previous association studies of meat quality, known QTL for meat colour in pigs, and genes functionally related to meat colour. Being consistent to Study 1, the results showed additional evidence that adaptive LASSO may help to detect extra candidate genomic regions and achieve higher detection power compared with other approaches.

2.4 References

Ayala, J.E., Bracy, D.P., James, F.D., Julien, B.M., Wasserman, D.H., and Drucker, D.J. 2009. The glucagon-like peptide-1 receptor regulates endogenous glucose production and muscle glucose uptake independent of its incretin action. Endocrinology **150**: 1155–1164.

Ayling, C.M., Moreland, B.H., Zanelli, J.M., and Schulster, D. 1989. Human growth hormone treatment of hypophysectomized rats increases the proportion of type-1 fibres in skeletal muscle. J. Endocrinol. **123**: 429–35.

Berk, R.A. 2008. Statistical learning from a regression perspective. Springer New York, New York, NY.

Bühlmann, P., and van de Geer, S. 2011. Statistics for High-Dimensional Data. Springer Berlin Heidelberg, Berlin, Heidelberg.

Chang, K.C., da Costa, N., Blackley, R., Southwood, O., Evans, G., Plastow, G., Wood, J.D., and Richardson, R.I. 2003. Relationships of myosin heavy chain fibre types to meat quality traits in traditional and modern pigs. Meat Sci. **64**: 93–103.

Chen, Y.-C., Wu, B.-K., Chu, C.-Y., Cheng, C.-H., Han, H.-W., Chen, G.-D., Lee, M.-T., Hwang, P.-P., Kawakami, K., Chang, C.-C., and Huang, C.-J. 2010. Identification and characterization of alternative promoters of zebrafish Rtn-4/Nogo genes in cultured cells and zebrafish embryos. Nucleic Acids Res. **38**: 4635–4650.

Choe, J.H., Choi, Y.M., Lee, S.H., Shin, H.G., Ryu, Y.C., Hong, K.C., and Kim, B.C. 2008. The relation between glycogen, lactate content and muscle fiber type composition, and their influence on postmortem glycolytic rate and pork quality. Meat Sci. **80**: 355–362.

Ciobanu, D., Bastiaansen, J., Malek, M., Helm, J., Woollard, J., Plastow, G., and Rothschild, M. 2001. Evidence for new alleles in the protein kinase adenosine monophosphate-activated gamma(3)-subunit gene associated with low glycogen content in pig skeletal muscle and improved meat quality. Genetics **159**: 1151–62.

Clarke, N.F., Ilkovski, B., Cooper, S., Valova, V.A., Robinson, P.J., Nonaka, I., Feng, J.-J., Marston, S., and North, K. 2007. The pathogenesis of ACTA1-related congenital fiber type disproportion. Ann. Neurol. **61**: 552–561.

Close, A.F., Rouillard, C., and Buteau, J. 2013. NR4A orphan nuclear receptors in glucose homeostasis: A minireview. Diabetes Metab. **39**: 478–484.

Doyle, M.E., and Egan, J.M. 2007. Mechanisms of action of glucagon-like peptide 1 in the pancreas. Pharmacol. Ther. **113**: 546–593.

Dunne, P.G., Monahan, F.J., O'Mara, F.P., and Moloney, A.P. 2009. Colour of bovine subcutaneous adipose tissue: A review of contributory factors, associations with carcass and meat quality and its potential utility in authentication of dietary history. Meat Sci. **81**: 28–45.

Eggert, J.M., Depreux, F.F.S., Schinckel, A.P., Grant, A.L., and Gerrard, D.E. 2002. Myosin heavy chain isoforms account for variation in pork quality. Meat Sci. **61**: 117–126.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome J. **4**: 250.

Estelle, J., Gil, F., Vazquez, J.M., Latorre, R., Ramirez, G., Barragan, M.C., Folch, J.M., Noguera, J.L., Toro, M.A., and Perez-Enciso, M. 2008. A quantitative trait locus genome scan for porcine muscle fiber traits reveals overdominance and epistasis. J. Anim. Sci. **86**: 3290–3299.

Florini, J.R., Ewton, D.Z., and Coolican, S.A. 1996. Growth hormone and the insulin-like growth factor system in myogenesis. Endocr. Rev. **17**: 481–517.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V., Weiler, J., O'Brien, P., and MacLennan, D. 1991. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science **253**: 448–451.

Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D., and Province, M.A. 2009. Avoiding the high Bonferroni penalty in genome-wide association studies. Genet. Epidemiol. **34**.

Gentry, J.G., McGlone, J.J., Miller, M.F., and Blanton, J.R. 2004. Environmental effects on pig performance, meat quality, and muscle characteristics. J. Anim. Sci. **82**: 209.

Gjerlaug-Enger, E., Kongsro, J., Aass, L., Odegård, J., and Vangen, O. 2011. Prediction of fat quality in pig carcasses by near-infrared spectroscopy. Animal **5**: 1829–41.

Goddard, M.E., and Hayes, B.J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. **10**: 381–391.

Guàrdia, M.D., Estany, J., Balasch, S., Oliver, M.A., Gispert, M., and Diestre, A. 2009. Risk assessment of skin damage due to pre-slaughter conditions and RYR1 gene in pigs. Meat Sci. **81**: 745–751.

Guttridge, D.C. 2000. NF-kappa B-Induced loss of MyoD messenger RNA: possible role in muscle decay and cachexia. Science **289**: 2363–2366.

Hennebry, A., Berry, C., Siriett, V., O'Callaghan, P., Chau, L., Watson, T., Sharma, M., and Kambadur, R. 2008. Myostatin regulates fiber-type composition of skeletal muscle by regulating MEF2 and MyoD gene expression. AJP Cell Physiol. **296**: C525–C534.

Hu, Z.-L., Park, C.A., Wu, X.-L., and Reecy, J.M. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. **41**: D871–D879.

Huang, W.H., Ma, Z.X., Xu, Z.Y., Xiong, Y.Z., and Zuo, B. 2011. Detection of novel SNPs and mapping of the fatness QTL on pig chromosome 7q1.1-1.4 region. Genet. Mol. Res. **10**: 3090–3097.

Hughes, S.M., Koishi, K., Rudnicki, M., and Maggs, A.M. 1997. MyoD protein is differentially accumulated in fast and slow skeletal muscle fibres and required for normal fibre type balance in rodents. Mech. Dev. **61**: 151–163.

James, G., Witten, D., Hastie, T., and Tibshirani, R. 2013. An Introduction to Statistical Learning. Springer New York, New York, NY.

Jiang, X., and Neapolitan, R.E. 2012. Mining pure, strict epistatic interactions from high-dimensional datasets: ameliorating the curse of dimensionality. PLoS One **7**: e46771.

Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D., Winkler, C.A., and O'Brien, S.J. 2010. Accounting for multiple comparisons in a genome-wide association study

(GWAS). BMC Genomics **11**: 724.

Jozsef, L., Tashiro, K., Kuo, A., Park, E.J., Skoura, A., Albinsson, S., Rivera-Molina, F., Harrison, K.D., Iwakiri, Y., Toomre, D., and Sessa, W.C. 2014. Reticulon 4 is necessary for endoplasmic reticulum tubulation, STIM1-Orai1 coupling, and store-operated calcium entry. J. Biol. Chem. **289**: 9380–95.

Kawaguchi, T., Itou, M., Taniguchi, E., and Sata, M. 2014. Exendin‐4, a glucagon‐like peptide‐1 receptor agonist, modulates hepatic fatty acid composition and Δ‐5‐desaturase index in a murine model of non‐alcoholic steatohepatitis. Int. J. Mol. Med. **34**: 782–787.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, a. D. 2002. The Human Genome Browser at UCSC. Genome Res. 12: 996–1006.

Klont, R.E., Brocks, L., and Eikelenboom, G. 1998. Muscle fibre type and meat quality. Meat Sci. **49**: S219–S229.

Krämer, N., Schäfer, J., and Boulesteix, A.-L. 2009. Regularized estimation of large-scale gene association networks using graphical Gaussian models. BMC Bioinformatics **10**: 384.

Laing, N.G., Clarke, N.F., Dye, D.E., Liyanage, K., Walker, K.R., Kobayashi, Y., Shimakawa, S., Hagiwara, T., Ouvrier, R., Sparrow, J.C., Nishino, I., North, K.N., and Nonaka, I. 2004. Actin mutations are one cause of congenital fibre type disproportion. Ann. Neurol. **56**: 689–694.

Langen, R.C.J., Van Der Velden, J.L.J., Schols, A.M.W.J., Kelders, M.C.J.M., Wouters, E.F.M., and Janssen-Heininger, Y.M.W. 2004. Tumor necrosis factor-alpha inhibits myogenic differentiation through MyoD protein destabilization. FASEB J. **18**: 227–37.

Lee, E.A., Kim, J.M., Lim, K.S., Ryu, Y.C., Jeon, W.M., and Hong, K.C. 2012. Effects of variation in porcine MYOD1 gene on muscle fiber characteristics, lean meat production, and meat quality traits. Meat Sci. **92**: 36–43.

Leng, C., Tran, M.N., and Nott, D. 2014. Bayesian adaptive Lasso. Ann. Inst. Stat. Math. **66**: 221–244.

Li, C., Liu, D., Zhou, G., Xu, X., Qi, J., Shi, P., and Xia, T. 2012. Meat quality and cooking attributes of thawed pork with different low field NMR T21. Meat Sci. **92**: 79–83.

Li, J., and Ma, S. 2013. Survival Analysis in Medicine and Genetics. CRC Press, Boca Raton, FL.

Lindahl, G., Enfält, A.-C., Seth, G. von, Joseli, Å., Hedebro-Velander, I., Andersen, H.J., Braunschweig, M., Andersson, L., and Lundström, K. 2004. A second mutant allele (V199I) at the PRKAG3 (RN) locus—II. Effect on colour characteristics of pork loin. Meat Sci. **66**: 621–627.

Long, X., Creemers, E.E., Wang, D.-Z., Olson, E.N., and Miano, J.M. 2007. Myocardin is a bifunctional switch for smooth versus skeletal muscle differentiation. Proc. Natl. Acad. Sci. **104**: 16570–16575.

Loughna, P.T., and Bates, P.C. 1994. Interactions between growth hormone and nutrition in hypophysectomized rats: skeletal muscle myosin heavy chain mRNA levels. Biochem. Biophys. Res. Commun. **198**: 97–102.

Lu, J., Tan, J., Shatadal, P., and Gerrard, D.E. 2000. Evaluation of pork color by using computer vision. Meat Sci. **56**: 57–60.

Mancini, R.A., and Hunt, M.C. 2005. Current research in meat color. Meat Sci. **71**: 100–121.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., and Visscher, P.M. 2009. Finding the missing heritability of complex diseases. Nature **461**:

747–53.

Miar, Y., Plastow, G.S., Moore, S.S., Manafiazar, G., Charagu, P., Kemp, R.A., Van Haandel, B., Huisman, A.E., Zhang, C.Y., McKay, R.M., Bruce, H.L., and Wang, Z. 2014. Genetic and phenotypic parameters for carcass and meat quality traits in commercial crossbred pigs. J. Anim. Sci. **92**: 2869–2884.

Milan, D. 2000. A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science **288**: 1248–1251.

Mucha, S., Pszczoła, M., Strabel, T., Wolc, A., Paczyńska, P., and Szydlowski, M. 2011. Comparison of analyses of the QTLMAS XIV common dataset. II: QTL analysis. BMC Proc. **5**: S2.

Nogueiras, R., Perez-Tilve, D., Veyrat-Durebex, C., Morgan, D. a, Varela, L., Haynes, W.G., Patterson, J.T., Disse, E., Pfluger, P.T., Lopez, M., Woods, S.C., DiMarchi, R., Dieguez, C., Rahmouni, K., Rohner-Jeanrenaud, F., and Tschop, M.H. 2009. Direct control of peripheral lipid deposition by CNS GLP-1 receptor signaling Is mediated by the sympathetic nervous system and blunted in diet-Induced obesity. J. Neurosci. **29**: 5916–5925.

Oksbjerg, N., Gondret, F., and Vestergaard, M. 2004. Basic principles of muscle development and growth in meat-producing mammals as affected by the insulin-like growth factor (IGF) system. Domest. Anim. Endocrinol. **27**: 219–240.

Otto, G., Roehe, R., Looft, H., Thoelking, L., Knap, P.W., Rothschild, M.F., Plastow, G.S., and Kalm, E. 2007. Associations of DNA markers with meat quality traits in pigs with emphasis on drip loss. Meat Sci. **75**: 185–195.

te Pas, M.F.W., Kruijt, L., Pierzchala, M., Crump, R.E., Boeren, S., Keuning, E., Hoving-Bolink, R., Hortós, M., Gispert, M., Arnau, J., Diestre, A., and Mulder, H.A. 2013. Identification of

proteomic biomarkers in M. Longissimus dorsi as potential predictors of pork quality. Meat Sci. **95**: 679–687.

Pearen, M.A., Eriksson, N.A., Fitzsimmons, R.L., Goode, J.M., Martel, N., Andrikopoulos, S., and Muscat, G.E.O. 2012. The nuclear receptor, Nor-1, markedly increases type II oxidative muscle fibers and resistance to fatigue. Mol. Endocrinol. **26**: 372–84.

Pearen, M.A., Goode, J.M., Fitzsimmons, R.L., Eriksson, N.A., Thomas, G.P., Cowin, G.J., Wang, S.-C.M., Tuong, Z.K., and Muscat, G.E.O. 2013. Transgenic muscle-specific Nor-1 expression regulates multiple pathways that effect adiposity, metabolism, and endurance. Mol. Endocrinol. **27**: 1897–917.

Pearen, M.A., Myers, S.A., Raichur, S., Ryall, J.G., Lynch, G.S., and Muscat, G.E.O. 2008. The orphan nuclear receptor, NOR-1, a target of beta-adrenergic signaling, regulates gene expression that controls oxidative metabolism in skeletal muscle. Endocrinology **149**: 2853–65.

Pearen, M.A., Ryall, J.G., Maxwell, M.A., Ohkura, N., Lynch, G.S., and Muscat, G.E.O. 2006. The orphan nuclear receptor, NOR-1, is a target of β-adrenergic signaling in skeletal muscle. Endocrinology **147**: 5217–5227.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35: D61–D65.

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

Ramayo-Caldas, Y., Mach, N., Esteve-Codina, A., Corominas, J., Castelló, A., Ballester, M.,

Estellé, J., Ibáñez-Escriche, N., Fernández, A.I., Pérez-Enciso, M., and Folch, J.M. 2012. Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. BMC Genomics **13**: 547.

Rohrer, G.A., Nonneman, D.J., Miller, R.K., Zerby, H., and Moeller, S.J. 2012. Association of single nucleotide polymorphism (SNP) markers in candidate genes and QTL regions with pork quality traits in commercial pigs. Meat Sci. **92**: 511–518.

Rosenvold, K., Lærke, H.N., Jensen, S.K., Karlsson, A.H., Lundström, K., and Andersen, H.J. 2002. Manipulation of critical quality indicators and attributes in pork through vitamin E supplementation, muscle glycogen reducing finishing feeding and pre-slaughter stress. Meat Sci. **62**: 485–496.

Ruiter, A., Jongbloed, A.W., van Gent, C.M., Danse, L.H., and Metz, S.H. 1978. The influence of dietary mackerel oil on the condition of organs and on blood lipid composition in the young growing pig. Am. J. Clin. Nutr. **31**: 2159–2166.

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

Schuenke, M.D., Kopchick, J.J., Hikida, R.S., Kraemer, W.J., and Staron, R.S. 2008. Effects of growth hormone overexpression vs. growth hormone receptor gene disruption on mouse hindlimb muscle fiber type composition. Growth Horm. IGF Res. **18**: 479–486.

Seideman, S.C., Cross, H.R., Smith, G.C., and Durland, P.R. 1984. Factors associated with fresh meat color: a review. J. Food Qual. **6**: 211–237.

Shi, G., Boerwinkle, E., Morrison, A.C., Gu C., C.C., Chakravarti, A., and Rao, D.C. 2011. Mining gold dust under the genome wide significance level: A two-stage approach to analysis of GWAS. Genet. Epidemiol. **35**: 111–118.

Škrlep, M., Kavar, T., Santé-Lhoutellier, V., and Čandek-Potokar, M. 2009. Effect of I199V polymorphism on PRKAG3 gene on carcass and meat quality traits in Slovenian commercial pigs. J. Muscle Foods **20**: 367–376.

Sonesson, A.K., de Greef, K.H., and Meuwissen, T.H.E. 1998. Genetic parameters and trends of meat quality, carcass composition and performance traits in two selected lines of large white pigs. Livest. Prod. Sci. **57**: 23–32.

van der Steen, H.A.M., Prall, G.F.W., and Plastow, G.S. 2005. Application of genomics to the pork industry. J. Anim. Sci. **83**: E1–E8.

Sun, W., Ibrahim, J.G., and Zou, F. 2010. Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. Genetics **185**: 349–59.

Sutendra, G., Dromparis, P., Wright, P., Bonnet, S., Haromy, A., Hao, Z., McMurtry, M.S., Michalak, M., Vance, J.E., Sessa, W.C., and Michelakis, E.D. 2011. The role of Nogo and the mitochondria-endoplasmic reticulum unit in pulmonary hypertension. Sci. Transl. Med. **3**: 88ra55.

Svegliati-Baroni, G., Saccomanno, S., Rychlicki, C., Agostinelli, L., De Minicis, S., Candelaresi, C., Faraci, G., Pacetti, D., Vivarelli, M., Nicolini, D., Garelli, P., Casini, A., Manco, M., Mingrone, G., Risaliti, A., Frega, G.N., Benedetti, A., and Gastaldelli, A. 2011. Glucagon-like peptide-1 receptor activation stimulates hepatic lipid oxidation and restores hepatic signalling alteration induced by a high-fat diet in nonalcoholic steatohepatitis. Liver Int. **31**: 1285–1297.

Szydlowski, M., and Paczyńska, P. 2011. QTLMAS 2010: simulated dataset. BMC Proc. **5**: S3.

Tang, Z., Li, Y., Wan, P., Li, X., Zhao, S., Liu, B., Fan, B., Zhu, M., Yu, M., and Li, K. 2007. LongSAGE analysis of skeletal muscle at three prenatal stages in Tongcheng and Landrace

pigs. Genome Biol. **8**: R115.

Tsika, R.W., Schramm, C., Simmer, G., Fitzsimons, D.P., Moss, R.L., and Ji, J. 2008. Overexpression of TEAD-1 in transgenic mouse striated muscles produces a slower skeletal muscle contractile phenotype. J. Biol. Chem. **283**: 36154–36167.

Veum, V.L., Dankel, S.N., Gjerde, J., Nielsen, H.J., Solsvik, M.H., Haugen, C., Christensen, B.J., Hoang, T., Fadnes, D.J., Busch, C., Våge, V., Sagen, J. V, and Mellgren, G. 2012. The nuclear receptors NUR77, NURR1 and NOR1 in obesity and during fat loss. Int. J. Obes. **36**: 1195–1202.

de Vries, A.G., Faucitano, L., Sosnicki, A., and Plastow, G.S. 2000. The use of gene technology for optimal development of pork meat quality. Food Chem. **69**: 397–405.

Waldmann, P., Mészáros, G., Gredler, B., Fuerst, C., and Sölkner, J. 2013. Evaluation of the lasso and the elastic net in genome-wide association studies. Front. Genet. **4**: 270.

Wang, W., Meng, Q., Hu, X., and Li, N. 2009. Genetic variation and association of insulin-like growth factor binding protein-3 with performance in Swine. Biochem. Genet. **47**: 315–21.

Warriss, P.D., Brown, S.N., and Paściak, P. 2006. The colour of the adductor muscle as a predictor of pork quality in the loin. Meat Sci. **73**: 565–569.

Wilborn, B.S., Kerth, C.R., Owsley, W.F., Jones, W.R., and Frobish, L.T. 2004. Improving pork quality by feeding supranutritional concentrations of vitamin D. J. Anim. Sci. **82**: 218.

Wood, J.D. 1984. Fat deposition and the quality of fat tissue in meat animals. *in* J. Wiseman, ed. Fats in Animal Nutrition. Butterworth-Heinemann, London, UK.

Wood, J.D., Richardson, R.I., Nute, G.R., Fisher, A.V., Campo, M.M., Kasapidou, E., Sheard, P.R., and Enser, M. 2004. Effects of fatty acids on meat quality: a review. Meat Sci. **66**: 21–32.

Wu, T.T., Chen, Y.F., Hastie, T., Sobel, E., and Lange, K. 2009. Genome-wide association analysis by lasso penalized logistic regression. Bioinformatics **25**: 714–721.

Xu, S., Zhu, D., and Zhang, Q. 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. **111**: 12456–12461.

Yang, H., Wang, G., Sun, H., Shu, R., Liu, T., Wang, C., Liu, Z., Zhao, Y., Zhao, B., Ouyang, Z., Yang, D., Huang, J., Zhou, Y., Li, S., Jiang, X., Xiao, Z., Li, X.-J., and Lai, L. 2014. Species-dependent neuropathology in transgenic SOD1 pigs. Cell Res. **24**: 464–481.

Zhan, H., and Xu, S. 2012. Adaptive ridge regression for rare variant detection. PLoS One **7**: e44173.

Zhang, C., Wang, Z., Bruce, H., Kemp, R.A., Charagu, P., Miar, Y., Yang, T., and Plastow, G. 2015. Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs. BMC Genet. **16**: 33.

Zhang, D., Wang, X., Li, Y., Zhao, L., Lu, M., Yao, X., Xia, H., Wang, Y., Liu, M.-F., Jiang, J., Li, X., and Ying, H. 2014. Thyroid hormone regulates muscle fiber type conversion via miR-133a1. J. Cell Biol. **207**: 753–766.

Zou, H. 2006. The adaptive LASSO and its oracle properties. J. Am. Stat. Assoc. **101**: 1418–1429.

Table 2.1 Comparison between the adaptive LASSO and previously reported methods in detection power and precision

Method	**Detection Power[a]**	**Precision[b]**
Adaptive Lasso	**28 / 37 (76%)**	**44%**
BVSM[c]	10 / 37 (27%)	89%
BayesC[c]	15 / 37 (41%)	75%
PLSR[c]	2 / 37 (5%)	8%
GRAMMAR[c]	5 / 37 (14%)	56%
Haplotype inference[c]	7 / 37 (19%)	70%
DHGLM[c]	11 / 37 (30%)	78%
BayesCPi[c]	16 / 37 (43%)	87%

[a]The fraction of the 37 simulated QTLs that are close to at least one detected position

[b]The fraction of the detected position that are close to at least one simulated QTL.

[c]Analyses were conducted previously by different groups and summarized by Szydlowski and Paczyńska (2011)

Table 2.2 List of the 18 meat colour traits used in Study 2

#	Position	Measurements		
		Minolta L*	Minolta a*	Minolta b*
1	longissimus dorsi muscle on the fresh loin	FMCOL L*	FMCOL a*	FMCOL b*
2	thawed loin muscle	TMCOL L*	TMCOL a*	TMCOL b*
3	gluteus medius muscle on the fresh ham surface	GMCOL L*	GMCOL a*	GMCOL b*
4	quadriceps femoris muscle on the fresh ham surface	QFCOL L*	QFCOL a*	QFCOL b*
5	iliopsoas muscle on the fresh ham surface	ICOL L*	ICOL a*	ICOL b*
6	subcutaneous fat above the longissimus dorsi muscle on the fresh loin	FCOL L*	FCOL a*	FCOL b*

Table 2.3 Associations between the meat colour traits and candidate regions.

Chr	Start (Mbp)	End (Mbp)	TMCOL L*	TMCOL a*	TMCOL b*	FMCOL L*	FMCOL a*	FMCOL b*	FCOL L*	FCOL a*	FCOL b*	GMCOL L*	GMCOL a*	GMCOL b*	QFCOL L*	QFCOL a*	QFCOL b*	ICOL L*	ICOL a*	ICOL b*
1	32	36				*SA*		*SA*								*SA*		*S*		
1	268	270	*SA*								*SA*							*S*		
2	44	46		*A*			*SA*													
2	50	54		*SA*		*A*	*SA*	*SA*					*SA*	*SA*				*S*		
2	154	156					*SA*	*S*				*S*	*SA*	*SA*						
3	14	16	*SA*										*A*	*SA*						
3	90	92													*SA*		*SA*	*S*		
7	0	2				*A*	*SA*	*SA*							*S*			*S*		
7	26	28	*SA*			*SA*		*S*										*S*		*S*
7	38	40					*SA*				*SA*									
9	60	62													*SA*		*A*			

Table 2.3 (Continued)

Chr	Start (Mbp)	End (Mbp)	TMCOL L*	TMCOL a*	TMCOL b*	FMCOL L*	FMCOL a*	FMCOL b*	FCOL L*	FCOL a*	FCOL b*	GMCOL L*	GMCOL a*	GMCOL b*	QFCOL L*	QFCOL a*	QFCOL b*	ICOL L*	ICOL a*	ICOL b*
10	48	50	*SA*			*A*														
12	14	16				*A*		*SA*								*S*				
12	60	62		*A*			*SA*													
13	88	90					*SA*	*SA*												
14	64	66		*A*			*SA*											*S*		
15	130	134			*SA*	*A*	*S*	*S*				*SA*	*S*	*SA*	*S*	*SA*	*SA*	*S*		
16	82	84	*SA*					*S*								*SA*		*S*		
17	46	48	*S*		*SA*		*SA*													
18	54	58				*SA*	*SA*	*SA*						*SA*						

Note: Associations detected with the adaptive LASSO are indicated as *A*, and those detected in the single-marker analysis are indicated as *S*.

Table 2.4 Candidate regions associated with multiple meat colour traits

Chr[a]	Pos[b]	# Asso (1)[c]	# Asso (2)[d]	#QTL[e]	Gene[f]	Distance[g]
1	32-36 Mbp	3	4	17		
1	268-270 Mbp	2	3	14	*NR4A3*	808 kbp
2	44-46 Mbp	2	1	9	*MYOD1*	66 kbp
2	50-54 Mbp	6	6	11	*TEAD1*	160 kbp
2	154-156 Mbp	3	5	11		
3	14-16 Mbp	3	2	4		
3	90-92 Mbp	2	3	4	*RTN4*	173 kbp
7	0-2 Mbp	3	3	0		
7	26-28 Mbp	2	6	5	*TNF*	100 kbp
7	38-40 Mbp	2	3	36	*GLP1R*	4 kbp
9	60-62 Mbp	2	1	1		
10	48-50 Mbp	2	1	3		
12	14-16 Mbp	2	2	1	*GH1*	72 kbp
12	60-62 Mbp	2	1	6	*MYOCD*	254 kbp
13	88-90 Mbp	2	2	0		
14	64-66 Mbp	2	2	1	*ACTA1*	349 kbp
15	130-134 Mbp	6	10	28	*PRKAG3*	30 kbp
16	82-84 Mbp	3	4	4		

[a] The chromosome that the candidate region was located

[b] The location of the candidate region. (1 Mbp = 10^6 bp)

[c] The number of traits that were associated with the region in the adaptive LASSO

[d] The number of traits that were associated with the region in the single marker association

[e] The number of the related QTL overlapping with the region

[f] Gene(s) that may be functionally connected to meat colour in the region

[g] The distance between the gene in column *Gene* (if applicable) and the closest SNP that was significant in the single marker association. (1 kbp = 10^3 bp)

Table 2.5 Previous estimates (Miar et al. 2014) of genetic (below diagonal) and phenotypic (above diagonal) correlations, heritabilities (diagonal), and their standard error of estimates among the meat colour traits for the same population of pigs.

Parameters of fat colour traits were not presented as it was not included in the previous study.

TRAIT	TMCOL L*	TMCOL a*	TMCOL b*	FMCOL L*	FMCOL a*	FMCOL b*	GMCOL L*	GMCOL a*	GMCOL b*	QFCOL L*	QFCOL a*	QFCOL b*	ICOL L*	ICOL a*	ICOL b*
TMCOL	**0.28**	**–0.19**	**0.65**	**0.23**	0.03	**0.15**	**0.22**	0.02	**0.18**	**0.18**	–0.00	**0.14**	**0.12**	–0.01	**0.10**
L*	**(0.06)**	**(0.03)**	**(0.02)**	**(0.03)**	(0.03)	**(0.03)**	**(0.03)**	(0.03)	**(0.03)**	**(0.03)**	(0.03)	**(0.03)**	**(0.03)**	(0.03)	**(0.03)**
TMCOL	**–0.36**	**0.26**	**0.43**	0.01	**0.37**	**0.19**	–0.02	**0.25**	**0.10**	0.00	**0.23**	**0.11**	0.02	**0.11**	0.06
a*	**(0.16)**	**(0.09)**	**(0.04)**	(0.03)	**(0.03)**	**(0.04)**	(0.03)	**(0.04)**	**(0.05)**	(0.03)	**(0.03)**	**(0.04)**	(0.03)	**(0.03)**	(0.04)
TMCOL	**0.60**	**0.59**	**0.31**	**0.17**	**0.19**	**0.21**	**0.16**	**0.13**	**0.18**	**0.15**	**0.11**	**0.17**	**0.08**	0.02	**0.08**
b*	**(0.10)**	**(0.16)**	**(0.06)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	(0.03)	**(0.03)**
FMCOL	0.20	–0.21	–0.03	**0.31**	**0.23**	**0.77**	**0.34**	0.04	**0.30**	**0.30**	–0.01	0.24	**0.24**	**0.06**	**0.23**
L*	(0.16)	(0.17)	(0.16)	**(0.06)**	**(0.03)**	**(0.01)**	**(0.02)**	(0.03)	**(0.02)**	**(0.02)**	(0.03)	(0.02)	**(0.03)**	**(0.03)**	**(0.03)**
FMCOL	**–0.30**	**0.76**	0.19	**–0.40**	**0.36**	**0.72**	0.02	**0.42**	**0.18**	**0.08**	**0.29**	**0.17**	0.04	**0.19**	**0.11**
a*	**(0.15)**	**(0.16)**	(0.14)	**(0.15)**	**(0.06)**	**(0.01)**	(0.03)	**(0.02)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	(0.03)	**(0.03)**	**(0.03)**

Table 4 (Continued)

TRAIT	TMCOL L*	TMCOL a*	TMCOL b*	FMCOL L*	FMCOL a*	FMCOL b*	GMCOL L*	GMCOL a*	GMCOL b*	QFCOL L*	QFCOL a*	QFCOL b*	ICOL L*	ICOL a*	ICOL b*
FMCOL	–0.10	0.32	0.11	**0.51**	**0.46**	**0.20**	**0.22**	**0.24**	**0.31**	**0.24**	**0.13**	**0.26**	**0.20**	**0.14**	**0.24**
b*	(0.18)	(0.19)	(0.18)	**(0.12)**	**(0.13)**	**(0.06)**	**(0.03)**	**(0.03)**	**(0.02)**	**(0.02)**	**(0.03)**	**(0.02)**	**(0.03)**	**(0.03)**	**(0.03)**
GMCOL	**0.35**	–0.29	0.03	**0.45**	**–0.42**	–0.06	**0.22**	0.04	**0.80**	**0.31**	–0.01	**0.24**	**0.24**	–0.04	**0.18**
L*	**(0.17)**	(0.18)	(0.17)	**(0.15)**	**(0.16)**	(0.20)	**(0.05)**	(0.03)	**(0.01)**	**(0.02)**	(0.03)	**(0.02)**	**(0.03)**	(0.03)	**(0.03)**
GMCOL	–0.19	**0.43**	0.14	**–0.30**	**0.55**	0.13	**–0.42**	**0.38**	**0.48**	**0.09**	**0.32**	**0.19**	**0.12**	**0.26**	**0.21**
a*	(0.14)	**(0.16)**	(0.14)	**(0.14)**	**(0.11)**	(0.17)	**(0.15)**	**(0.06)**	**(0.02)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.02)**	**(0.03)**
GMCOL	0.23	0.13	0.32	0.33	0.03	0.39	**0.56**	**0.34**	**0.12**	**0.25**	**0.13**	**0.29**	**0.26**	**0.10**	**0.30**
b*	(0.21)	(0.23)	(0.21)	(0.19)	(0.21)	(0.21)	**(0.14)**	**(0.17)**	**(0.05)**	**(0.02)**	**(0.03)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**
QFCOL	**0.40**	–0.28	0.06	**0.66**	–0.18	0.28	**0.74**	–0.20	0.36	**0.19**	–0.03	**0.74**	**0.24**	0.03	**0.19**
L*	**(0.17)**	(0.20)	(0.29)	**(0.14)**	(0.18)	(0.20)	**(0.15)**	(0.17)	(0.22)	**(0.05)**	(0.03)	**(0.01)**	**(0.03)**	(0.03)	**(0.03)**
QFCOL	**–0.40**	**0.69**	0.09	–0.16	**0.53**	0.19	–0.18	**0.52**	0.24	**–0.46**	**0.27**	**0.53**	**–0.06**	**0.23**	**0.06**
a*	**(0.15)**	**(0.18)**	(0.16)	(0.16)	**(0.13)**	(0.18)	(0.18)	**(0.12)**	(0.21)	**(0.17)**	**(0.06)**	**(0.02)**	**(0.03)**	**(0.02)**	**(0.03)**

Table 4 (Continued)

TRAIT	TMCOL L*	TMCOL a*	TMCOL b*	FMCOL L*	FMCOL a*	FMCOL b*	GMCOL L*	GMCOL a*	GMCOL b*	QFCOL L*	QFCOL a*	QFCOL b*	ICOL L*	ICOL a*	ICOL b*
QFCOL	0.04	0.19	0.27	**0.67**	0.19	**0.71**	**0.70**	0.05	**0.87**	**0.67**	0.20	**0.10**	**0.13**	**0.12**	**0.18**
b*	(0.23)	(0.23)	(0.21)	**(0.17)**	(0.21)	**(0.19)**	**(0.22)**	(0.21)	**(0.25)**	**(0.13)**	(0.21)	**(0.04)**	**(0.03)**	**(0.02)**	**(0.02)**
ICOL L*	0.16	0.03	0.03	**0.39**	–0.15	0.30	0.24	–0.06	**0.36**	0.09	–0.18	0.10	**0.32**	**0.17**	**0.83**
	(0.16)	(0.17)	(0.15)	**(0.14)**	(0.15)	(0.17)	(0.16)	(0.14)	**(0.18)**	(0.18)	(0.16)	(0.22)	**(0.06)**	**(0.03)**	**(0.01)**
ICOL a*	–0.21	0.16	–0.13	0.06	**0.43**	**0.44**	–0.34	**0.58**	0.11	0.01	**0.38**	0.21	–0.29	**0.16**	**0.52**
	(0.19)	(0.20)	(0.19)	(0.19)	**(0.16)**	**(0.19)**	(0.20)	**(0.15)**	(0.25)	(0.22)	**(0.17)**	(0.26)	(0.19)	**(0.05)**	**(0.02)**
ICOL b*	0.12	0.03	–0.01	**0.31**	–0.00	**0.48**	0.13	0.04	**0.46**	0.04	–0.11	0.31	**0.92**	0.00	**0.26**
	(0.17)	(0.18)	(0.16)	**(0.15)**	(0.16)	**(0.17)**	(0.18)	(0.15)	**(0.18)**	(0.19)	(0.17)	(0.39)	**(0.03)**	(0.21)	**(0.06)**

Note: Significant correlations are highlighted in bold ($P < 0.05$).

Figure 2.1 Manhattan plot from the GWAS for QT (QTLMAS 2010 dataset).

Each blue data point represents a 1-Mb SNP window. Red Points indicate QTLs along the genome, and the variance of their effects are represented by red lines (solid lines for additive effects and dashed lines for non-additive effect).

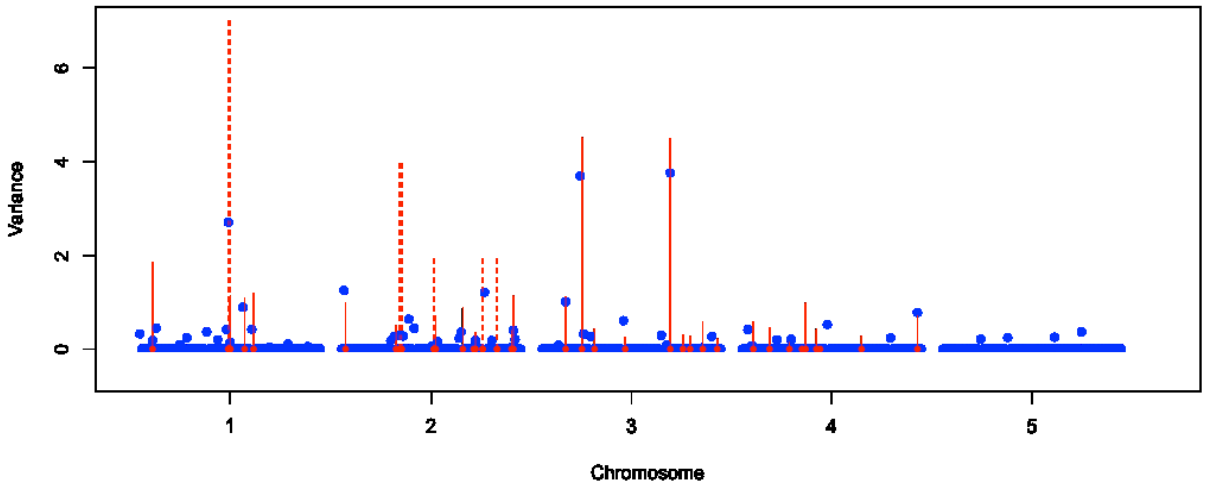
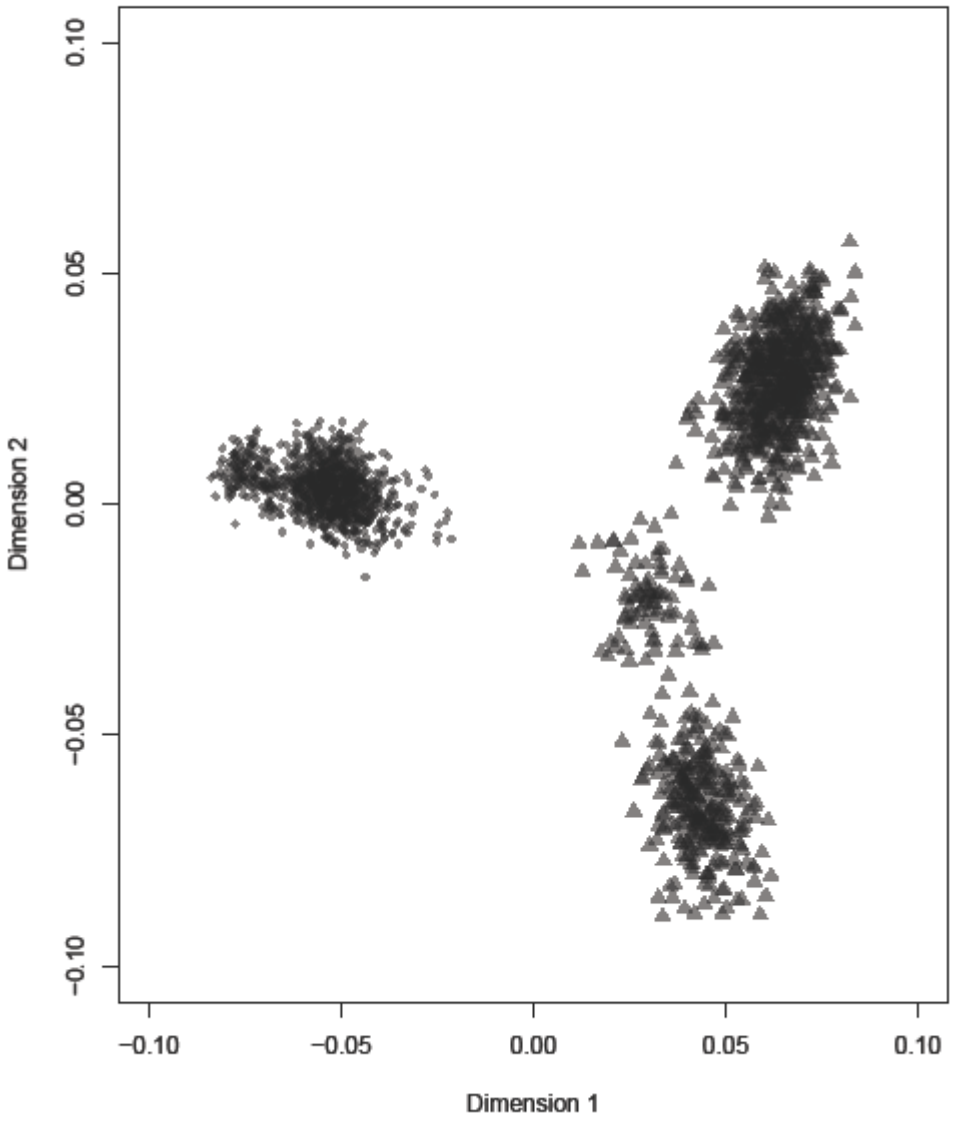


Figure 2.2 Multidimensional scaling result for assessing the structure of the population. The shape of the points (circle or triangle) indicates the source of the animals (Hypor Inc. or Genesus Inc.).



Chapter 3 A GWAS of fetal response to type 2 PRRSV challenge

In Chapter 2, we discussed the application of the adaptive LASSO in the GWAS. Even though the method resulted in an improved detection power, it failed to provide a quantitative measurement (such as P-value) about the probability that the detected associations were actually true (i.e. not happened by chance). Such measurement can help to determine the relative importance of multiple associations. In this Chapter, we introduce a GWAS of fetal response to type 2 PRRSV challenge, where we applied permutation tests to improve the precision of the original LASSO, and reported empirical P-values. The contents of this Chapter have been published in *Scientific Reports*[1]. I also helped co-author a review related to the GWAS[2].

3.1 Background

Porcine reproductive and respiratory syndrome is one of the most serious threats to pig production worldwide (Islam et al. 2013). Caused by PRRS virus (PRRSV), family Arteriviridae (Faaberg et al. 2012), it results in reproductive failure in sows, such as late-term abortion, premature delivery, stillborn or mummified fetuses, elevated preweaning mortality, and respiratory disease in neonatal and young pigs. Because of its widespread distribution, high mortality in infected herds, and poor performance in recovered herds (Rossow 1998), this disease is economically very important for

[1] Tianfu Yang, James Wilkinson, Zhiquan Wang, Andrea Ladinig, John Harding and Graham Plastow, 2016. "A genome-wide association study of fetal response to type 2 porcine reproductive and respiratory syndrome virus challenge", Scientific Reports, 6: 20305

[2] John C.S. Harding, Andrea Ladinig, Predrag Novakovic, Susan E. Detmer, Jamie M. Wilkinson, Tianfu Yang, Joan K. Lunney, and Graham S. Plastow. 2017. "Novel Insights into Host Responses and Reproductive Pathophysiology of Porcine Reproductive and Respiratory Syndrome Caused by PRRSV-2." (in press)

the swine industry. The estimated annual economic loss caused by PRRS in the US is more than 500 million US dollars, and its impact increased during the years 2005-2013 (Neumann et al. 2005; Holtkamp et al. 2013).

At present, developing effective vaccines against a wide range of PRRSV strains is still a challenge, as there is a gap in the knowledge of PRRS biology, pathogenesis and immunity (Karniychuk et al. 2012). Genomic tools may provide an alternative opportunity to explore the mechanisms behind PRRS, and to select animals resistant, or with reduced susceptibility to PRRSV (Lunney and Chen 2010). For example, a major quantitative trait locus (QTL) on SSC4 associated with host response to PRRSV has been reported (Boddicker et al. 2012) in a genome-wide association study (GWAS). Further analysis showed that estimated breeding values (EBV) based on this QTL were sufficiently accurate for potential use in animal selection to reduce the effects of PRRSV in growing pigs (Boddicker et al. 2014). In addition, a QTL on SSC7 associated with reproductive traits and PRRS IgG antibody response was discovered, and was subsequently confirmed in an unrelated population (Serão et al. 2014; Serao et al. 2014). However, since these important advances in our understanding of PRRS host responses were obtained from piglets post-weaning and dams, respectively, knowledge of the genetic basis of fetal response to PRRSV infection is still scarce.

To explore the mechanisms of reproductive PRRS, an experimental pregnant gilt challenge model (PGM) was undertaken (Ladinig et al. 2014d), and several phenotypic responses in dams and fetuses associated with PRRS severity were revealed (Ladinig et al. 2014a, 2014c, 2014b, 2015). The objective of the current study was to explore the genetic basis of fetal response to PRRSV infection, including viral load in thymus (VLT), viral load in endometrium (VLE), fetal death (FD) and fetal viability (FV), through GWAS using fetal samples from this challenge model.

3.2 Materials and methods

3.2.1 Animal resources

Samples and data used in the current study were obtained from a PRRS pregnant gilt challenge model, previously described in detail (Ladinig et al. 2014d). The experiment was approved by the University of Saskatchewan's Animal Research Ethics Board and adhered to the Canadian Council on Animal Care guidelines for humane animal use (permit #20110102). In brief, 114 purebred Landrace gilts (Fast Genetics Inc., Spiritwood, Canada) were artificially inseminated using homospermic semen from one of 24 Yorkshire boars, and inoculated with type 2 NVSL 97-7895 PRRSV on gestation day 85 ± 1 and 19 similarly mock inoculated gilts. All were humanely euthanized 21 days post inoculation (dpi). In total, 1,422 fetuses obtained from the PRRSV-inoculated gilts were categorized based on their preservation status as: viable (VIA, n = 697), meconium-stained (MEC, n = 125), decomposed (DEC, n = 111), autolysed (AUT, n = 459), or mummified (MUM, n = 30). All AUT and MUM fetuses were excluded from the present analyses as poor DNA yield and quality prevented genotyping. VIA were alive until termination and externally normal, whereas MEC were alive but clearly showed pathologic changes. It is estimated that DEC died 3-5 days prior to termination based on their fetal size and primarily normal external appearance (more than 50% white skin, lack of generalized subcutaneous edema and emphysema).

3.2.2 Phenotypic data

The viral load (VL; target RNA concentration per mg tissue) in fetal thymus (VLT) and endometrium (VLE) were measured in VIA, MEC and DEC fetuses using an in-house quantitative real-time PCR (qRT-PCR) (Ladinig et al. 2014d) targeted at a highly conserved region of the C-

terminal end of ORF7 of NVSL 97-7895. Endometrium was collected from the umbilical stump of each fetus and included the adherent fetal placental layers. RNA was extracted from 10-20 mg tissue using the RNeasy extraction kit (Qiagen, Toronto, Canada) as per the manufacturer's instructions. However, previous analysis showed that both VLT and VLE were considerably lower in DEC fetuses than in VIA and MEC, likely due to viral RNA degradation during the period of decomposition (Ladinig et al. 2014d). As the degree of RNA degradation was hard to measure and model, the VLT and VLE in DEC fetuses (n = 111) were excluded from the association analysis. The random errors of raw phenotypic values for VL traits were assumed to follow a lognormal distribution (Islam et al. 2013), and were log-transformed (base 10) before the analysis (a phenotypic value of 0 was given to negative records). Two binary traits were also defined based on the fetal preservation: 1) fetal death (FD), where all DEC fetuses were coded as "dead" and all VIA and MEC fetuses were coded as "live", and 2) fetal viability (FV), where all VIA fetuses were coded as "viable" and all MEC and DEC fetuses were coded as "non-viable". We presumed that the genetic basis behind these two binary traits were different, so they were analyzed separately.

The phenotype of VLT and VLE for genotyped fetuses was summarized in Table 3.5 and Figure 3.1. It is noticeable that the VLT and VLE are not symmetrically distributed, but appear to follow a mixture of two distributions, one with high density near zero and the other roughly normal but spread over higher values.

Some other traits were also measured as potential environmental effects on fetal response to PRRSV. Inoculated dams were genotyped for the WUR10000125 SNP (Dam-WUR-SNP), which was associated with the SSC4 QTL (Boddicker et al. 2012). Their serum viral load was also measured at day 0, 2, 6, 21 post-inoculation (Dam-VL-0, 2, 6, 21), using qRT-PCR and the same

primers used for fetal viral load measurement. The area under the curve (Dam-VL-AUC0-21) was calculated from 0 to 21 dpi.

3.2.3 Genomic data

In total, 928 fetuses with high DNA quality were genotyped using the PorcineSNP60 Genotyping BeadChip v2 (Illumina, San Diego, CA, USA) containing 61,565 SNPs. SNPs were filtered out when they: 1) had a call rate less than 90%; 2) had a minor allele frequency (MAF) less than 0.05; or 3) demonstrated a significant deviation from Hardy-Weinberg equilibrium (HWE) with a Chi^2-value higher than 600. After the quality control, 45,255 SNPs remained in the dataset, with a missing call rate of 0.32%.

3.2.4 Population structure

To assess the potential impact of population stratification, the population structure of the fetuses was tested before the association analysis using Plink 1.90 (Chang et al. 2015) and R/rrBLUP (Endelman 2011). Genetic difference among fetuses was measured as pairwise identity-by-state (IBS) Hamming distances in Plink, and a MDS analysis was conducted to construct the 2-D plot, which shows the first 2 dimensions of the population structure. A PCA by the eigenvalue decomposition of marker-based relationship matrix was also performed using R/rrBLUP. As there was no evidence that the population was substantially stratified, the population structure was not modelled in the following analysis.

3.2.5 Model

A generalized linear model was used in the analyses:

$$f[E(y)] = \mu + X\beta + Zg$$

Two link functions (*f*) were used to transform the expectation of phenotypes (*E(y)*) to the linear predictor. For continuous phenotypes (VLE, VLT), an identity link function was used which was equivalent to a general linear model. For the two preservation traits (FD, FV), a logit link function was used to model the binary phenotype. μ was the intercept. *X* was the design matrix for environmental fixed effects, and β was the vector of environmental fixed effects, which included two litter factors (litter size and litter fetal mortality rate), three factors measuring maternal disease status (Dam-WUR-SNP, Dam-VL-21, Dam-VL-AUC0-21), three factors related to maternal uterine environment (relative fetal position within uterine horn increasing incrementally from tip to body [POS], nINF, nDEAD), and two other factors (fetal sex, experimental repetition/group). *Z* was the design matrix associated with SNP effects (*g*). In the design matrix, genotypes were coded as 1/0/-1 for genotype AA/AB/BB, respectively. All SNP effects (*g*) were treated as fixed effects.

The model was fitted using the least absolute shrinkage and selection operator (LASSO), which was reported to be appropriate for fitting fixed effects models in GWAS and genomic prediction (Wu et al. 2009; Usai et al. 2009; Ayers and Cordell 2010; Xu et al. 2014). The analysis was performed using R/glmnet package (Friedman et al. 2010), and the tuning parameter (λ) was selected to minimize the mean square error (MSE) in a 10-fold cross validation (CV) for each run. As the package does not accept design matrix with missing values, all missing genotypes were imputed as the overall average value of the marker. After the fitting of the model, the total genetic variance explained by all 45,255 SNPs ($\sigma^2_{g,total}$) and the genetic variance explained by each single marker ($\sigma^2_{g,marker}$) were calculated using the estimated SNP effects (Fernando and Garrick 2013).

3.2.6 Permutation test

Since R/glmnet has a very high computational efficiency (Friedman et al. 2010), the significance of those non-zero SNP effects were tested through permutation, by randomly shuffling the phenotype while keeping the genotype intact thereby destroying the association between the phenotype and genotype. Then, the permutated data were analysed with the same procedure to generate a null-distribution that was used to determine empirical test criteria underlying the null hypothesis (Hayes 2013). One thousand runs were conducted for each trait. In each run, the $\sigma^2_{g,marker}$ for each marker was calculated. The highest $\sigma^2_{g,marker}$ in each run ("highest-σ^2") was used to construct the null distribution. For each SNP, its *P*-value was calculated as the proportion of highest-σ^2 that were greater than the genetic variance explained by that SNP. When the proportion was equal to zero, it implied a *P*-value less than 0.001. The 99^{th} and 95^{th} percentiles (corresponding to an empirical *P*-value of 0.01 and 0.05) of the null distribution were used as the two critical values to test the result for that trait. Given the limited sample size in the current dataset (less than 1,000 fetuses) and our relatively robust test strategy, the 90^{th} percentile (corresponding to an empirical *P*-value of 0.1) was also calculated to explore more potential SNPs.

3.2.7 Transcriptomic analyses

To provide a functional context, GWAS results were compared with that of two differential gene expression experiments, in order to detect genes close to the candidate SNPs whose expression was altered in response to PRRSV infection. The results of the transcriptomic analyses came from a related study (Wilkinson et al. 2016) that investigated the fetal transcriptomic response to PRRSV infection, and used the same population of fetuses. In brief, for each of the two

experiments, the gene expression profiles of two groups of fetuses (n = 12 per group) were compared. The first experiment compared fetuses from mock-inoculated control gilts (CON) to viable, qRT-PCR positive fetuses from PRRSV-inoculated gilts (INF). The second experiment compared INF fetuses to viable, uninfected (qRT-PCR negative) fetuses from PRRSV-inoculated gilts (UNINF). All differentially expressed genes located within a 4Mbp window (2Mbp upstream or downstream) of any of the candidate SNPs were identified. The results may provide support for possible QTLs in those regions, and aid in the identification of the causative SNP if it affects gene expression. Although those genes without differential expression may also contain causative polymorphisms, this functional analysis was mainly focused on those differentially expressed genes, as they are more likely to play a role in the fetal response to PRRSV.

The location of all SNPs were based on *Illumina Pig 60k SNPs mapped to pig genome assembly 10.2* (Groenen et al. 2012). The searching of genes near candidate SNPs was performed using BioMart, an online service that integrated the information from top bioinformatic databases, and R/Bioconductor package biomaRt, an R interface for BioMart (Durinck et al. 2005, 2009).

3.3 Results and discussion

3.3.1 Population structure

Figure 3.2 presents the result of multi-dimension scaling (MDS) and provides a visualization of the pairwise genetic distances among the fetuses. There was no evidence of population stratification based on the 2-dimensional plot. This conclusion was also supported by the principal component analysis (PCA) result. The eigenvalues for the top three principal components were only 2.49%, 2.45% and 2.26% of the summation of all eigenvalues, which implies a lack of population stratification.

3.3.2 Association results

In the regression analysis, 24 candidate SNPs were found associated with at least one of the four traits (VLT, VLE, FD, FV). Twenty-two could be mapped to the porcine genome and are located across 10 chromosomes. We also evaluated the contribution of these SNPs to the phenotypic variation, both individually and collectively. The results are grouped below by the four traits analyzed.

3.3.2.1 Viral load - thymus (VLT)

Results of the associations between SNPs and VLT were summarized in Table 3.1 and Figure 3.3a. One single nucleotide polymorphism (SNP) on each of *Sus scrofa* chromosome (SSC) 1 (ASGA0005344, $P = 0.021$) and SSC14 (DIAS0000654, $P = 0.044$) were significantly associated, and another SNP on SSC12 (ASGA0055300, $P = 0.069$) showed suggestive association with viral load. The three associated SNPs explained 5.58%, 4.16%, and 3.45% of $\sigma^2_{g,total}$ (total genetic variance explained by all SNPs passed quality control, see Methods section), respectively; and 13.0% collectively.

For these three associated SNPs, a simplified linear model was used to investigate their genetic contribution. In this model, only two non-zero environmental effects, number of adjacent PRRSV-infected fetuses (nINF) and number of adjacent dead fetuses (nDEAD), and the three SNPs were fitted. All three SNPs were significantly associated with VLT ($P < 0.001$) in the regression. The least-square (LS) means for the SNPs were also calculated (Figure 3.4). For the SNP ASGA0005344 on SSC1, the difference in LS mean between the two homozygous genotypes was 1.35 ($\log_{10}$ copies/mg). For the SNP ASGA0055300 on SCC12, the LS mean for the

heterozygotes was obviously deviated from the average of the LS means for homozygotes, which implies the existence of a dominance effect. However, it should be noted that the number of fetuses with genotype AA is relatively small and the estimate of viral load may be an underestimate so that the true effect is additive.

We also explored the distribution of the phenotypic values, to determine the interaction among the three SNPs. For each SNP, we determined the favoured allele, which was associated with a lower VLT. Fetuses were grouped based on how many favoured alleles they had across the three SNPs. The distribution for each group is shown in Figure 3.5. It is evident that the proportion of fetuses with high VLT decreases when more favoured alleles were present in individual fetuses, which supports the conclusion that the effect of the three SNPs was additive.

3.3.2.2 Viral load - endometrium (VLE)

No SNPs were significantly or suggestively associated with VLE (Figure 3.3b; $P > 0.1$). However, the top SNP on SSC15 (ALGA0115095, located at ~134Mbp) was very close to a candidate SNP associated with FD.

3.3.2.3 Fetal death (FD)

SNPs associated with FD are summarized in Table 3.2 and Figure 3.3c. Sixteen SNPs were significantly associated with FD: seven SNPs with $P < 0.01$ and nine other SNPs with $P < 0.05$. Four additional SNPs had a suggestive association ($P < 0.1$). The *P*-values for each SNP were listed in Table 3.2. Two SNPs were unmapped in the latest genome map (build 10.2). In total, the 18 mapped candidate SNPs accounted for between approximately 1% and 4.4% of the genetic

variance ($\sigma^2_{g,total}$) each and 35.4% of the variance overall, while the 7 most significant SNPs explain 20% of the genetic variance.

To explore if the associated genes interacted, the fetuses were grouped based on the number of favoured alleles they had and three plots were generated (Figure 3.6) corresponding to sets of mapped SNPs: 1) seven SNPs with *P*-value lower than 0.01, 2) fifteen SNPs with *P*-value lower than 0.05, and 3) eighteen SNPs with *P*-value lower than 0.1. The proportion of viable (VIA), meconium-stained (MEC) and decomposed (DEC) fetuses was then compared for fetuses grouped by the number of favoured alleles they possessed. Whereas VIA fetuses appeared developmentally normal, MEC fetuses showed early signs of PRRS-related pathology, and DEC had died an estimated 3-5 days prior to termination (Ladinig et al. 2014d). Across all three plots, the proportion of DEC decreased as the number of favoured alleles increased. This trend was consistent regardless of the number of SNPs used in the analyses or their level of significance. This could be partly due to the low number of fetuses with an extreme number of favoured alleles in the dataset. Thus, a larger population may help to more accurately reveal the distribution for fetuses with extreme genotypes. In addition, the number of favoured alleles did not appear to affect the proportion of MEC fetuses, which support our assumption that the two traits, FD and FV, may each have a different genetic basis to a degree.

3.3.2.4 Fetal viability (FV)

Results of GWAS for FV were summarized in Table 3.3 and Figure 3.3d. Only one SNP on SSC7 was found to be significant ($P = 0.018$), which accounted for a considerable amount (34.0%) of total genetic variance ($\sigma^2_{g,total}$).

3.3.3 Overlap with previously discovered QTL

Some of the SNPs were found to be linked to candidate regions identified in previous studies. For example, the associated SNP located at 97-98 Mbp on SSC7 was very close to one of the candidate regions found associated with percentage of piglets born dead (PBD) in a commercial farm experiencing a PRRS outbreak (Serao et al. 2014). Even though PBD was a trait measured at the litter-level and was associated with the genotype of sows, it is still noteworthy as both FD and PBD relate to fetal death. Based on a pigQTLdb search (Hu et al. 2013), seven of the SNPs identified here were located within QTLs reported to be associated with pig disease resistance. One SNP associated with VLT is within a QTL of C3c serum concentration, which is a measurement of complement activity in innate immunity (Phatsara et al. 2007). Six SNPs associated with FD are located in previously described QTL that are related to leukocyte subset percentage, interferon level, Toll-like receptor level, pathogen count and reproductive traits (Phatsara et al. 2007; Galina-Pantoja et al. 2009; Lu et al. 2011b, 2011a; Onteru et al. 2012). More information about the pigQTLdb comparison is summarized in Table 3.4. However, we did not find any associated SNPs within the major QTL detected on SSC4 associated with viral load and growth rate in a PRRS nursery pig model (Boddicker et al. 2012) nor the genomic region on SSC7 associated with antibody response in a PRRS reproductive outbreak (Serao et al. 2014).

3.3.4 Potential link with PRRS

An association suggests that there are mutations in a gene or genes in these regions that explain the observed phenotype. In order to investigate this further, we identified genes within the candidate regions to determine if there was support for a functional basis of the observation.

We defined 21 regions for the 22 candidate SNPs with map locations (SNP ASGA0055300 and M1GA0017106 are close together and therefore share the same candidate region). Then we searched for genes that were differentially expressed in an ongoing related transcriptomic analyses of fetal response to PRRSV infection (see Methods section). For one of the candidate regions on SSC1 (~197Mbp), all four genes identified in this region were excluded due to their very low level of expression in fetal thymus (in the related transcriptomic analysis). Candidate regions on SSC13 (~64Mbp) and SSC14 (~86Mbp) do not harbour any differentially expressed genes in the transcriptomic analysis. However, we found differentially expressed genes in all of the other 18 candidate regions. In the following sections, we propose some hypotheses about the potential functional links between those differentially expressed genes and the fetal response to PRRSV challenge.

3.3.4.1 Mechanisms of reproductive PRRS

Recent pathogenesis and immunological studies provide new information about the biology of fetal response to PRRSV infection. PRRSV has the capability to replicate in the fetus at any stage of gestation, as demonstrated by direct intra-fetal inoculation, but in a natural infection must first cross from dam to fetus via the placenta. The precise conditions required for transplacental infection are not fully understood, but are largely restricted to late in gestation (Karniychuk and Nauwynck 2013). In the current study, gilts were inoculated in late gestation (day 85 ±1), which resulted in the death of about 40% of fetuses (Ladinig et al. 2014d). Previous studies suggested that PRRSV may cross the placental barrier through maternal macrophage migration (Karniychuk and Nauwynck 2013). Moreover, larger fetuses, which tend to have larger placentae, exhibit higher PRRS viral load and therefore appear to be more susceptible to PRRSV infection (Ladinig et al.

2014a). It seems likely that damage to the placental attachment site contributes to the reproductive pathology observed in PRRS. PRRSV-induced macrophage apoptosis in the placenta may lead to focal detachment and degeneration of the fetal placenta, resulting in fetal death (Karniychuk et al. 2011). However, events in the fetus itself may also influence the outcome of fetal infection. It has also been shown that both the status of adjacent fetuses, and the presence of PRRSV RNA (particularly at high levels in the fetus), are associated with fetal death, which suggests key roles for inter-fetal transmission and viral replication within fetuses (Ladinig et al. 2015). Fetuses are immunocompetent as early as 79 days of gestation (Butler et al. 2014). However, it is reported that immunity may remain dysregulated even in fetuses surviving *in utero* infection of PRRSV (Aasted et al. 2002; Nielsen et al. 2003). These results provide the basis for investigating the potential function of the regions identified here.

3.3.4.2 RIG-I pathway

Deconjugation of ubiquitin and ISG15 (IFN-stimulated gene product of 15kDa) is involved in one of the modulation mechanisms by which PRRSV evades host immune responses (Frias-Staheli et al. 2007). As ISG15 plays an important role in antivirus defence through protein ISGylation, deconjugation of ISG15 leads to further inhibition of downstream signalling and innate immune responses, such as NF-kappa-B activation (Sun et al. 2012). In one of our candidate genomic regions, we found the gene *UBA7* (also known as *UBE1L*, encoding Ubiquitin-Like Modifier Activating Enzyme 7, located at ~36Mbp on SSC13). The encoded enzyme catalyzes the conjugation of ISG15, and is critical in the protein ISGylation process (Madani et al. 2002). These findings suggest the possibility that a mutation in this region of the genome may alter the

expression of *UBA7* or the function of the encoded enzyme, and therefore modulate the host response to PRRSV.

3.3.4.3 Monocyte/macrophage lineage cells

PRRSV shows a strong tropism for monocyte/macrophage lineage cells, and it is reported that the differentiation and activation of these cells critically affect their susceptibility (Duan et al. 1997). In our candidate regions, we found a set of genes critically involved in processes related to the differentiation and activation of monocytes to macrophages. They are gene *ACKR2* (also known as *CCBP2*, encoding atypical chemokine receptor 2, located at ~29Mbp on SSC13), *CSF1* (also known as *M-CSF*, encoding macrophage colony-stimulating factor 1, located at ~121Mbp on SSC4), *MST1* (also known as *MSP*, encoding macrophage stimulating 1, located at ~35Mbp on SSC13), and *MST1R* (also known as *RON*, encoding macrophage stimulating 1 receptor, located at ~36Mbp on SSC13). Gene variants in these regions may result in changes to related pathways and ultimately to differences in the host's innate and adaptive immune responses.

For example, ACKR2 is one of the receptors of CCL2 (Lee et al. 2014), chemokine (C-C Motif) ligand 2 (also known as MCP1). CCL2 is a key pro-inflammatory chemokine involved in the activation of monocytes (Ford et al. 2014) and ACKR2 may act antagonistically due to its ability to scavenge extracellular chemokines including CCL2 (Lee et al. 2014). A previous study also suggested associations between circulating monocyte count and two missense variants within the gene *ACKR2* in humans (Crosslin et al. 2013). In addition, Ladinig et al. previously reported for this challenge study that serum CCL2 levels in the gilts following PRRSV challenge were positively related to VL in serum and lung, but were not associated with the odds of fetal death

(Ladinig et al. 2014c). A hypothesis based on these results is that a mutation within gene *ACKR2* results in fetal death, through the interference of ACKR2 in the functioning of CCL2.

It is reported that CSF1 is involved in the maturation and differentiation of monocytes into macrophages (Pixley and Stanley 2004), and CSF1 injections significantly increase macrophages and circulating monocytes in mice (Hume et al. 1988).

Macrophage stimulating 1 (MST1) and its receptor (MST1R) establish the MSP-RON signalling system (Lutz and Correll 2003) and mediate second messenger pathways within macrophages, including the Phosphatidylinositol-3-kinase (PI3K) pathway and mitogen-activated protein kinase (MAPK) pathway (Yao et al. 2013), both of which are reported to be involved in host response to PRRSV (Hou et al. 2012; Wang et al. 2014). The interaction between MST1 and MST1R may strongly modulate the production of interleukins (ILs), especially IL-12, IL-15 and IL-18 (Morrison et al. 2004). All three ILs have been reported to modulate the activities of NK cells, such as their apoptosis, development and survival (Gately et al. 1998; Waldmann and Tagaya 1999).

3.3.4.4 Natural killer (NK) cells

It is reported that NK cell-mediated cytotoxicity was significantly decreased in pigs infected with PRRSV VR2332, a prototype type 2 PRRSV strain, and it appears to result in the suppression of IFN-γ production (Manickam et al. 2013). In our candidate regions, we found two genes that may be related to the activity of NK cells. They are *CXCR2* (encoding chemokine [C-X-C Motif] Receptor 2, located at ~133Mbp on SSC15) and *NKTR* (encoding natural killer triggering receptor, located at ~29Mbp on SSC13, although this gene was not differentially expressed in the

transcriptomic analysis). These may play protective roles in PRRSV-induced immunosuppression, or be related to the modulation mechanism.

The product of *CXCR2* is a receptor that binds to a set of chemokines including IL-8 (also named CXCL8), which was reported to be a key factor in virus-induced selective chemotaxis of NK cells in humans (Burke et al. 2008). A previous study suggested that IL-8 is one of the important cytokines involved in the clearance of virus from serum in PRRSV-infected pigs (Lunney et al. 2010). In the related cytokine profiling study by Ladinig et al., however, IL-8 level was not associated with viral load or fetal death, although it was significantly increased in PRRSV-stimulated peripheral blood mononuclear cells (PBMC) from infected pigs (Ladinig et al. 2014c).

NKTR is a cyclophilin-related protein, and the gene is exclusively expressed in NK cells (Anderson et al. 1993). The protein is believed to have an important role in NK cell cytotoxicity, and induce the production of IFN-γ (Frey et al. 1991). As NK cell cytotoxicity and production of IFN-γ are both involved in PRRSV modulation of host innate immune response, it is tempting to speculate that NKTR may be part of this process.

3.3.4.5 T-cells

PRRSV is able to modulate host T-cell responses. One mechanism is reported to be upregulation of the frequency of Foxp3+ T-regulatory cells (T_{reg}s), which secrete IL-10 and transforming growth factor β (TGFB) that suppress the host immune response (Manickam et al. 2013). A protein in the integrin family, αEβ7-integrin (also known as CD103), can help to retain T_{reg}s and therefore may be key to this process (Campbell and Ziegler 2007). We found the gene *ITGB7* (encoding β7-integrin, one of the two components of αEβ7-integrin) located in the candidate region at ~19Mbp on SSC5. Further *ITGB7* is one of the target genes of Foxp3+, and it is possible that Foxp3+ has a

direct effect on the expression of *ITGB7* (Campbell and Ziegler 2007). In this case, *ITGB7* may be involved in the PRRSV-induced upregulation of Foxp3+ T_{reg}s. On the other hand, β7-integrin is also related to the production of α4β7-integrin, which is involved in T-cell migration (Agace 2006). Gene *TMPO* (encoding thymopentin, located at ~90Mbp on SSC5), may also have an impact on the PRRSV-induced modulation of host adaptive immune response, as thymopentin may be involved in the regulation of T helper cells, both Th1 and Th2, and their related cytokines, such as IFN-γ (Cillari et al. 1992).

3.3.4.6 IRF3/7 signalling

In a candidate region on SSC15 (~145Mbp), we also found genes encoding three members of the nuclear antigen SP100 family (SP100, SP110 and SP140), which may be related to the suppression of interferon (IFN) production by PRRSV. It is reported that PRRSV is able to inhibit the activation of interferon regulatory factor 3 (IRF3), and thereby suppress the synthesis of interferon-β (IFN-β) (Sun et al. 2012). A previous study showed that SP100 may play a significant role in enhancing the production of IFN-β in IRF3/7 signalling (Schmid et al. 2014). Further evidence also suggests a role of the SP100 family in antiviral response through promyelocytic leukemia protein nuclear body (PML-NB) (Bloch et al. 2000), especially for viruses whose proteins localize to the nucleolus, such as PRRSV (Hiscox 2002).

3.3.4.7 Apoptosis and JNK signalling

Apoptosis is a critical process in the pathogenesis of PRRSV and the fetal response to PRRSV challenge. Evidence implies that PRRSV may be able to regulate the progress of apoptosis enhancing viral replication (Huo et al. 2013). It was also proposed that the PRRSV-induced

apoptosis at fetal implantation sites is a primary mechanism of fetal death (Karniychuk et al. 2011). Previous studies showed that c-Jun NH(2)-terminal kinase (JNK) pathway is critical to apoptosis, and that activation of the JNK pathway is required for PRRSV-induced apoptosis (Yin et al. 2012). The JNK pathway can be activated by either of two MAPK kinases (MAP2Ks), and these two MAP2Ks can be activated by a total of fourteen MAPK kinase kinases (MAP3Ks) (Dhanasekaran and Reddy 2008). In our candidate genomic regions, we found two genes encoding kinases in this list. They are gene *MAP2K4* (encoding MAPK kinases 4, located at ~60Mbp on SSC12) and *MAP3K12* (encoding MAPK kinase kinases 12, located at ~19Mbp on SSC5). Neither showed differential expression in the transcriptomic analysis. MAP3K12 is not a kinase for MAP2K4, but activates another MAP2K, MAP2K7. We also found gene *MAPKAPK3* (encoding MAPK-activated protein kinase 3, located at ~36Mbp on SSC13) in a candidate region, which was also not differentially expressed. MAPKAPK3 catalyzes the phosphorylation of heat shock proteins B1 (HSPB1) (Paul et al. 2010), which is able to reduce the activity of JNK and thereby protect stressed cells from apoptosis (Kennedy et al. 2014). Mutations of any of these three genes may have an effect on the activation of JNK and PRRSV-induced apoptosis, potentially leading to differential fetal responses, especially fetal death.

3.3.4.8 Response to secondary infection

In the candidate regions we also found a set of genes that function in the host response to pathogens, but have not been reported to be involved specifically in the response to single strand RNA viruses like PRRSV. These genes are *TLR9* (encoding toll-like receptor 9, located at ~38Mbp on SSC13), *TREX1* (encoding the prime repair exonuclease 1, located at ~35Mbp on SSC13), and

SLC11A1 (also known as *NRAMP1*, encoding solute carrier family 11 member 1, located at ~133Mbp on SSC15).

As a member of the TLR family, TLR9 recognizes CpG motifs common to both bacterial and viral DNA and RNA:DNA hybrids (Rigby et al. 2014). The stimulation of TLR9 triggers the production of type I interferons (IFN), cytokines that play an important role in controlling viral infections (Akira and Takeda 2004). Protein TREX1 is targeted as a microbial evasion strategy. This exonuclease can digest cytoplasmic single-stranded DNA decreasing its concentration thereby avoiding the stimulation of innate immune response (Yan et al. 2010). Protein SLC11A1 protects macrophages and is reported to be involved in resistance to bacterial infection in pigs (Peracino et al. 2006).

Given that PRRSV is able to modulate the host immune response, increasing the host's susceptibility to secondary infections (Calzada-Nova et al. 2010), these genes may have an indirect impact on fetal responses to PRRSV challenge, although they would not necessarily be expected to be associated with fetal outcome in this challenge model.

3.3.5 Limitations and future work

Although the potential genomic regions and their genes and pathways discussed above may provide important clues for the fine mapping of specific causative mutations, additional research is required to confirm the roles of any these genes. It should also be noted that although this GWAS was based on a relatively large number of fetuses, only one type 2 PRRSV strain and one termination time-point were used in the challenge experiment. Thus, the results may be specific to the experiment. Careful validation with different time-points and PRRSV strains, including type 1

PRRSV, are needed before generalizing the results. Some of this validation work is underway, including a new experiment with type 2 PRRSV using earlier time-point(s) for termination.

3.4 Conclusions

In this GWAS of fetal response to PRRSV challenge, specifically fetal viral load (VLT, VLE) and fetal death and viability (FD, FV), we found 21 candidate regions located on 10 chromosomes, with four of these regions being on SSC13 and three regions on SSC7. Eighteen of the 21 candidate regions harbour genes showing differential expression associated with fetal PRRSV infection in a related transcriptomic study, and 7 candidate regions overlap with previously reported QTLs involved in fetal health and host responses to pathogens. Within these regions, we found genes that are involved in a variety of immune processes, including cytokine signalling, leukocyte activities, and innate immunity, and a number of them are functionally linked to known PRRSV-related immune response pathways. The results may provide new evidence to help explain the genetic basis of the fetal response to PRRSV infection and may ultimately lead to alternative control strategies to reduce the impact of reproductive PRRS. However, it should be noted that only one type 2 PRRSV strain was studied; effects would need to be tested with other strains.

3.5 References

Aasted, B., Bach, P., Nielsen, J., and Lind, P. 2002. Cytokine profiles in peripheral blood mononuclear cells and lymph node cells from piglets infected in utero with porcine reproductive and respiratory syndrome virus. Clin. Diagn. Lab. Immunol. **9**: 1229–34.

Agace, W.W. 2006. Tissue-tropic effector T cells: generation and targeting opportunities. Nat. Rev. Immunol. **6**: 682–692.

Akira, S., and Takeda, K. 2004. Toll-like receptor signalling. Nat. Rev. Immunol. **4**: 499–511.

Anderson, S.K., Gallinger, S., Roder, J., Frey, J., Young, H.A., and Ortaldo, J.R. 1993. A cyclophilin-related protein involved in the function of natural killer cells. Proc. Natl. Acad. Sci. U. S. A. **90**: 542–546.

Ayers, K.L., and Cordell, H.J. 2010. SNP selection in genome-wide and candidate gene studies via penalized logistic regression. Genet. Epidemiol. **34**: 879–891.

Bloch, D.B., Nakajima, A., Gulick, T., Chiche, J.D., Orth, D., de La Monte, S.M., and Bloch, K.D. 2000. Sp110 localizes to the PML-Sp100 nuclear body and may function as a nuclear hormone receptor transcriptional coactivator. Mol. Cell. Biol. **20**: 6138–46.

Boddicker, N., Waide, E.H., Rowland, R.R.R., Lunney, J.K., Garrick, D.J., Reecy, J.M., and Dekkers, J.C.M. 2012. Evidence for a major QTL associated with host response to Porcine Reproductive and Respiratory Syndrome Virus challenge. J. Anim. Sci. **90**: 1733–1746.

Boddicker, N.J., Bjorkquist, A., Rowland, R.R., Lunney, J.K., Reecy, J.M., and Dekkers, J.C. 2014. Genome-wide association and genomic prediction for host response to porcine reproductive and respiratory syndrome virus infection. Genet. Sel. Evol. **46**: 18.

Burke, S.M., Issekutz, T.B., Mohan, K., Lee, P.W.K., Shmulevitz, M., and Marshall, J.S. 2008. Human mast cell activation with virus-associated stimuli leads to the selective chemotaxis of natural killer cells by a CXCL8-dependent mechanism. Blood **111**: 5467–5476.

Butler, J.E., Lager, K.M., Golde, W., Faaberg, K.S., Sinkora, M., Loving, C., and Zhang, Y.I. 2014. Porcine reproductive and respiratory syndrome (PRRS): an immune dysregulatory pandemic. Immunol. Res. **59**: 81–108.

Calzada-Nova, G., Schnitzlein, W., Husmann, R., and Zuckermann, F.A. 2010. Characterization of the cytokine and maturation responses of pure populations of porcine plasmacytoid

dendritic cells to porcine viruses and toll-like receptor agonists. Vet. Immunol. Immunopathol. **135**: 20–33.

Campbell, D.J., and Ziegler, S.F. 2007. FOXP3 modifies the phenotypic and functional properties of regulatory T cells. Nat. Rev. Immunol. **7**: 305–310.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience **4**: 7.

Cillari, E., Milano, S., and Pathology, G. 1992. Thymopentin reduces the susceptibility of aged mice to cutaneous leishmaniasis by modulating CD4 T-cell subsets. Immunology **76**: 362–366.

Crosslin, D.R., McDavid, A., Weston, N., Zheng, X., Hart, E., de Andrade, M., Kullo, I.J., McCarty, C.A., Doheny, K.F., Pugh, E., Kho, A., Hayes, M.G., Ritchie, M.D., Saip, A., Crawford, D.C., Crane, P.K., Newton, K., Carrell, D.S., Gallego, C.J., Nalls, M.A., Li, R., Mirel, D.B., Crenshaw, A., Couper, D.J., Tanaka, T., van Rooij, F.J.A., Chen, M.-H., Smith, A. V., Zakai, N.A., Yango, Q., Garcia, M., Liu, Y., Lumley, T., Folsom, A.R., Reiner, A.P., Felix, J.F., Dehghan, A., Wilson, J.G., Bis, J.C., Fox, C.S., Glazer, N.L., Cupples, L.A., Coresh, J., Eiriksdottir, G., Gudnason, V., Bandinelli, S., Frayling, T.M., Chakravarti, A., van Duijn, C.M., Melzer, D., Levy, D., Boerwinkle, E., Singleton, A.B., Hernandez, D.G., Longo, D.L., Witteman, J.C.M., Psaty, B.M., Ferrucci, L., Harris, T.B., O'Donnell, C.J., Ganesh, S.K., Larson, E.B., Carlson, C.S., and Jarvik, G.P. 2013. Genetic variation associated with circulating monocyte count in the eMERGE Network. Hum. Mol. Genet. **22**: 2119–2127.

Dhanasekaran, D.N., and Reddy, E.P. 2008. JNK signaling in apoptosis. Oncogene **27**: 6245–6251.

Duan, X., Nauwynck, H.J., and Pensaert, M.B. 1997. Effects of origin and state of differentiation and activation of monocytes/macrophages on their susceptibility to porcine reproductive and

respiratory syndrome virus (PRRSV). Arch. Virol. **142**: 2483–2497.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics **21**: 3439–3440.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. **4**: 1184–1191.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome J. **4**: 250.

Faaberg, K.S., Balasuriya, U.B., Brinton, M.A., Gorbalenya, A.E., Leung, F..-C., Nauwynck, H., Snijder, E.J., Stadejek, T., Yang, H., and Yoo, D. 2012. Arteriviridae. *in* A.M.Q. King, E. Lefkowitz, M.J. Adams, and E.B. Carstens, eds. Virus Taxonomy. Elsevier.

Fernando, R.L., and Garrick, D. 2013. Bayesian Methods Applied to GWAS. *in* C. Gondro, J. van der Werf, and B. Hayes, eds. Genome-wide association studies and genomic prediction. Humana Press, Totowa, NJ.

Ford, L.B., Cerovic, V., Milling, S.W.F., Graham, G.J., Hansell, C.A.H., and Nibbs, R.J.B. 2014. Characterization of conventional and atypical receptors for the chemokine CCL2 on mouse leukocytes. J. Immunol. **193**: 400–411.

Frey, J.L., Bino, T., Kantor, R.R., Segal, D.M., Giardina, S.L., Roder, J., Anderson, S., and Ortaldo, J.R. 1991. Mechanism of target cell recognition by natural killer cells: characterization of a novel triggering molecule restricted to CD3- large granular lymphocytes. J. Exp. Med. **174**: 1527–1536.

Frias-Staheli, N., Giannakopoulos, N. V, Kikkert, M., Taylor, S.L., Bridgen, A., Paragas, J., Richt,

J.A., Rowland, R.R., Schmaljohn, C.S., Lenschow, D.J., Snijder, E.J., García-Sastre, A., and Virgin, H.W. 2007. Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. Cell Host Microbe **2**: 404–16.

Friedman, J., Hastie, T., and Tibshirani, R. 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. **33**: 124–31.

Galina-Pantoja, L., Siggens, K., van Schriek, M.G.M., and Heuven, H.C.M. 2009. Mapping markers linked to porcine salmonellosis susceptibility. Anim. Genet. **40**: 795–803.

Gately, M.K., Renzetti, L.M., Magram, J., Stern, A.S., Adorini, L., Gubler, U., and Presky, D.H. 1998. The interleukin-12/interleukin-12-receptor system: role in normal and pathologic immune responses. Annu. Rev. Immunol. **16**: 495–521.

Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A.F., Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R.C., Clee, C., Crooijmans, R.P.M.A., Dawson, H.D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G.J., Fowler, K.E., Fredholm, M., Fritz, E., Gilbert, J.G.R., Giuffra, E., Gorodkin, J., Griffin, D.K., Harrow, J.L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S.J., Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H.A., Li, Y., Liu, W., Loveland, J.E., Lu, Y., Lunney, J.K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M.P., Narayan,

J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J.M., Rodriguez-Zas, S., Rohrer, G.A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J.G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B.R., Sperber, G., Stadler, P., Sweedler, J. V., Tafer, H., Thomsen, B., Wali, R., Wang, J., Wang, J., White, S., Xu, X., Yerle, M., Zhang, G., Zhang, J., Zhang, J., Zhao, S., Rogers, J., Churcher, C., and Schook, L.B. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature **491**: 393–398. Nature Publishing Group.

Hayes, B. 2013. Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). *in* C.Gondro; J. van der Werf; B. Hayes, ed. Genome-wide association studies and genomic prediction. Humana Press, Totowa, NJ.

Hiscox, J.A. 2002. The nucleolus – a gateway to viral infection? Arch. Virol. **147**: 1077–1089.

Holtkamp, D.J., Kliebenstein, J.B., Neumann, E.J., Zimmerman, J.J., Rotto, H.F., Yoder, T.K., Wang, C., Yeske, P.E., Mowrer, C.L., and Haley, C. a. 2013. Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. J. Swine Heal. Prod. **21**: 72–84.

Hou, J., Wang, L., Quan, R., Fu, Y., Zhang, H., and Feng, W. 2012. Induction of interleukin-10 is dependent on p38 mitogen-activated protein kinase pathway in macrophages infected with porcine reproductive and respiratory syndrome virus. Virol. J. **9**: 165.

Hu, Z.-L., Park, C.A., Wu, X.-L., and Reecy, J.M. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. **41**: D871–D879.

Hume, D.A., Pavli, P., Donahue, R.E., and Fidler, I.J. 1988. The effect of human recombinant

macrophage colony-stimulating factor (CSF-1) on the murine mononuclear phagocyte system in vivo. J. Immunol. **141**: 3405–9.

Huo, Y., Fan, L., Yin, S., Dong, Y., Guo, X., Yang, H., and Hu, H. 2013. Involvement of unfolded protein response, p53 and Akt in modulation of porcine reproductive and respiratory syndrome virus-mediated JNK activation. Virology **444**: 233–240.

Islam, Z.U., Bishop, S.C., Savill, N.J., Rowland, R.R.R., Lunney, J.K., Trible, B., and Doeschl-Wilson, A.B. 2013. Quantitative analysis of porcine reproductive and respiratory syndrome (PRRS) viremia profiles from experimental infection: a statistical modelling approach. PLoS One **8**: e83567.

Karniychuk, U.U., and Nauwynck, H.J. 2013. Pathogenesis and prevention of placental and transplacental porcine reproductive and respiratory syndrome virus infection. Vet. Res. **44**: 95.

Karniychuk, U.U., Saha, D., Geldhof, M., Vanhee, M., Cornillie, P., Van den Broeck, W., and Nauwynck, H.J. 2011. Porcine reproductive and respiratory syndrome virus (PRRSV) causes apoptosis during its replication in fetal implantation sites. Microb. Pathog. **51**: 194–202.

Karniychuk, U.U., Saha, D., Vanhee, M., Geldhof, M., Cornillie, P., Caij, A.B., De Regge, N., and Nauwynck, H.J. 2012. Impact of a novel inactivated PRRS virus vaccine on virus replication and virus-induced pathology in fetal implantation sites and fetuses upon challenge. Theriogenology **78**: 1527–1537.

Kennedy, D., Jäger, R., Mosser, D.D., and Samali, A. 2014. Regulation of apoptosis by heat shock proteins. IUBMB Life **66**: 327–338.

Ladinig, A., Ashley, C., Detmer, S.E., Wilkinson, J.M., Lunney, J.K., Plastow, G., and Harding, J.C. 2015. Maternal and fetal predictors of fetal viral load and death in third trimester, type 2

porcine reproductive and respiratory syndrome virus infected pregnant gilts. Vet. Res. **46**: 107.

Ladinig, A., Foxcroft, G., Ashley, C., Lunney, J.K., Plastow, G., and Harding, J.C.S. 2014a. Birth weight, intrauterine growth retardation and fetal susceptibility to porcine reproductive and respiratory syndrome virus. PLoS One **9**: e109541.

Ladinig, A., Gerner, W., Saalmüller, A., Lunney, J.K., Ashley, C., and Harding, J.C. 2014b. Changes in leukocyte subsets of pregnant gilts experimentally infected with porcine reproductive and respiratory syndrome virus and relationships with viral load and fetal outcome. Vet. Res. **45**: 128.

Ladinig, A., Lunney, J.K., Souza, C.J., Ashley, C., Plastow, G., Harding, J.C., Gerner, W., Saalmüller, A., Lunney, J.K., Ashley, C., and Harding, J.C. 2014c. Cytokine profiles in pregnant gilts experimentally infected with porcine reproductive and respiratory syndrome virus and relationships with viral load and fetal outcome. Vet. Res. **45**: 113.

Ladinig, A., Wilkinson, J., Ashley, C., Detmer, S.E., Lunney, J.K., Plastow, G., and Harding, J.C.S. 2014d. Variation in fetal outcome, viral load and ORF5 sequence mutations in a large scale study of phenotypic responses to late gestation exposure to type 2 porcine reproductive and respiratory syndrome virus. PLoS One **9**: e96104.

Lee, K.M., Danuser, R., Stein, J. V, Graham, D., Nibbs, R.J., and Graham, G.J. 2014. The chemokine receptors ACKR2 and CCR2 reciprocally regulate lymphatic vessel density. EMBO J. **33**: 2564–2580.

Lu, X., Gong, Y.F., Liu, J.F., Wang, Z.P., Hu, F., Qiu, X.T., Luo, Y.R., and Zhang, Q. 2011a. Mapping quantitative trait loci for cytokines in the pig. Anim. Genet. **42**: 1–5.

Lu, X., Liu, J.-F., Gong, Y.-F., Wang, Z.-P., Liu, Y., and Zhang, Q. 2011b. Mapping quantitative

trait loci for T lymphocyte subpopulations in peripheral blood in swine. BMC Genet. **12**: 79.

Lunney, J.K., and Chen, H. 2010. Genetic control of host resistance to porcine reproductive and respiratory syndrome virus (PRRSV) infection. Virus Res. **154**: 161–169.

Lunney, J.K., Fritz, E.R., Reecy, J.M., Kuhar, D., Prucnal, E., Molina, R., Christopher-Hennings, J., Zimmerman, J., and Rowland, R.R.R. 2010. Interleukin-8, interleukin-1β, and interferon-γ levels are linked to PRRS virus clearance. Viral Immunol. **23**: 127–134.

Lutz, M.A., and Correll, P.H. 2003. Activation of CR3-mediated phagocytosis by MSP requires the RON receptor, tyrosine kinase activity, phosphatidylinositol 3-kinase, and protein kinase C zeta. J. Leukoc. Biol. **73**: 802–14.

Madani, N., Millette, R., Platt, E.J., Marin, M., Kozak, S.L., Bloch, D.B., and Kabat, D. 2002. Implication of the Lymphocyte-Specific Nuclear Body Protein Sp140 in an Innate Response to Human Immunodeficiency Virus Type 1. J. Virol. **76**: 11133–11138.

Manickam, C., Dwivedi, V., Patterson, R., Papenfuss, T., and Renukaradhya, G.J. 2013. Porcine reproductive and respiratory syndrome virus induces pronounced immune modulatory responses at mucosal tissues in the parental vaccine strain VR2332 infected pigs. Vet. Microbiol. **162**: 68–77.

Morrison, A.C., Wilson, C.B., Ray, M., and Correll, P.H. 2004. Macrophage-stimulating protein, the ligand for the stem cell-derived tyrosine kinase/RON receptor tyrosine kinase, inhibits IL-12 production by primary peritoneal macrophages stimulated with IFN-gamma and lipopolysaccharide. J. Immunol. **172**: 1825–32.

Neumann, E.J., Kliebenstein, J.B., Johnson, C.D., Mabry, J.W., Bush, E.J., Seitzinger, A.H., Green, A.L., and Zimmerman, J.J. 2005. Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States. J. Am. Vet.

Med. Assoc. **227**: 385–92.

Nielsen, J., Bøtner, A., Tingstedt, J.-E., Aasted, B., Johnsen, C.., Riber, U., and Lind, P. 2003. In utero infection with porcine reproductive and respiratory syndrome virus modulates leukocyte subpopulations in peripheral blood and bronchoalveolar fluid of surviving piglets. Vet. Immunol. Immunopathol. **93**: 135–151.

Onteru, S.K., Fan, B., Du, Z.Q., Garrick, D.J., Stalder, K.J., and Rothschild, M.F. 2012. A whole-genome association study for pig reproductive traits. Anim. Genet. **43**: 18–26.

Paul, C., Simon, S., Gibert, B., Virot, S., Manero, F., and Arrigo, A.-P. 2010. Dynamic processes that reflect anti-apoptotic strategies set up by HspB1 (Hsp27). Exp. Cell Res. **316**: 1535–1552.

Peracino, B., Wagner, C., Balest, A., Balbo, A., Pergolizzi, B., Noegel, A.A., Steinert, M., and Bozzaro, S. 2006. Function and mechanism of action of Dictyostelium Nramp1 (Slc11a1) in bacterial infection. Traffic **7**: 22–38.

Phatsara, C., Jennen, D.G.J., Ponsuksili, S., Murani, E., Tesfaye, D., Schellander, K., and Wimmers, K. 2007. Molecular genetic analysis of porcine mannose-binding lectin genes, MBL1 and MBL2, and their association with complement activity. Int. J. Immunogenet. **34**: 55–63.

Pixley, F.J., and Stanley, E.R. 2004. CSF-1 regulation of the wandering macrophage: complexity in action. Trends Cell Biol. **14**: 628–638.

Rigby, R.E., Webb, L.M., Mackenzie, K.J., Li, Y., Leitch, A., Reijns, M. a M., Lundie, R.J., Revuelta, A., Davidson, D.J., Diebold, S., Modis, Y., MacDonald, A.S., and Jackson, A.P. 2014. RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. EMBO J. **33**: 542–558.

Rossow, K.D. 1998. Porcine reproductive and respiratory syndrome. Vet. Pathol. **35**: 1–20.

Rothschild, M.F., Hu, Z.L., and Jiang, Z. 2007. Advances in QTL mapping in pigs. Int. J. Biol. Sci. **3**: 192–197.

Schmid, S., Sachs, D., and TenOever, B.R. 2014. Mitogen-activated protein kinase-mediated licensing of interferon regulatory factor 3/7 reinforces the cell response to virus. J. Biol. Chem. **289**: 299–311.

Serão, N.V.L., Kemp, R.A., Mote, B.E., Harding, J.C.S., Willson, P., Bishop, S.C., Plastow, G.S., and Dekkers, J.C.M. 2014. Whole-genome scan and validation of regions previously associated with PRRS antibody response and growth rate using gilts under health challenge in commercial settings. *in* Proceedings of 10th World Congress of Genetics Applied to Livestock Production. ASAS, Vancouver, British Columbia, Canada.

Serao, N.V.L., Matika, O., Kemp, R.A., Harding, J.C.S., Bishop, S.C., Plastow, G.S., and Dekkers, J.C.M. 2014. Genetic analysis of reproductive traits and antibody response in a PRRS outbreak herd. J. Anim. Sci. **92**: 2905–2921.

Sun, Y., Han, M., Kim, C., Calvert, J.G., and Yoo, D. 2012. Interplay between Interferon-Mediated Innate Immunity and porcine reproductive and respiratory syndrome virus. Viruses **4**: 424–446.

Usai, M.G., Goddard, M.E., and Hayes, B.J. 2009. LASSO with cross-validation for genomic selection. Genet. Res. (Camb). **91**: 427–436.

Waldmann, T.A., and Tagaya, Y. 1999. The multifaceted regulation of interleukin-15 expression and the role of this cytokine in NK cell differentiation and host response to intracellular pathogens. Annu. Rev. Immunol. **17**: 19–49.

Wang, X., Zhang, H., Abel, A.M., Young, A.J., Xie, L., and Xie, Z. 2014. Role of

phosphatidylinositol 3-kinase (PI3K) and Akt1 kinase in porcine reproductive and respiratory syndrome virus (PRRSV) replication. Arch. Virol. **159**: 2091–2096.

Wilkinson, J.M., Bao, H., Ladinig, A., Hong, L., Stothard, P., Lunney, J.K., Plastow, G.S., and Harding, J.C.S. 2016. Genome-wide analysis of the transcriptional response to porcine reproductive and respiratory syndrome virus infection at the maternal/fetal interface and in the fetus. BMC Genomics **17**: 383.

Wu, T.T., Chen, Y.F., Hastie, T., Sobel, E., and Lange, K. 2009. Genome-wide association analysis by lasso penalized logistic regression. Bioinformatics **25**: 714–721.

Xu, S., Zhu, D., and Zhang, Q. 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. **111**: 12456–12461.

Yan, N., Regalado-Magdos, A.D., Stiggelbout, B., Lee-Kirsch, M.A., and Lieberman, J. 2010. The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. Nat. Immunol. **11**: 1005–1013.

Yao, H.-P., Zhou, Y.-Q., Zhang, R., and Wang, M.-H. 2013. MSP–RON signalling in cancer: pathogenesis and therapeutic potential. Nat. Rev. Cancer **13**: 466–481. Nature Publishing Group.

Yin, S., Huo, Y., Dong, Y., Fan, L., Yang, H., Wang, L., Ning, Y., and Hu, H. 2012. Activation of c-Jun NH(2)-terminal kinase is required for porcine reproductive and respiratory syndrome virus-induced apoptosis but not for virus replication. Virus Res. **166**: 103–108.

Table 3.1 Summary of fetal viral load measured in fetal thymus (VLT) and fetal viral load measured in endometrium (VLE) for genotyped fetuses

The PRRS virus concentration (target RNA copies / mg) has been log-transformed (base 10). Zeros were given to negative records.

Trait	No. of observations	Mean	Standard Deviation	Minimum value	Maximum value
VLT	817	3.06	3.20	0	8.80
VLE	817	3.35	2.35	0	7.61

Table 3.2 Top SNPs associated with viral load in fetal thymus (VLT)

The locations of SNPs were based on Illumina Pig 60k SNPs mapped to pig genome assembly 10.2 (http://www.animalgenome.org/repository/pig/). Empirical P-values were calculated with a 1,000-run permutation analysis. $\sigma^2_{g,marker}$ means the genetic variance explained by the single marker, and $\sigma^2_{g,total}$ means the total genetic variance explained by all 45,255 SNPs.

Marker ID	**Location**		***P***	$\boldsymbol{\sigma^2_{g,marker}/\ \sigma^2_{g,total}}$ **(%)**
	Chromosome	**Position (bp)**		
ASGA0005344	1	197,479,988	0.021	5.58
ASGA0055300	12	59,921,056	0.069	3.45
DIAS0000654	14	85,991,839	0.044	4.16

Table 3.3 Top SNPs associated with fetal death (FD)

The locations of SNPs were based on Illumina Pig 60k SNPs mapped to pig genome assembly 10.2 (http://www.animalgenome.org/repository/pig/). Empirical *P*-values were calculated with a 1,000-run permutation analysis. $\sigma^2_{g,marker}$ means the genetic variance explained by the single marker, and $\sigma^2_{g,total}$ means the total genetic variance explained by all 45,255 SNPs.

Marker ID	Location		*P*	$\sigma^2_{g,marker}/\sigma^2_{g,total}$ (%)
	Chromosome	Position (bp)		
MARC0003250	1	144,155,620	0.005	2.33
ASGA0021980	4	119,682,308	0.073	0.98
MARC0042986	5	18,071,917	< 0.001	4.35
ASGA0026553	5	87,726,880	0.002	2.60
DRGA0007745	7	71,531,707	0.012	1.80
ASGA0035226	7	97,430,037	< 0.001	3.36
ALGA0053793	9	78,113,024	0.048	1.16
ALGA0061607	11	26,560,250	0.036	1.32
MARC0089129	11	77,708,385	0.026	1.53
M1GA0017106	12	59,174,296	0.027	1.49
MARC0077450	12	61,568,245	0.074	0.97
ALGA0069106	13	28,741,495	0.005	2.29
ASGA0057175	13	36,517,366	0.033	1.36
ALGA0070448	13	63,937,862	0.002	2.60
ALGA0070951	13	82,891,929	< 0.001	3.97

M1GA0018773	14	65,422,608	0.018	1.66
MARC0055746	15	132,949,028	0.065	1.04
ALGA0087932	15	144,250,933	0.015	1.73
MARC0076503	-	-	0.047	1.19
ALGA0032154	-	-	0.067	1.02

Table 3.4 Top SNPs associated with fetal viability (FV)

The locations of SNPs were based on Illumina Pig 60k SNPs mapped to pig genome assembly 10.2 (http://www.animalgenome.org/repository/pig/). Empirical *P*-values were calculated with a 1,000-run permutation analysis. $\sigma^2_{g,marker}$ means the genetic variance explained by the single marker, and $\sigma^2_{g,total}$ means the total genetic variance explained by all 45,255 SNPs.

Marker ID	**Location**		***P***	**$\sigma^2_{g,marker}/\ \sigma^2_{g,total}$ (%)**
	Chromosome	**Position (bp)**		
DRGA0008048	7	109,279,352	0.018	34.6

Table 3.5 QTLs overlapping with candidate regions

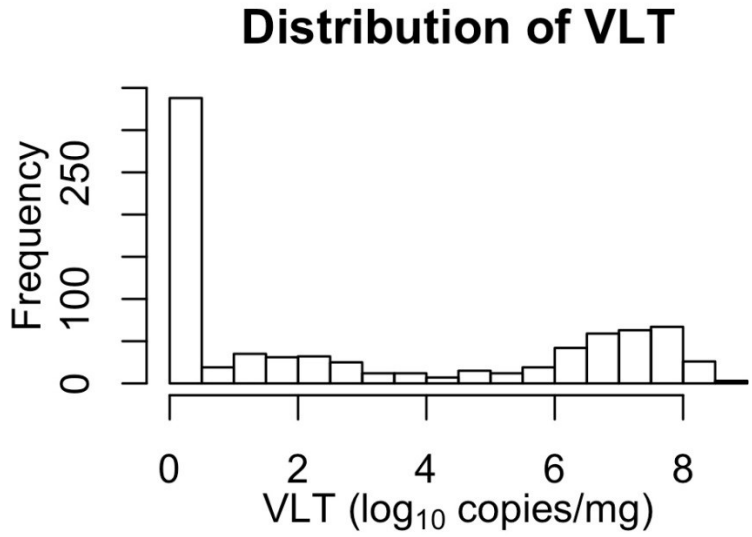
The QTL information was based on pigQTLdb (Hu et al. 2013). Only QTLs spanning no more than 30Mbp were listed in the table, as those QTLs spanning a larger region provide limited support. Associated traits include viral load measured in fetal thymus (VLT) and fetal death (FD).

SNP ID (Associated Trait)	**QTL Information**		
	Chromosome	**Span (bp)**	**Trait**
ASGA0021980 (FD)	4	119,299,797 -120,509,810	CD4-positive, CD8-negative leukocyte percentage; CD4-positive leukocyte percentage
ASGA0026553 (FD)	5	85,756,451 - 90,130,141	Interferon-gamma level
MARC0089129 (FD)	11	73,438,559 - 78,480,320	CD4-positive leukocyte percentage
ASGA0057175 (FD)	13	36,281,164 - 36,541,751	Mummified pigs
DIAS0000654 (VLT)	14	83,873,248 - 93,372,131	C3c concentration
M1GA0018773 (FD)	14	53,465,935 - 81,745,465	Salmonella count in liver and spleen; Salmonella count in liver
MARC0055746 (FD)	15	135,157,314 -149,797,711	Toll-like receptor 2 level

Table 3.6 Genes potentially linked to PRRSV infection

Related Biological Processes	Genes
RIG-I pathway	*UBA7*
Monocyte/macrophage lineage cells activity	*ACKR2, CSF1, MST1, MST1R*
Natural killer (NK) cells activity	*CXCR2, NKTR*
T-cells activity	*ITGB7, TMPO*
IRF3/7 signaling	*SP100, SP110, SP140*
Apoptosis and JNK signaling	*MAP2K4, MAP3K12, MAPKAPK3*
Response to secondary infection	*TLR9, TREX1, SLC11A1*

Figure 3.1 Distribution of viral load in fetal thymus (VLT) and endometrium (VLE) of live fetuses. Concentration of NVSL 97-7985 PRRSV RNA (log 10) per mg tissue measured by in-house quantitative real-time PCR.



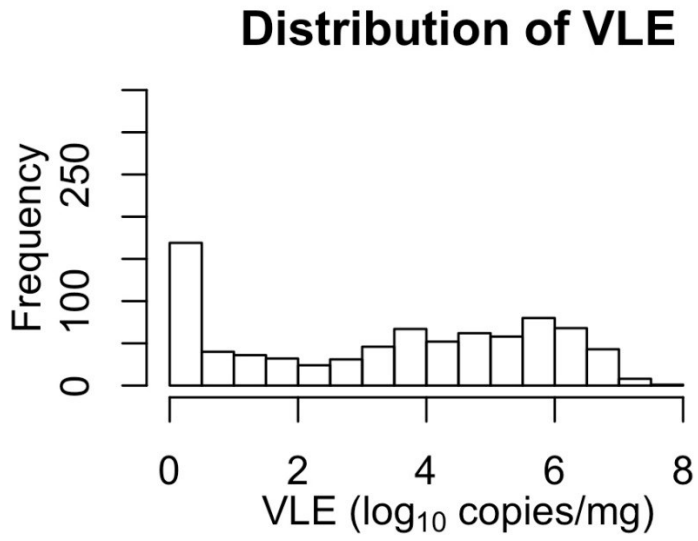


Figure 3.2 Population structure of the fetuses.

The genetic distance was calculated using the genotype of 45,255 SNPs with Hamming distance. The plot was built with 2-D multi-dimension scaling (MDS), and presents the top 2 dimensions (C1 and C2) of the population structure. Colors of the points represent fetal preservation status: green for viable (VIA), red for meconium-stained (MEC), and black for decomposed (DEC). No substantial stratification was evident.

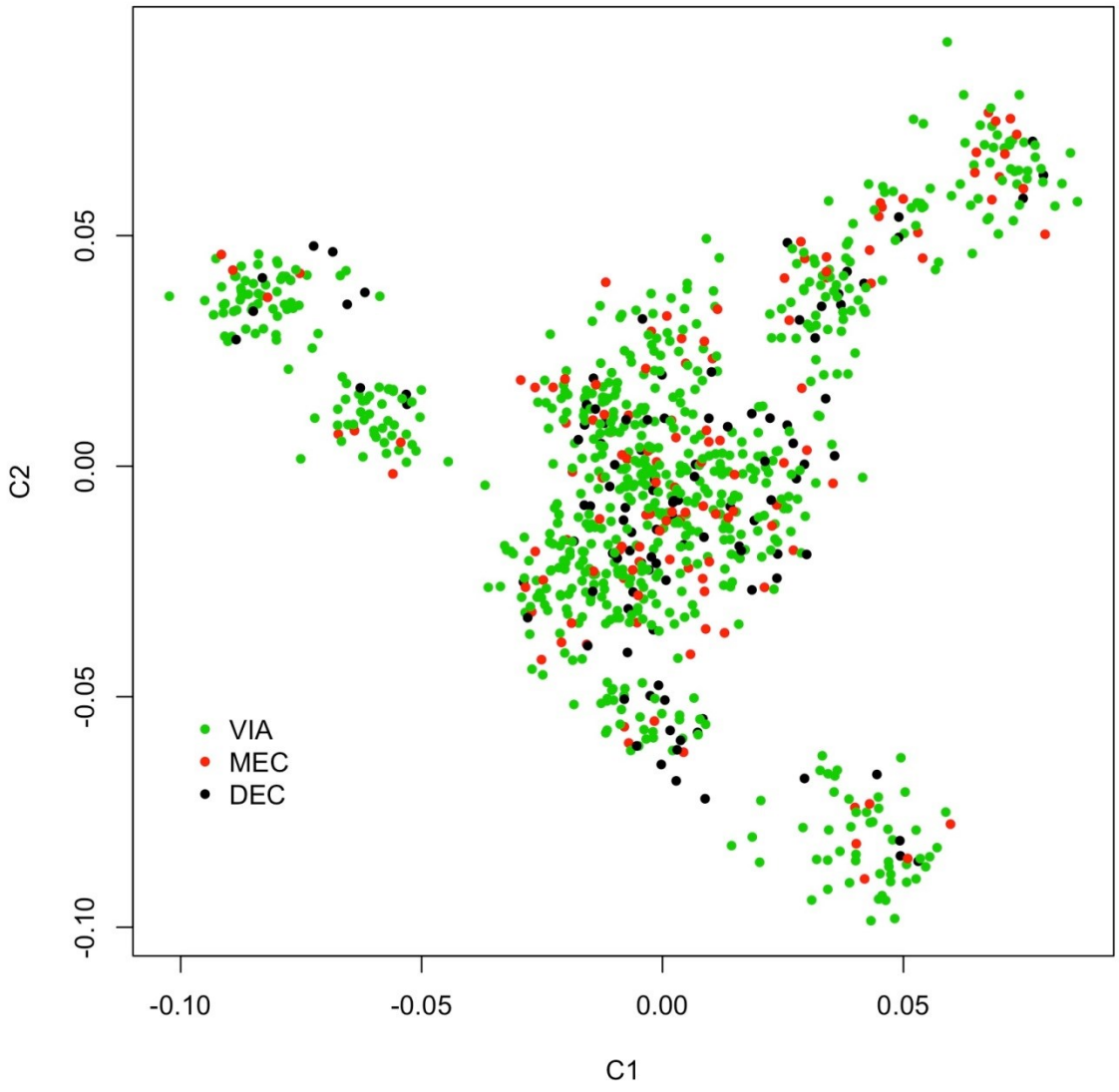


Figure 3.3 Manhattan plot for viral load in fetal thymus (VLT), viral load in fetal endometrium (VLE), fetal death (FD) and fetal viability (FV)

The association analysis was conducted with the least absolute shrinkage and selection operator (LASSO). A generalized linear model was used. The Y-axes shows the percentage of total genetic variance (calculated as the variance of GEBV calculated using 45,255 SNPs) that can be explained by each single SNP. Thresholds corresponding to different empirical *P*-values were calculated with a 1,000-run permutation analysis.

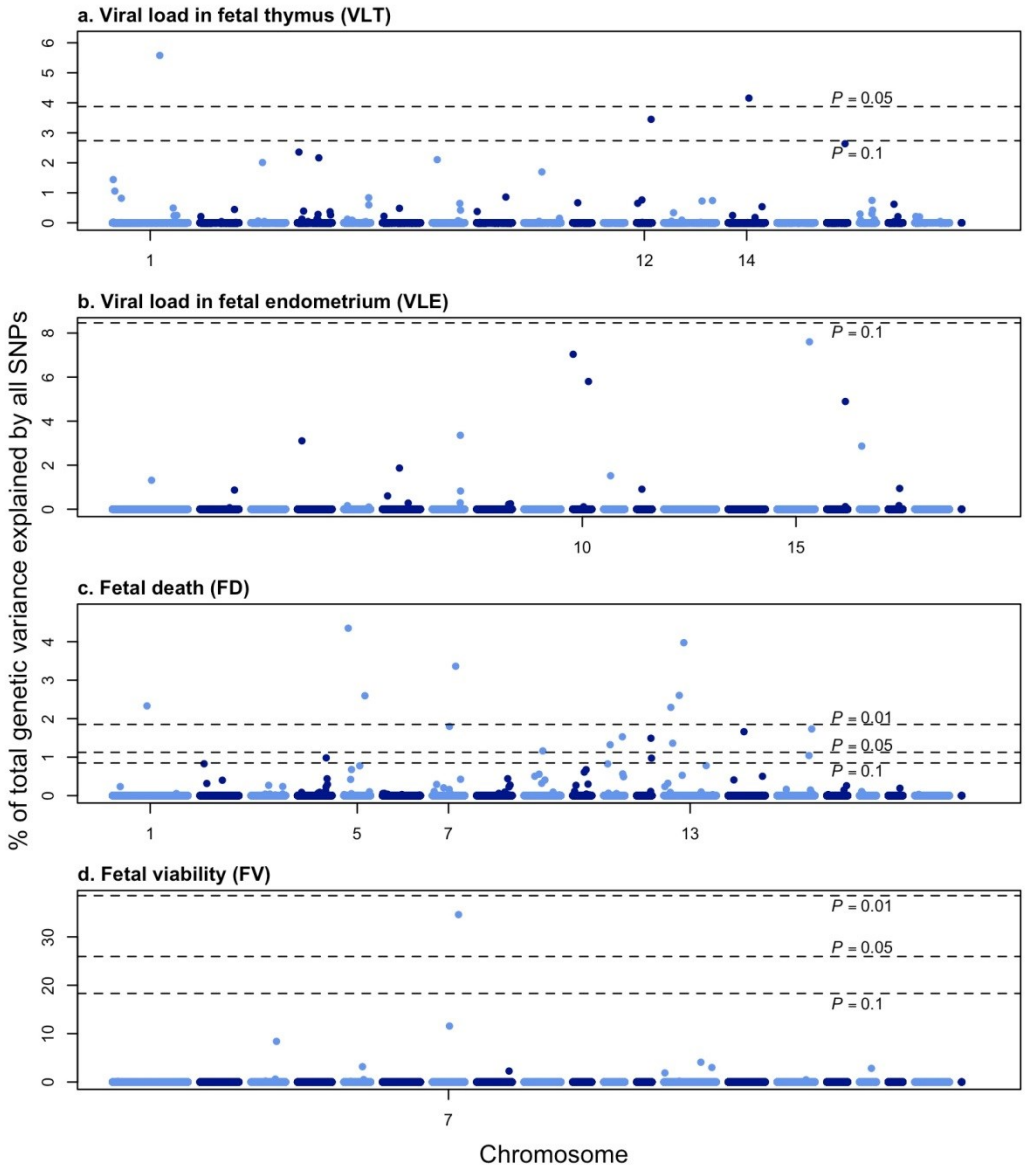


Figure 3.4 Least-square (LS) means of viral load in fetal thymus (VLT) for individuals with different genotypes for each of three SNPs showing significant ($P < 0.05$) or suggestive association ($P < 0.1$) with VLT.

The error bars represent the 95% confidence interval. The numbers above the X-axis represent the total number of animals with that genotype.

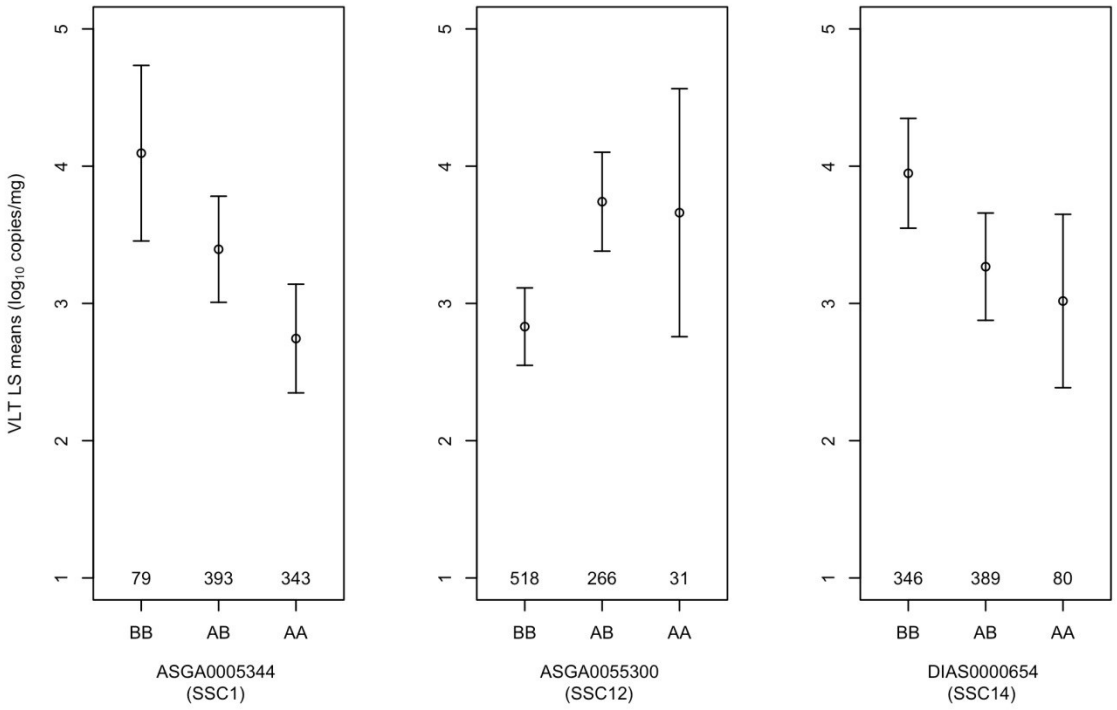


Figure 3.5 Distribution of viral load in fetal thymus (VLT) grouped by number of favoured alleles present in live fetuses.

For each of the three SNPs showing significant ($P < 0.05$) or suggestive ($P < 0.1$) association with VLT, a favoured allele was determined. For each individual fetus, the total number of favoured alleles across the three SNPs was determined. The Y-axis represents the proportion of the fetuses (number of fetuses in that VL window divided by the number of all the fetuses in the group); X-axis represents PRRSV RNA concentration (logarithm 10 target copies per mg). The distribution shows the trend that a lower proportion of individuals had high VLT when more favoured alleles were present.

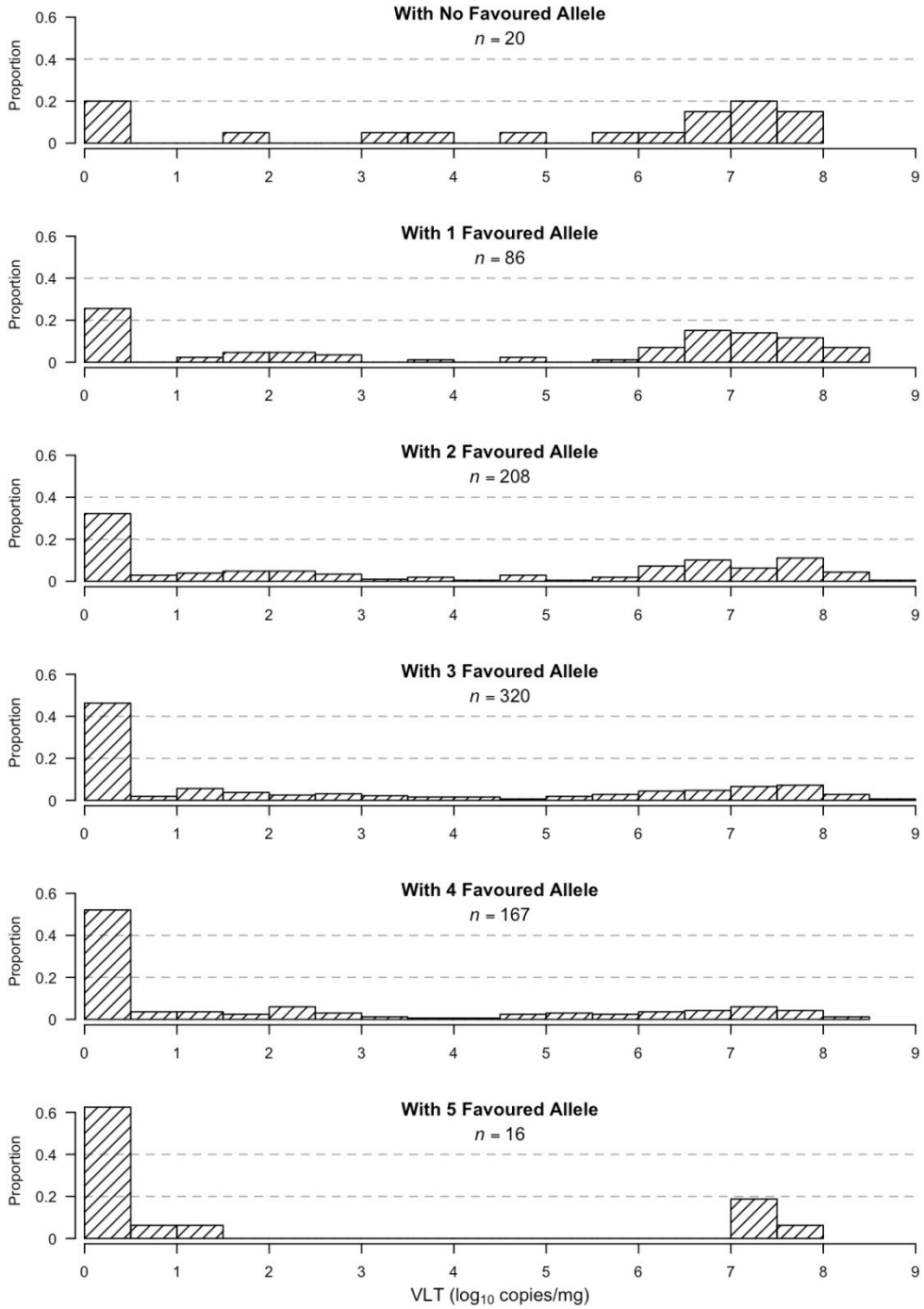
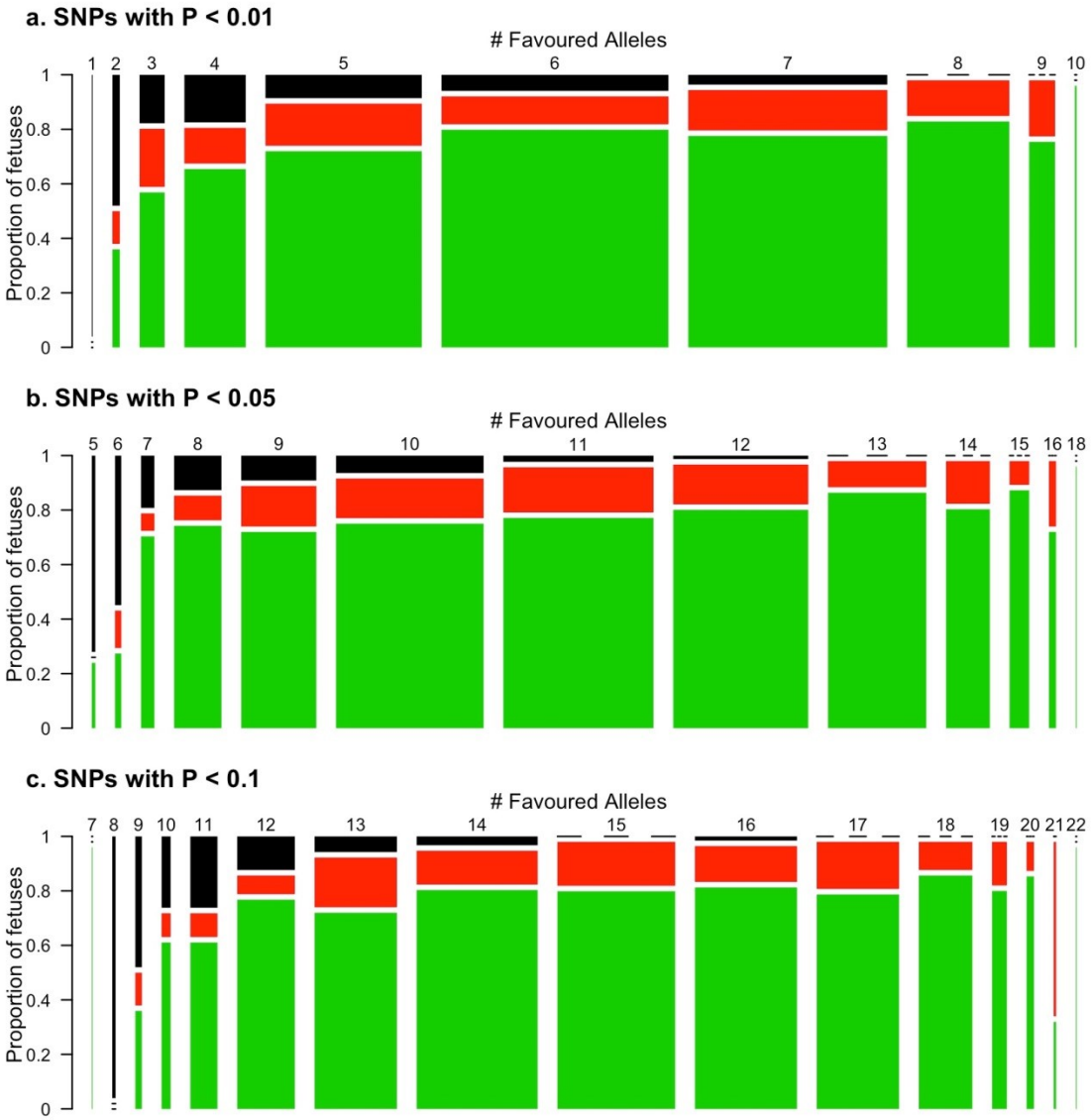

With No Favoured Allele
n = 20
With 1 Favoured Allele
n = 86
With 2 Favoured Allele
n = 208
With 3 Favoured Allele
n = 320
With 4 Favoured Allele
n = 167
With 5 Favoured Allele
n = 16
Proportion
0.6
0.4
0.2
0
0 1 2 3 4 5 6 7 8 9
VLT (log10 copies/mg)

Figure 3.6 Distribution of fetal preservation status grouped by number of favoured alleles.

For each of the SNPs significantly ($P < 0.05$) or suggestively ($P < 0.1$) associated with fetal death (FD), a favoured allele was determined. For each individual fetus, the total number of favoured alleles was determined, and the results displayed for set of SNPs based on their level of significance in the asociation analysis: a) 7 SNPs with $P < 0.01$, b) 15 SNPs with P-values < 0.05 or < 0.01, c) 18 SNPs with P-values < 0.1, < 0.05 or < 0.01. In each plot (A, B, C, $n = 928$), the area of each bar is proportional to the number of fetuses, with preservation status represeted by colour (green for viable [VIA], red for meconium-stained [MEC] and black for decomposed/dead [DEC]). The distribution shows the trend that the proportion of dead fetuses (DEC, black rectangles) decreases when more favoured alleles are present. However, the same trend was evident regardless of the number of SNPs included in the analyses (i.e. plots A, B, C look similar), and the number of favoured alleles did not affect the proportion of MEC fetuses (red rectangles), which remained relatively constant.


a. SNPs with P < 0.01
Favoured Alleles
1 2 3 4 5 6 7 8 9 10
Proportion of fetuses
0 0.2 0.4 0.6 0.8 1
b. SNPs with P < 0.05
Favoured Alleles
5 6 7 8 9 10 11 12 13 14 15 16 18
Proportion of fetuses
0 0.2 0.4 0.6 0.8 1
c. SNPs with P < 0.1
Favoured Alleles
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Proportion of fetuses
0 0.2 0.4 0.6 0.8 1

Chapter 4 A Transmission disequilibrium test of fetal autolysis in porcine reproductive and respiratory syndrome virus challenge

In the genome-wide association study (GWAS) described in Chapter 3, we explored the genetic basis of the variation in fetal response to type 2 PRRSV within non-autolysed fetuses. In this chapter, we describe a genome-wide transmission disequilibrium test (TDT), which aimed to detect candidate genetic variation associated with fetal autolysis. The results were validated subsequently with genotypes from some of the autolyzed fetuses.

I also contributed to the analysis of two related studies (Appendix 3). Along with the contingency-table test described in Section 4.3.6, the two studies composed a minor part of my work in the Ph.D. program, which aimed to detect candidate genomic regions associated with binary traits, using the contingency table test (see Section 1.2.1.4.1) and data from populations with relatively small size.

4.1 Background

A GWAS using the pregnant gilt model (PGM) (Ladinig et al. 2014b) was described in Chapter 3. In the PGM, fetuses with discoloured skin and liquefied internal organs were labelled as autolysed (AUT). It was hypothesized that the AUT fetuses died before the decomposed (DEC) fetuses, and we were interested if there was any genetic factor underlying the early death of these AUT fetuses. They might be even more susceptible to the type 2 PRRSV than the DEC fetuses due to genetics, or just happened to be infected earlier and therefore died earlier. If any associations can be found between fetal autolysis and genetic variants, they may help to confirm some of the detected

genomic regions in Chapter 3. In addition, the new associations and potential functional connection may shed new light on a general genetic basis of the fetal response to type 2 PRRSV.

However, the regression-based method described in Chapter 3 was not suitable for the analysis of fetal autolysis at first, due to the lack of the genomic data of AUT fetuses. For the binary traits (fetal viability and fetal death) in Chapter 3, the generalized linear model and regression require the genomic data from fetuses with each of the two different phenotypes. For fetal viability, we used data from both viable and non-viable fetuses. Similarly, for fetal death, we used data from both dead fetuses and live fetuses. However, this was not feasible for fetal autolysis, as the AUT fetuses did not have genotype records due to poor DNA yield and quality from these fetuses. Available genomic data included genotypes from the challenged gilts, the service sires, and the non-autolysed fetuses. Therefore, a TDT was conducted instead to detect possible candidate genomic regions without the genotypes of AUT fetuses.

The TDT was first introduced in detail in 1993 (Spielman et al. 1993), and has been widely applied in family-based genetic analysis since then. By taking advantage of the genotypes of parents, the TDT is able to detect genetic associations even when data are not available for offspring with one of the two possible phenotypic outcomes (Spielman et al. 1994). The basic idea of the TDT is to compare the observed genotypic frequencies in offspring to their expectations, which can be calculated using the genotype of the parents. If the observed genotypic frequencies do not fit their corresponding expectations (i.e. transmission disequilibrium), it may imply that there is an association between the genotype and the trait (Ott et al. 2011). The original TDT was also extended to other situations, such as sib-TDT (Spielman and Ewens 1998) and sib-disequilibrium test (SDT) (Horvath and Laird 1998), to make better use of different types of family data. A principal limitation of the TDT is that it is not very robust to undetected genotyping errors,

which can be solved to a degree by TDTae, a robust extension of the original TDT (Barral et al. 2005).

Its ability for analysing datasets with only affected individuals gives the TDT an advantage in studies of human disease, where the data of affected individuals are usually more accessible than those from unaffected individuals (Spielman et al. 1994). For example, in the aforementioned TDT study, Spielman (1993) detected association between a genetic marker and diabetes using data from only affected children. In a study about schizophrenia published in 1996, the TDT was applied to a dataset from only 178 patients but no unaffected individuals, and an associated polymorphism was detected (Li et al. 1996). The TDT has also been successfully applied to animal QTL mapping. For example, a generalized TDT was used to test the effect of gene MC4R on production traits in pigs in 2003 (Hernández-Sánchez et al. 2003). Other similar ideas about genotype frequency disequilibrium have also been applied in animal genomics, such as a study about the absence of homozygous haplotypes in cattle in 2012 (VanRaden et al. 2011), where the expected genotype frequencies were calculated based on population data and absence of recessive homozygotes were detected.

In the current study, we conducted a TDT of fetal autolysis and applied raw genotyping data (fluorescence intensity data) in avoiding false positive results. Following this, a new DNA extraction approach resulted in genotypes for 155 of the AUT fetuses, and they were used as a test population to validate the results from the TDT.

4.2 Materials

4.2.1 Phenotypic data

Data used in the current study were also obtained from the PGM described in a previous study (Ladinig et al. 2014b) and Chapter 3. Phenotypic data were collected from the same 1422 fetuses obtained from the PRRSV-inoculated gilts, as described in Chapter 3. In addition to the viable (VIA, n = 697), meconium-stained (MEC, n = 125) and decomposed (DEC, n = 111) fetuses used in the GWAS described in Chapter 3, the AUT fetuses (AUT, n = 459) were also included in the current analysis. All the AUT fetuses were dead at termination, with more than 50% of surface discoloured and liquefied internal organs. Based on the severity of their autolysis, it has been hypothesized that the AUT fetuses were dead for at least 1 week prior to termination, and that they died before DEC fetuses (Ladinig et al. 2014a).

Pedigree information was available for the fetus population. In total, 873 non-autolysed fetuses from 105 litters had records of both of their parents, and 155 autolysed fetuses from 74 litters had records of both of their parents.

4.2.2 Genomic data

In addition to the genotype data from the 979 non-autolysed fetuses (VIA, MEC or DEC) used in the GWAS described in Chapter 3, 107 gilts and 24 service sires were also genotyped using the PorcineSNP60 Genotyping BeadChip v2 (Illumina, San Diego, CA, USA) containing 61,565 SNPs. For AUT fetuses, a new procedure for DNA extraction from autolysed samples was proposed by the Genomics and Proteomics Unit in the Department of Agricultural, Food and Nutritional Science, University of Alberta. Although the new procedure was more laborious, essentially using phenol/chloroform DNA extraction, it successfully enabled the genotyping for

some of the AUT fetuses. In total 155 AUT fetuses were genotyped using the Geneseek-Neogen GPPHD 80K SNP chip containing 68,528 SNPs. Based on the rs ID of SNPs, we found 47,874 common SNPs in the two genotype datasets.

4.3 Statistical methods

Methods used in the analysis included 1) the detection of genotyping errors in Mendelian inheritance (MI), 2) the detection of transmission disequilibrium, and 3) validation tests for the detected SNPs using litter information, raw genotyping data (fluorescence intensity data) and available data from AUT fetuses.

4.3.1 Genotyping errors in Mendelian inheritance (MI)

We first scanned the genomic data for possible genotyping errors in Mendelian inheritance. Normally, the alleles for a specific SNP in an individual must be received from its parents, unless *de novo* mutations have happened at that locus. The probability of *de novo* mutations at a given locus is extremely low and can be ignored. Given the genotypes of the parents, the possible genotype of the offspring for this SNP must follow Mendel's law of segregation, otherwise it may be a genotyping error. For each SNP, we calculated the possible genotype of offspring based on the genotype combination of parents for every trio. If the offspring in a trio showed a genotype which is impossible in Mendelian inheritance (Table 4.1), we labelled this trio as "MI error" for the SNP and removed the trio from the following TDT (treated as missing genotype). The "MI error score" was also calculated for each SNP as the proportion of "MI errors" trios in all trios, which was used as one of the measurements for the reliability of the genotype data.

4.3.2 Transmission disequilibrium test (TDT)

The basic idea of TDT in the current study is to scan for transmission disequilibrium within trios of non-autolysed fetuses. For each SNP, we can calculate its expected genotype frequencies in both non-autolysed fetuses and autolysed fetuses based on their parents' genotype. If a SNP has no association with fetal autolysis, the observed genotype frequencies should obey expectation in both fetus populations. However, if one or more genotypes is associated with autolysed fetuses, it would show enrichment in autolysed fetuses and appear less in non-autolysed fetuses. Deviation in genotype frequencies in non-autolysed fetuses was what we were trying to detect in the analysis.

For the 873 trios with non-autolysed fetuses, we calculated the expected genotype frequencies in the fetuses for every SNP as:

$$\begin{bmatrix} p_{exp,AA} \\ p_{exp,AB} \\ p_{exp,BB} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_i p_{exp,AA,i} \\ \frac{1}{n}\sum_i p_{exp,AB,i} \\ \frac{1}{n}\sum_i p_{exp,BB,i} \end{bmatrix},$$

where $p_{exp,AA,i}$, $p_{exp,AB,i}$, $p_{exp,BB,i}$ is the expected genotype frequency in the ith trio, and n is the number of trios with valid genotype for the SNP. The expected genotype frequencies were then compared with the observed genotype in the fetuses using Pearson's chi-squared test:

$$\chi^2 = \sum_{geno} \frac{\left(\frac{O_{geno}}{n} - p_{exp,geno}\right)^2}{p_{exp,geno}}, \qquad (geno = AA, AB \text{ and } BB)$$

where O_{geno} is the observed number of fetus genotypes. The chi-squared statistic measures the degree to which the observed genotype frequency fit the expected value. For most of the SNPs, all three genotypes (AA, AB and BB) were possible in the fetuses, as the expected frequencies for the three genotypes were all non-zero. In this case, the statistic would follow a chi-squared distribution with two degrees of freedom. For the other SNPs, only two of the three genotypes are possible for the fetuses, then the degree of freedom would be set as one. A P-value was calculated for each SNP based on the determined distribution for statistical hypothesis testing. False discovery rates (FDR) were used to adjust the p-values for multiple testing. All SNPs with a FDR lower than 0.1 were selected as candidate SNPs that had transmission distortion.

4.3.3 Candidate SNP test using call frequency and MI error score

As genotyping quality is one of our main concerns, we used the MI error score and call frequency for all candidate SNPs for validation. As described before, MI error score measures how the genotype is obviously wrong in a family-based test, and the error may be introduced in producing raw data or genotype calling. The call frequency is the proportion of all genotyped samples that had a genotype assigned during genotyping, which can serve as one of the measurements of the reliability of genotyping. All candidate SNPs were evaluated based on these two values, and SNPs with relatively low quality were removed from further analysis.

4.3.4 Litter-level test

The detected transmission disequilibrium was checked at a litter-level. In the TDT described above, we detected SNPs showing transmission disequilibrium at the population level (i.e. shows disequilibrium when all litters are combined together). One assumption is that a SNP associated

with fetal autolysis should show transmission disequilibrium in all litters evenly. Extreme transmission disequilibrium in only a few litters is more likely to be a false discovery due to genotyping error of the parents in those litters. For each candidate SNP, we calculated the disequilibrium in each litter, and tested the parents using raw genotyping data as described in the next section.

4.3.5 Candidate SNP validation using raw genotyping data

Raw genotyping data provided another way for us to evaluate the possibility that the detected associations were due to genotyping error, especially clustering error. In the analysis, we plotted the normalized intensity R (the total intensity of both alleles) to the normalized theta (the ratio between the intensity of the two alleles), both of which had been used in the genotyping clustering, and checked if an alternative clustering results could better explain the observation of genotype in fetuses. Based on our prior knowledge, we believe that genotyping error is more likely to happen than "real transmission disequilibrium". In this case, if the raw genotyping data showed the possibility of genotyping error, and the error could well explain the transmission disequilibrium we had observed, we would treat it as "false discovery due to genotyping error".

The possibility of "false discovery due to genotyping error" was assessed using the likelihood function below:

$$L(geno_{parents}) = P(geno_{child} \mid geno_{parents}),$$

which equals the probability of the observed genotype of the fetuses given their parents' genotype. This likelihood function describes how a certain set of parents' genotype can explain the observation in the fetuses' genotype. In our calculation, the likelihood function can be calculated as the production of the probabilities for all possible combination of parents' genotype

$$L(geno_{parents}) = \prod_i P(geno_{child,i}|geno_{parents,i}).$$

By adjusting the threshold of clustering, we were able to evaluate how good an alternative clustering is using likelihood ratio (LR) and log of the odds (LOD) as

$$LOD = \log_{10} LR = \log_{10} \frac{L(geno_{parents})_A}{L(geno_{parents})_0},$$

where LR is the ratio of the likelihood of alternative clustering ($L(geno_{parents})_A$) to that of original clustering ($L(geno_{parents})_0$).

4.3.6 Candidate SNP validation using trios with autolysed fetuses

The candidate SNPs were also tested in trios with autolysed fetuses. For those candidate SNPs found in TDT, we presume that their deviation in genotype frequencies within autolysed fetuses was reflected in the association between their genotype frequencies with fetal autolysis, i.e. the genotype frequencies are significantly different between autolysed and non-autolysed fetuses. By using the genomic data of the 155 autolysed fetuses, we tested the hypothesis and evaluated our result from the first part of the analysis (the TDT).

In addition, as another method of testing association between genomic regions and fetal autolysis, an association study was conducted using a small subpopulation of 280 fetuses. We applied stratified sampling in selecting the animals. i.e., for each autolysed fetus with available genotype record, a viable full-sibling was selected as a control (if there is). In total, 140 viable-autolysed sibling pairs were selected. The matched case-control design made use of the genomic data of available autolysed fetuses, however, the relatively small population size may limit its statistical power. Pearson's chi-squared test were used to detect genomic regions that showed different genotypic frequencies between viable and autolysed fetuses.

4.4 Results and discussion

4.4.1 Transmission disequilibrium test (TDT) – autosomes

On autosomes, 54,846 SNPs were tested for transmission disequilibrium. The result was shown in Figure 4.1 as a Manhattan plot. Using a criterion of FDR less than 0.1, Sixteen SNPs were found significant in the statistical test (Table 4.2). One of the SNPs (MARC0013088) was deemed to be unreliable due to obvious clustering error; four of the SNPs showed low call rate (less than 20%); one of the SNPs showed high MI error rate (higher than 20%); three SNPs showed good call rate but non-zero MI error rate, and the other 7 SNPs shows high call rate without any MI error. These seven SNPs were then tested with raw genotyping data and trios with autolysed fetuses.

4.4.2 Transmission disequilibrium test (TDT) – allosomes

For the X chromosome, 1,557 SNPs were tested for transmission disequilibrium. Fifty SNPs were found to have significant disequilibrium in transmission. The distribution of the call frequency and MI error frequency of these SNPs was shown in Figure 4.2. None of the 50 significant SNPs

showed a good call frequency and MI error score. The highest "valid call rate" is only 49.5%, and all significant SNPs showed MI error to some degree. For the Y chromosome, we found 9 SNPs in the genotyping panel. No further analysis was conducted as none of them have a good call frequency in males (call rate < 5%).

4.4.3 Litter-level test and raw genotyping data analysis

For the 7 SNPs showing high call rate without any MI error, we conducted a litter-level test and checked the raw genotyping data. For 5 of the 7 SNPs, we found the transmission disequilibrium was mainly driven by a few litters, and the raw genotyping data also showed the possibility of clustering error. They are ASGA0076852, MARC0113898, ASGA0097612, MARC0111071, and MARC0038959 (Figure 4.3). For all the five SNPs, only two genotypes appeared in fetuses. In the R-Theta plot of their fluorescent intensity, however, we can find that there seemed to be three clusters for each of these SNPs, as the AB group is formed by two groups of points. The likelihood analysis showed that clustering error can greatly explain the transmission disequilibrium for these five SNPs. For each of them, we tested all possible clustering, and the one that maximized the LOD value was treated as the best "alternative clustering" (Figure 4.3). The maximum LOD values for the five SNPs are listed in Table 4.3. For the other two SNPs (ASGA0004971 and MARC0058459), there was no obvious sign that the transmission disequilibrium was mainly driven by only a few litters. In the R-Theta plot (Figure 4.4), all three genotypes appeared as discrete clusters. There is no evidence that the genotype was biased due to clustering error.

Previous studies has reported that undetected genotyping error may contribute to false positives in TDT (Mitchell et al. 2003). Genotyping errors occurring for the parental samples may lead to incorrect expectation of genotypic frequencies for the offspring, and those errors happening

in offspring may result in incorrect observed genotypic frequencies. In both of the two cases, the goodness of fit between expectation and observation decreases, which causes false transmission disequilibrium. Genotyping errors can be modelled as a random event in some cases (Douglas et al. 2002; Sobel et al. 2002), and those models has been reported to help in the TDT (Gordon et al. 2001, 2004). However, the genotyping errors in the current study (those due to mis-clustering) were not absolutely random, as incorrect genotypes were more likely to have extreme Theta value in the raw genotyping data. We took advantage of this property and detected five candidate SNPs that may be false. The result may also imply the possibility of a model of genotyping error that is based on raw genotyping data, and its application in the TDT.

Another difference between the current study and a typical application of TDT is that the family size in the current study (i.e. litter size) was larger. In this case, the negative effect of genotyping error in parents on the results was also increased, as the genotype of a parent was used to calculate the expected genotype frequencies for offspring in one or more litters. On the other hand, the larger family size provided more information in inferring mis-clustering in the parents, which makes the analysis of raw genotyping data more reliable.

4.4.4 Candidate SNP validation using trios with autolysed fetuses

The two SNPs (ASGA0004971 and MARC0058459) were further tested using trios with autolysed fetuses. The first one (ASGA0004971) did not have any valid call in any of the 155 autolysed fetuses, and it was therefore not possible to confirm its genotype frequency. The same pattern were observed in the autolysed fetuses for the other SNP (MARC0058459), which did not meet our expectation from the TDT. This result, however, may imply the possibility that there was a transmission disequilibrium in all born fetuses (including non-autolysed and autolysed fetuses),

which may be irrelevant to the PRRSV challenge in the current study. This is discussed further below in relation to the genes identified in this region.

Four SNPs showed significant associations in the small subpopulation of 280 fetuses, not including ASGA0004971 and MARC0058459. Evidence were found that the four associations were suspicious as MI were found for all of them. Further investigation may be conducted for these four associations, in the current study, however, they failed to provide support to the results of the TDT.

4.4.5 Functional analysis

We also explored the regions around the two SNPs for previously discovered QTLs or genes that are possibly related to the associated trait. All QTL information were obtained from PigQTLdb (Hu et al. 2013).

Two genes around SNP ASGA0004971 were found to be related to immune response and apoptosis, which can be related to PRRSV infection. One of them is BCL2, which encodes B-cell CLL/lymphoma 2. This protein has been reported to play a role in the regulation of apoptotic cell death, and its homologues has been used by virus (e.g. African swine fever virus) for regulating apoptosis (Neilan et al. 1993; Kennedy et al. 2014). It has been reported that BCL2 is not able to prevent the apoptosis induced by a recombinant vaccinia virus expressing GP5 of PRRSV (Suárez et al. 1996). However, a previous study also showed that PRRSV infection reduced BCL2 expression level (Lee and Kleiboeker 2007). The other gene close to ASGA0004971 is PHLPP1. This gene encodes pleckstrin homology leucine-rich repeat protein phosphatase 1, which also plays a role in apoptosis (Qiao et al. 2010) and immunomodulation (Alamuru et al. 2014).

For SNP MARC0058459, we found two QTLs overlapping with the SNP, with one for reproductive tract weight (Rosendo et al. 2012) and the other one for total number born alive (Mucha et al. 2013),. This region harbors an important gene related to reproductive traits, such as fetal loss, in mammals, which encodes leukemia inhibitory factor (LIF). This factor has been reported to be expressed in blastocysts (Murray et al. 1990), and plays a critical role in implantation of the blastocyst (Stewart 1994). The observed transmission disequilibrium that happened in all born fetuses may be explained by variants within this gene.

4.5 Conclusion

We conducted a TDT of fetal autolysis using data obtained from the PGM. Seven SNPs showed significant transmission disequilibrium, while having a relative good genotyping quality (with a high call rate and a low MI error rate). Raw genotyping data analysis suggested that five of the seven SNPs may be mis-clustered in genotyping the parents, which may result in false transmission disequilibrium for these five SNPs. The other two candidate SNPs showed valid transmission disequilibrium in the current study. Testing using AUT fetuses did not provide further support for the two SNPs. One of them was deemed to be associated with general fetal loss rather than the fetal autolysis due to PRRSV challenge. Genes and previous reported QTLs overlapping with the two candidate regions were found to be related to apoptosis, immune response and reproductive traits.

4.6 Reference

Alamuru, N.P., Behera, S., Butchar, J.P., Tridandapani, S., Kaimal Suraj, S., Babu, P.P., Hasnain, S.E., Ehtesham, N.Z., and Parsa, K.V.L. 2014. A novel immunomodulatory function of

PHLPP1: inhibition of iNOS via attenuation of STAT1 ser727 phosphorylation in mouse macrophages. J. Leukoc. Biol. **95**: 775–783.

Barral, S., Haynes, C., Levenstien, M. a, and Gordon, D. 2005. Precision and type I error rate in the presence of genotype errors and missing parental data: a comparison between the original transmission disequilibrium test (TDT) and TDTae statistics. BMC Genet. **6**: S150.

Douglas, J.A., Skol, A.D., and Boehnke, M. 2002. Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. Am. J. Hum. Genet. **70**: 487–95.

Gordon, D., Haynes, C., Johnnidis, C., Patel, S.B., Bowcock, A.M., and Ott, J. 2004. A transmission disequilibrium test for general pedigrees that is robust to the presence of random genotyping errors and any number of untyped parents. Eur. J. Hum. Genet. **12**: 752–761.

Gordon, D., Heath, S.C., Liu, X., and Ott, J. 2001. A transmission/disequilibrium test that allows for genotyping errors in the analysis of single-nucleotide polymorphism data. Am. J. Hum. Genet. **69**: 371–80.

Hernández-Sánchez, J., Visscher, P., Plastow, G., and Haley, C. 2003. Candidate gene analysis for quantitative traits using the transmission disequilibrium test: the example of the melanocortin 4-receptor in pigs. Genetics **164**: 637–44.

Horvath, S., and Laird, N.M. 1998. A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am. J. Hum. Genet. **63**: 1886–97.

Hu, Z.-L., Park, C.A., Wu, X.-L., and Reecy, J.M. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. **41**: D871–D879.

Kennedy, D., Jäger, R., Mosser, D.D., and Samali, A. 2014. Regulation of apoptosis by heat shock

proteins. IUBMB Life **66**: 327–338.

Ladinig, A., Foxcroft, G., Ashley, C., Lunney, J.K., Plastow, G., and Harding, J.C.S. 2014a. Birth weight, intrauterine growth retardation and fetal susceptibility to porcine reproductive and respiratory syndrome virus. PLoS One **9**: e109541.

Ladinig, A., Wilkinson, J., Ashley, C., Detmer, S.E., Lunney, J.K., Plastow, G., and Harding, J.C.S. 2014b. Variation in fetal outcome, viral load and ORF5 sequence mutations in a large scale study of phenotypic responses to late gestation exposure to type 2 porcine reproductive and respiratory syndrome virus. PLoS One **9**: e96104.

Lee, S.-M., and Kleiboeker, S.B. 2007. Porcine reproductive and respiratory syndrome virus induces apoptosis through a mitochondria-mediated pathway. Virology **365**: 419–34.

Li, T., Sham, P.C., Vallada, H., Xie, T., Tang, X., Murray, R.M., Liu, X., and Collier, D.A. 1996. Preferential transmission of the high activity allele of COMT in schizophrenia. Psychiatr. Genet. **6**: 131–3.

Mitchell, A.A., Cutler, D.J., and Chakravarti, A. 2003. Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. Am. J. Hum. Genet. **72**: 598–610.

Mucha, A., Ropka-Molik, K., Piórkowska, K., Tyra, M., and Oczkowicz, M. 2013. Effect of EGF, AREG and LIF genes polymorphisms on reproductive traits in pigs. Anim. Reprod. Sci. **137**: 88–92.

Murray, R., Lee, F., and Chiu, C.P. 1990. The genes for leukemia inhibitory factor and interleukin-6 are expressed in mouse blastocysts prior to the onset of hemopoiesis. Mol. Cell. Biol. **10**: 4953–6.

Neilan, J.G., Lu, Z., Afonso, C.L., Kutish, G.F., Sussman, M.D., and Rock, D.L. 1993. An African

swine fever virus gene with similarity to the proto-oncogene bcl-2 and the Epstein-Barr virus gene BHRF1. J. Virol. **67**: 4391–4.

Ott, J., Kamatani, Y., and Lathrop, M. 2011. Family-based designs for genome-wide association studies. Nat. Rev. Genet. **12**: 465–474.

Qiao, M., Wang, Y., Xu, X., Lu, J., Dong, Y., Tao, W., Stein, J., Stein, G.S., Iglehart, J.D., Shi, Q., and Pardee, A.B. 2010. Mst1 is an interacting protein that mediates PHLPPs' induced apoptosis. Mol. Cell **38**: 512–23.

Rosendo, A., Iannuccelli, N., Gilbert, H., Riquet, J., Billon, Y., Amigues, Y., Milan, D., and Bidanel, J.P. 2012. Microsatellite mapping of quantitative trait loci affecting female reproductive tract characteristics in Meishan x Large White F2 pigs. J. Anim. Sci. **90**: 37–44.

Sobel, E., Papp, J.C., and Lange, K. 2002. Detection and integration of genotyping errors in statistical genetics. Am. J. Hum. Genet. **70**: 496–508.

Spielman, R.S., and Ewens, W.J. 1998. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am. J. Hum. Genet. **62**: 450–8.

Spielman, R.S., McGinnis, R.E., and Ewens, W.J. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. **52**: 506–16.

Spielman, R.S., McGinnis, R.E., and Ewens, W.J. 1994. The transmission/disequilibrium test detects cosegregation and linkage. Am. J. Hum. Genet. **54**: 559-60–3.

Stewart, C.L. 1994. Leukaemia inhibitory factor and the regulation of pre-implantation development of the mammalian embryo. Mol. Reprod. Dev. **39**: 233–8.

Suárez, P., Díaz-Guerra, M., Prieto, C., Esteban, M., Castro, J.M., Nieto, A., and Ortín, J. 1996. Open reading frame 5 of porcine reproductive and respiratory syndrome virus as a cause of

virus-induced apoptosis. J. Virol. **70**: 2876–82.

VanRaden, P.M., Olson, K.M., Null, D.J., and Hutchison, J.L. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy Sci. **94**: 6153–6161.

Table 4.1 Genotype combinations treated as “MI error” (Autosome)

Genotype Combination of Parents	**“Impossible” Genotype of Offspring**
AA + AA	AB or BB
AA + AB	BB
AA + BB	AA or BB
AB + AB	(None)
AB + BB	AA
BB + BB	AA or AB

Table 4.2 Genotype combinations treated as “MI error” (X chromosome)

Sire’s Genotype	**Dam’s Genotype**	**“Impossible” Genotype of Offspring**	
		Male	female
A	AA	B	AB or BB
A	AB	(None)	BB
A	BB	A	AA or BB
B	AA	B	AA or BB
B	AB	(None)	AA
B	BB	A	AA or AB

Table 4.3 Genotype combinations treated as “MI error” (Y chromosome)

Genotype Combination of Sire	**“Impossible” Genotype of Male Offspring**
A	B
B	A

Table 4.4 Significant SNPs detected on autosomes with call rate and MI error

SNP	Chr	Position	#MI_error	#VALID	Comments
ASGA0004971	1	176,096,880	0	813	No MI error and high call rate
ASGA0009146	2	12,082,068	9	810	Low MI error rate
MARC0111071	2	29,644,443	0	881	No MI error and high call rate
MARC0113898	2	53,317,766	0	826	No MI error and high call rate
ASGA0010636	2	86,601,289	5	888	Low MI error rate
ALGA0116161	2	160,111,983	0	141	Low call rate
MARC0001070	3	32,120,448	0	141	Low call rate
ASGA0097612	6	34,035,458	0	873	No MI error and high call rate
MARC0026980	8	70,937,913	69	67	Low call rate
ASGA0093792	12	26,564,113	200	666	High MI error rate
ASGA0055965	13	11,778,021	112	28	Low call rate
ALGA0070715	13	73,041,071	13	857	Low MI error rate
MARC0013088	13	140,497,748	0	819	Clustering error
MARC0058459	14	52,227,402	0	888	No MI error and high call rate

Table 4.5 The maximum log of the odds (LOD) obtained in raw genotyping data analysis

SNP	Maximum LOD
MARC0113898	52.6
ASGA0076852	23.6
ASGA0097612	12.5
MARC0111071	2.13
MARC0038959	19.8

Figure 4.1 Manhattan plot for the transmission disequilibrium test (TDT) on autosomes

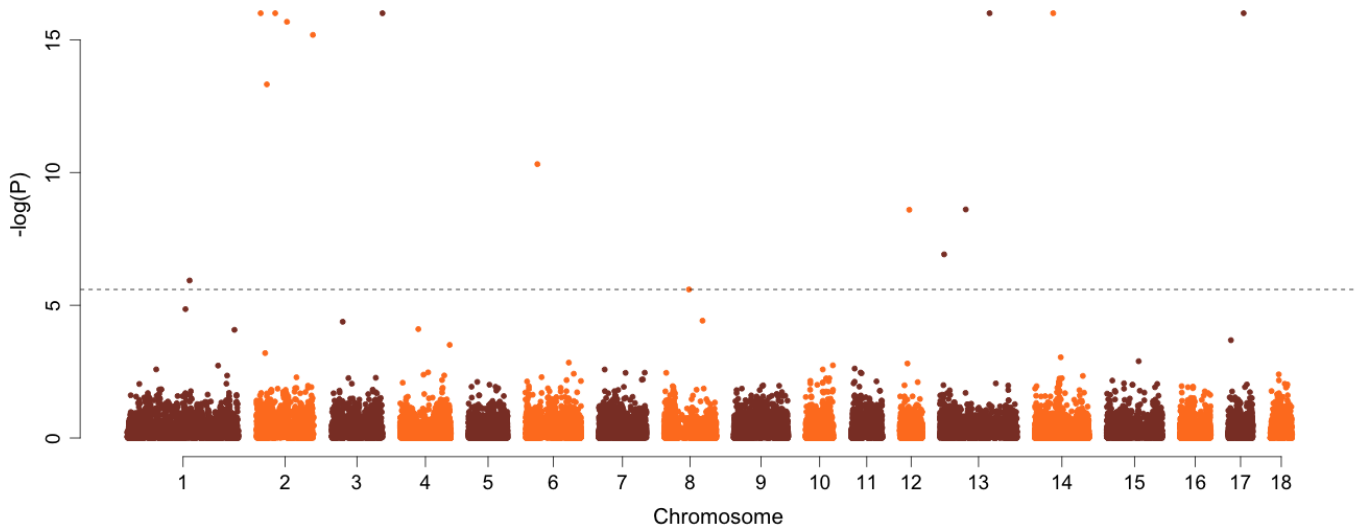


Figure 4.2 Histogram for the number of MI errors and valid calls of the significant SNPs on the X chromosome

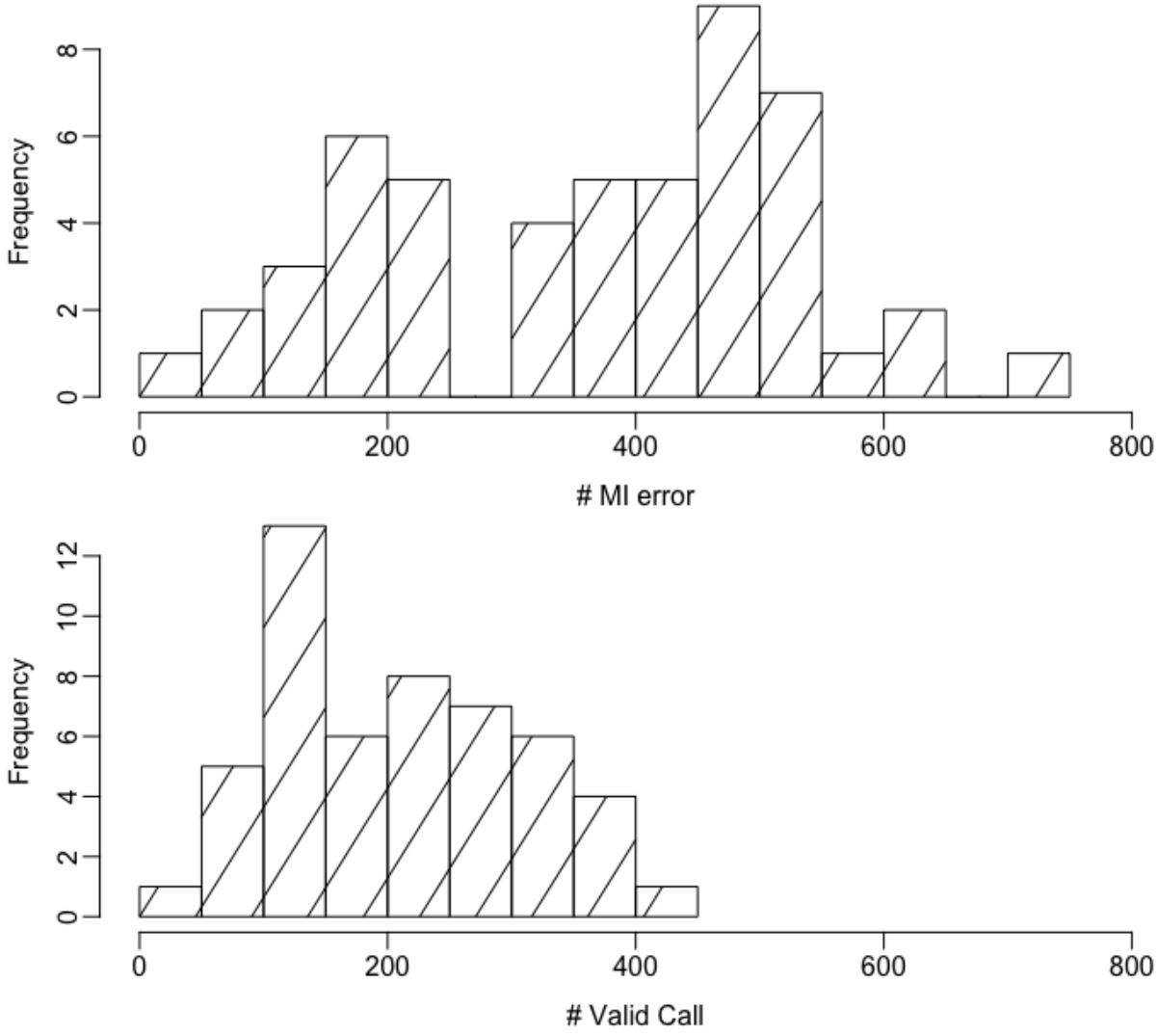
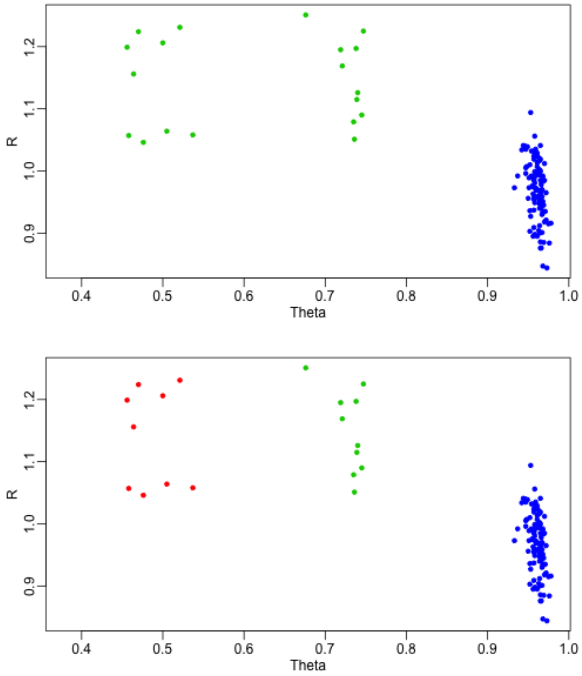


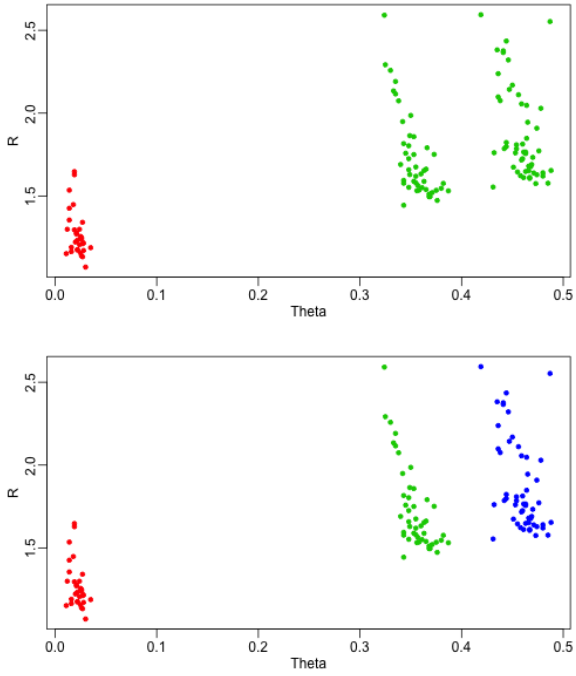
Figure 4.3 Raw genotyping data analysis (Part 1)

The upper plots show the original clustering of individuals for a certain SNP, and the lower plots show the clustering suggested by the family-based inference (i.e. maximizes the LOD value). Each point represents a parent, and is colored based on the genotype (red for AA, green for AB, blue for BB, and blue for missing).

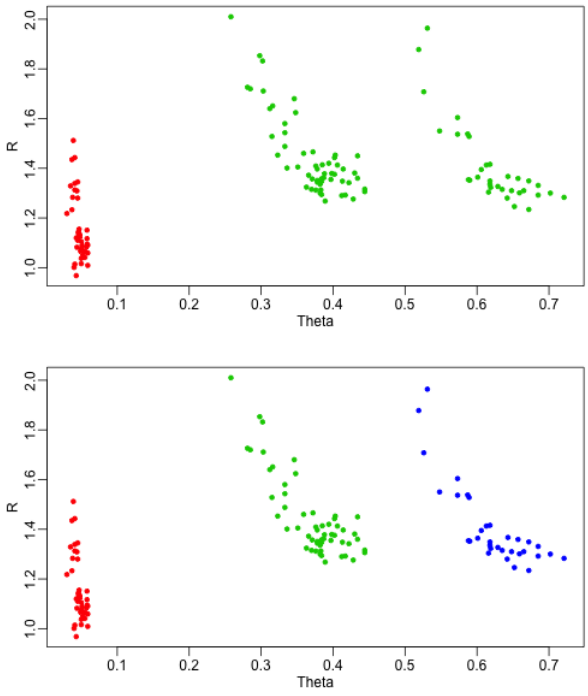
a. SNP: ASGA0076852



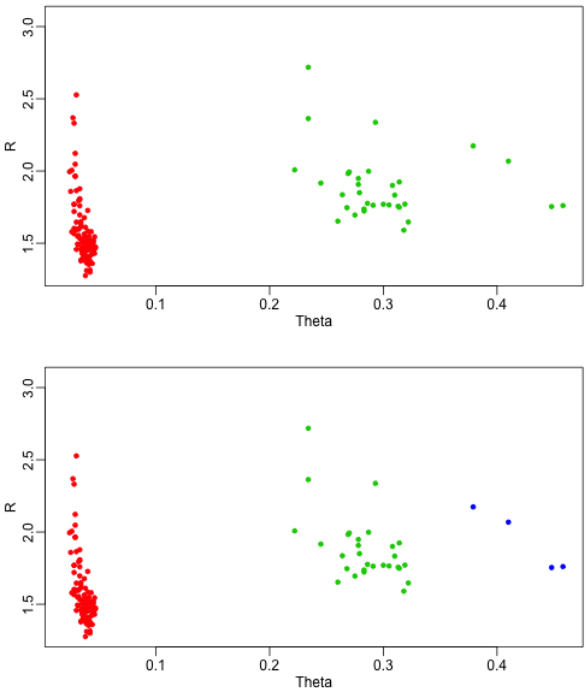
b. SNP: MARC0113898



c. SNP: ASGA0097612



d. SNP: MARC0111071



e. SNP: MARC0038959

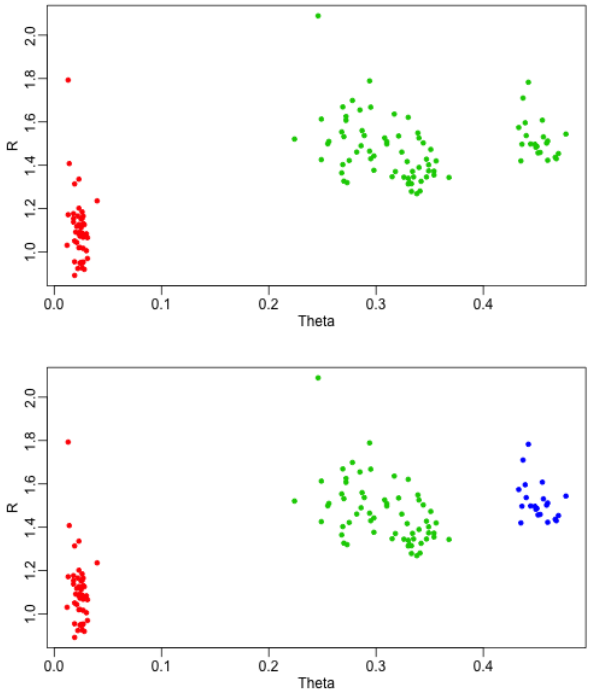
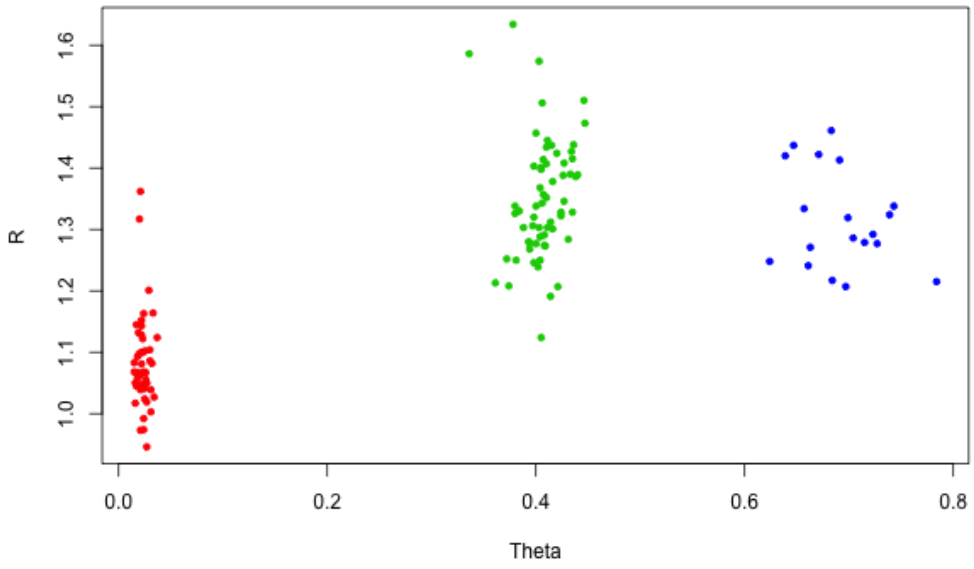


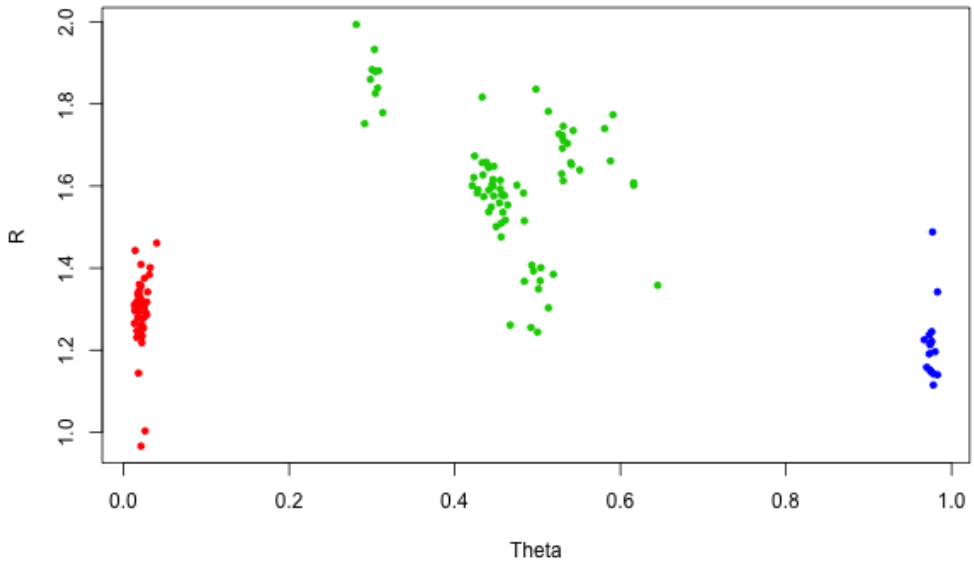
Figure 4.4 Raw genotyping data analysis (Part 2)

The R-Theta plots (upper plots) show the clustering of individuals for a certain SNP. Each point represents a parent, and is colored based on the genotype (black for AA, red for AB, green for BB, and blue for missing).

a. SNP: ASGA0004971



b. SNP: MARC0058459



Chapter 5 A statistical model for allele dosage data of sperm and its application in constructing individualized linkage maps

In Chapter 2 – 4, I mainly focused on how genomic data, especially SNP genotype data, can be used to explore the genetic basis of interesting traits in livestock production. In this chapter, we would like to introduce another approach to the use of cutting-edge biotechnologies and new types of genomic data, in order to deepen our understanding about the genome itself. Here we are going to propose a new method for linkage mapping, which may serve as a more efficient tool to construct better genome maps and discover new genomic variants on an individual basis. This approach may have potential for the development of personalized genomic-based treatment for human as well as animal health.

Preliminary results of this study has been published in the Proceedings of 10th World Congress on Genetics Applied to Livestock Production (WCGALP 2014). A manuscript based on this chapter has been prepared for submission.

5.1 Background

Linkage maps reveal the relative positions of genetic markers along the genome, using the degree of linkage as the measurement of the inter-marker distance. It has been more than 100 years since the first attempt of linkage mapping (Sturtevant 1913), and linkage maps are still one of the most essential tools in genetic and genomic studies. Their application includes serving as a scaffold in *de novo* DNA sequence assembly (Fierst 2015), providing information of genetic linkage in studies

about genomic evolution (Naruse et al. 2000) and recombination hotspots (Jeffreys et al. 2000; Auton and McVean 2007), and enabling linkage-based association studies (e.g. interval mapping) (Lander and Botstein 1989). Recent efforts to construct linkage maps mainly aim to make better use of new genotyping technologies such as RAD-sequencing (Gonen et al. 2014), as well as to apply the maps in sequence assembly (Xu et al. 2015).

Most linkage maps are constructed at a population level, where the relative position of genetic markers is inferred as an average result in a mapping population. This property is sufficient in some scenarios, such as building a reference genome to cover the majority of individuals for a species. Meanwhile, constructing a linkage map at an individual level (i.e. a linkage map specific for a certain individual) has an exclusive advantage as there can be significant variation in the relative positions of genetic markers between individuals even for the same species. Examples have been found in different species, such as human (Giglio et al. 2001), *C. elegans* (Hillier et al. 2008), and cattle (Marques et al. 2007; Snelling et al. 2007). An important one of these variations is chromosomal structural variations (CSVs) (Sudmant et al. 2015), such as rearrangement, insertion/deletion (indel), and copy number variations (CNVs). Individualized genome maps can serve as a way to type genome-wide CSVs, and play a critical role in many related genomic studies. Individualized linkage maps are also able to detect variation in recombination rate between individuals, which has been reported to be common in human and animals (Kong et al. 2010; Decker 2015). These variations may apply to animal breeding, as recombination may increase the response to artificial selection (Groenen et al. 2008).

Currently, few methods are perfectly suitable for constructing genome maps at an individualized level, even when methods for constructing physical maps are taken into consideration as well. Firstly, methods that require a mapping population, such as pedigree-based methods (Dear 2001) and linkage disequilibrium based methods (Fearnhead and Donnelly 2001), are not feasible for individualized genome mapping. Secondly, methods leading to low resolution may not be able to type minor chromosomal structural variants, and therefore have relatively limited application. Thirdly, methods with low-throughput, such as single sperm typing (Arnheim et al. 1991) and radiation-hybrid (RH) mapping (Cox et al. 1990), may not fit the needs in practice. For example, lower throughput means greater difficulty in screening a large number of individuals, and medical genomic tests may be expected and/or needed to produce timely results in some cases. Lastly and maybe most importantly, a good method for individualized genome mapping is supposed to have a low cost, which ensures the technology can be delivered successfully into practical applications. Even though some of the existing methods, such as optical mapping and long-read sequencing, provide relatively suitable solutions for individualized genome mapping, they are not yet perfect and there is still room for other competing approaches or technologies. Here we propose a strategy ("multiple sperm typing") for constructing individualized linkage maps using allele dosage data of sperm cells. This may potentially achieve a high throughput while keeping the cost at a desired level.

5.2 Methods

Generally, the construction of genetic linkage maps would include three parts: the estimation of pairwise inter-marker recombination rates, the grouping of markers, and the ordering of markers (Cheema and Dicks 2009). As inter-marker recombination rates carry information about the

relative positions of markers, the estimation of them greatly determines the resolution and accuracy of genome mapping. Our proposed methods mainly aim at improving the accuracy of the estimation.

The multiple sperm typing methods can be treated as a generalization of the traditional single sperm typing method (Arnheim et al. 1991), which genotypes individual sperm cells to infer the occurrence of recombination events during meiosis. Even though it performs well in constructing fine linkage maps, its low throughput greatly limited its application in practice (Pole et al. 2011). Instead, our new methods are designed to extract recombination information from a "sperm population" with a large number of individual sperm, without the need to genotype each single sperm cell. The estimation of recombination rate may achieve high accuracy as the number of sperm cells integrated into the analysis can be large.

In this section, we mainly describe the conclusions of the method, with the detailed statistical derivations provided in the Section 5.6. Here we first introduce allele dosage data, which is a new type of genomic data that has not previously been used in sperm typing. Single nucleotide polymorphisms (SNPs) are used as an example of genetic markers in the introduction and discussion. Then we propose two estimators of recombination rate under a simplified scenario. A standardization method is then introduced to fit a more general scenario. We also proposed a parametric hypothesis testing method to test for the existence of linkage, which can be used in marker grouping. Related topics about ordering markers will be discussed in later sections.

5.2.1 Allele dosage data

The SNP allelic dosage in a sample of sperm cells refers to the number of each possible allele on the genome. For pairwise recombination rate estimation, we only need to consider a pair of SNPs,

denoted as SNP A (with possible allele A1 and A2) and SNP B (with possible allele B1 and B2). For each sperm cell, there are four possible SNP genotypes at the two loci: A1B1, A1B2, A2B1, A2B2 (Table 1). The recombination rate between SNP A and SNP B, r_{AB}, can be estimated using a set of sperm samples. Now, we use n_s to denote the number of sperm samples, and $N_t = \left(n_{t,1}, n_{t,2}, \ldots, n_{t,n_s} \right)$ to denote the number of sperm cells in the samples. Then, the allele dosages for the four possible alleles (A1, B1, A2, and B2) in the sperm sample are denoted as N_{A1}, N_{B1}, N_{A2}, and N_{B2}, respectively. Similarly, in the i^{th} sperm sample, the number of the four possible alleles are denoted as $n_{A1,\mathrm{i}}$, $n_{B1,\mathrm{i}}$, $n_{A2,\mathrm{i}}$, and $n_{B2,\mathrm{i}}$. The relation between N_t and allele dosage is:

$$N_t = N_{A1} + N_{A2} = N_{B1} + N_{B2} \qquad \text{(Equation 5.1)}$$

5.2.2 A simplified scenario

In a simplified scenario, we assume that all sperm samples have the same number of sperm cells:

$$n_{t,1} = n_{t,2} = n_{t,n_s} = n_t \qquad \text{(Equation 5.2)}$$

Then we can estimate the recombination rate (r_{AB}) as a function of the correlation coefficient between N_{A1} and N_{B1} ($\rho_{N_{A1},N_{B1}}$):

$$r_{AB} = \begin{cases} \frac{1-\rho_{N_{A1},N_{B1}}}{2}, & if\ \rho_{N_{A1},N_{B1}} \geq 0 \\ \frac{1+\rho_{N_{A1},N_{B1}}}{2}, & if\ \rho_{N_{A1},N_{B1}} < 0 \end{cases} \qquad \text{(Equation 5.3)}$$

In the following description, we call Eq. 5.3 as Method I or Cor-Method. When allele A1 and B1 are on the same haplotype, it will be observed that $\rho_{N_{A1},N_{B1}} > 0$, then the first estimator, $\frac{1-\rho_{N_{A1},N_{B1}}}{2}$, is applicable, otherwise $\frac{1+\rho_{N_{A1},N_{B1}}}{2}$ should be used.

We can also estimate r_{AB} as a function of the variance of $N_{A1} - N_{B1}$ ($\sigma^2_{N_{A1}-N_{B1}}$):

$$r_{AB} = \begin{cases} \frac{\sigma^2_{N_{A1}-N_{B1}}}{n_t}, \ if \ \frac{\sigma^2_{N_{A1}-N_{B1}}}{n_t} \leq 0.5 \\ 1 - \frac{\sigma^2_{N_{A1}-N_{B1}}}{n_t}, if \ \frac{\sigma^2_{N_{A1}-N_{B1}}}{n_t} > 0.5 \end{cases} \quad \text{(Equation 5.4)}$$

We call Eq. 5.4 as Method II or Var-Method. Similar to Method I, we have two estimation equations for Method II. When allele A1 and B1 are on the same haplotype, the recombination rate can be estimated as $\frac{\sigma^2_{N_{A1}-N_{B1}}}{n_t} \leq 0.5$, otherwise $1 - \frac{\sigma^2_{N_{A1}-N_{B1}}}{n_t}$ should be used.

5.2.3 Standardized allele dosage data

When the assumption Eq. 5.2 is not true, we can standardize the allele dosage data for the estimation. The standardized allele dosage of allele A1, B1, A2, and B2 are denoted as N'_{A1}, N'_{B1}, N'_{A2}, and N'_{B2}, respectively, and the number are denoted as $n'_{A1,\mathrm{i}}$, $n'_{B1,\mathrm{i}}$, $n'_{A2,\mathrm{i}}$, and $n'_{B2,\mathrm{i}}$ for the i^{th} sperm sample. The standardization can be described as:

$$n'_{A1,\mathrm{i}} = (n_{A1,\mathrm{i}} - \frac{n_{t,i}}{2})/\sqrt{n_{t,i}}, \quad \text{(Equation 5.5)}$$

then Method I and Method II can be modified using the standardized data as:

$$r_{AB} = \begin{cases} \frac{1-\rho_{N'_{A1},N'_{B1}}}{2}, & if\ \rho_{N'_{A1},N'_{B1}} \geq 0 \\ \frac{1+\rho_{N'_{A1},N'_{B1}}}{2}, & if\ \rho_{N'_{A1},N'_{B1}} < 0 \end{cases} \quad \text{(Equation 5.6)}$$

and

$$r_{AB} = \begin{cases} \sigma^2_{N'_{A1}-N'_{B1}}, & if\ \sigma^2_{N'_{A1}-N'_{B1}} \leq 0.5 \\ 1-\sigma^2_{N'_{A1}-N'_{B1}}, & if\ \sigma^2_{N'_{A1}-N'_{B1}} > 0.5 \end{cases} \quad \text{(Equation 5.7)}$$

5.2.4 Marker grouping and Student t-test

In genome mapping, markers need to be grouped into linkage groups when the chromosome location is unknown. In our methods, we use a student's t-test to determine whether two markers are in the same linkage group or not. The null hypothesis (H_0) of the test is: there is no linkage between two given markers, which is equivalent to the event that the recombination rate between the two markers (r) is equal to 0.5. Based on our Method I, it is obvious that the correlation coefficient (ρ) is equal to zero when r is 0.5. The null hypothesis can be tested with the following *t* statistic (Gayen 1951):

$$t = \rho_{N'_{A1},N'_{B1}} \sqrt{\frac{n-2}{1-\rho^2_{N'_{A1},N'_{B1}}}}, \quad \text{(Equation 5.8)}$$

which follows a student's t distribution with a degree of freedom of $n-2$. The *t* statistic can be used to test the null hypothesis (no linkage). If the null hypothesis is rejected, an alternative hypothesis of existence of linkage can then be accepted.

5.2.5 Simulation experiments

Four simulation experiments were conducted to test the performance of our new method. In the first two experiments, we tested the accuracy of recombination rate estimation under a two-locus model. In the third experiment, we validate the null distribution of our estimators. In the fourth experiment, we simulated allele dosage data for 6,584 SNPs on porcine chromosome 1 (SSC1), estimated their inter-marker pairwise recombination rates, and tested the performance of our method in a more realistic genome mapping example. In the simulation experiments we assumed that allele dosage could be measured without error. The relaxation of this assumption will be discussed in Section 5.3.4.

5.2.5.1 Two-locus model and simulation experiments of estimation accuracy

In the first two experiments, we simulated allele dosage data for sperm samples where all sperm cells had a pseudo-genome with only two SNPs. The genotypes of each single sperm cell were simulated. However, the genotype of single sperm cells remained unknown and was not directly used in the estimation. Only allelic dosage data, the total number of each of the four possible alleles in each sperm sample, were available for the following analysis. The allele dosage data were simulated with different levels of sample number (n_s), sperm cell number in each sample (N_t), and true value of the recombination rate (r_{AB}) (Table 2a). For each combination of the parameters, both Method I and Method II were used to estimate $\mathrm{r_{AB}}$. The simulation and analysis was repeated 600 times. The performance of estimation was measured by mean square error (MSE), which is a joint measurement for accuracy and precision (Walther and Moore 2005).

$$\mathrm{MSE} = \sum(\mathrm{r_{AB}} - \mathrm{r_{AB,true}})^2/600 \qquad \text{(Equation 5.9)}$$

The differences between the first two experiments were the assumptions used in the simulation:

Experiment 1 In this experiment, all data were simulated under the simplified scenario described before, i.e. all sperm samples had the same number of sperm cells as the parameter N_t. For this experiment, we would like to evaluate Method I and Method II without standardizing the allele dosage data.

Experiment 2 In this experiment, the number of sperm cells is not identical to the parameter N_t, but followed a normal distribution whose expectation is the parameter N_t:

$$N_{t,simulated} \sim Normal\left(N_t, \left(\frac{N_t}{10}\right)^2\right) \qquad \text{(Equation 5.10)}$$

We set the coefficient of variation (CV) as 10%, so the standard deviation of the simulated cell number in each sample is $\frac{N_t}{10}$. With this experiment, we would like to test the performance and robustness of our method after the data standardization in a more realistic scenario (samples of unequal sperm cell number).

5.2.5.2 Hypothesis testing and null distribution

We also tested the null distribution of the estimated linkage using the two-locus model. As mentioned above, we can test the linkage between markers using a t-test, where the null distribution is a student's t distribution with a degree of freedom of $n_s - 2$. This statistical property may lead to a parametric way of testing linkage and grouping molecular markers into linkage groups.

To validate this statistical property, we simulated DNA quantification datasets under the two-locus model and repeated it 600 times, where the recombination rate between the two loci was set as 0.5 (i.e. no genetic linkage). Estimations were conducted using both Method I and Method II. The estimated recombination rates were used to calculate the *t* statistic and to build empirical null distributions. The observed null distributions were compared with the theoretical distribution through Q-Q (quantile-quantile) plots.

5.2.5.3 Genome map of SSC1

The fourth experiment was conducted to test how our method fits the objective of genome mapping. Instead of focusing on the absolute value of accuracy of the estimation of inter-marker recombination rates, this experiment was designed to test if the new estimation is good enough for linkage mapping in practice.

In this experiment, we simulated data of sperm samples from a heterozygous boar. In each sperm sample, there are 10^4 sperm cells whose genome contains only one chromosome. In total, 6583 heterozygous markers are simulated. Marker positions were set based on the position of real markers on the first chromosome of pigs (Illumina PorcineSNP60 SNP panel). True inter-marker recombination rates were calculated based on an approximation between recombination rate and physical distance (1 centimorgan = 10^6 base pairs). With this map, we simulated sperm samples, estimated inter-marker recombination rates, ordered SNPs, and constructed linkage maps for SSC1. The number of sperm samples were set as 15, 20, 25, 30, 35, 40, 45, and 50. For each setting of the sample number, we repeated the simulation and analysis 20 times. The estimation was conducted using Method II, as it outperformed Method I in the first two experiments. The ordering method used in the analysis is described in Section 5.7.

To evaluate the constructed maps, we reviewed the data to detect all misplacements of blocks of markers. For a certain block on the chromosome, if all the SNPs within this group are ordered perfectly, it is defined as a good block with a correct placement. Those blocks located between these "good blocks" are defined as misplacement blocks, and SNPs within misplacement blocks are ordered in an incorrect way. The size and number of these misplacement blocks can be used as a measurement of the quality of constructed linkage maps. The fewer and smaller the misplacement blocks there are, the better the mapping result.

5.3 Results and discussions

5.3.1 Estimation accuracy

The results of simulation experiment 1 and 2 were presented in Figure 1 and Figure 2, respectively. No evidence of an obvious difference in their accuracy was found between experiment 1 and 2. So we will not distinguish between them when we discuss how the three parameters (N_t, sample number, and inter-marker recombination rate) impacted the accuracy of the estimation.

As we expected increasing N_t improved the accuracy of estimation. Traditional sperm typing methods only make use of data from single sperm cells, which is equivalent to allele dosage data with only one sperm cell in each sperm sample. Through including more sperm cells in each sample, our method was able to outperform the traditional single sperm typing method in accuracy. As the recombination rate was essentially estimated using recombination events detected in the genotype data, a larger number of sperm cell means more available recombination events that can be used for estimation and therefore results in greater accuracy. The improvement, however, slowed down when N_t was higher than a certain value, which implied that the improvement that

can be achieved through increasing N_t has an upper bound, and we can optimize N_t to find a balance between the accuracy and the cost of data collection.

The degree of the improvement through increasing N_t also varied across different parameter combinations. When the inter-marker recombination rate was 0.001, the improvement started to slow down when N_t is higher than 10^3; when the inter-marker recombination rate was 0.1, increasing N_t did not really help after $N_t > 10^2$. It was possibly due to the different frequencies of recombination events. For a recombination rate of 0.001, the recombination event was very rare in the simulated data when N_t is low, so increasing N_t is very important and useful. When the recombination rate was 0.1, it was supposed to have a fair number of recombination events even when N_t was low. Thus, with relatively small N_t, the accuracy of estimation was good enough and did not leave much room for further improvement. This observation implies that the advantage of our method is greater when the inter-marker recombination rate is smaller, which is very important in constructing high-resolution genome maps.

Another way to improve the estimation accuracy is to increase sample number (n_s). Within the range of n_s we set in the two experiments (10 - 160), a higher n_s almost always led to higher accuracy. Increasing n_s had little effect on the relationship between the accuracy and N_t. However, high n_s usually means a higher cost for data collection, which will be an important factor during experimental design in practice.

The comparison between experiment 1 and 2 also provides information on the impact of N_t. In experiment 2, the assumptions were less strict than experiment 1 as the cell numbers across samples (N_t) were variable, and a standardization method was applied to adjust for that variability. As both Method I and Method II achieved an accuracy in experiment 2 that was very close to that

in experiment 1, we may conclude that our methods seem robust after data standardization, and they can be applied to a more realistic scenario where values of N_t are variable.

In terms of comparison between Method I and Method II, we can see that Method II outperformed Method I in most cases, especially when N_t is high (e.g. greater than 10^2).

5.3.2 Null distribution of the estimator and marker grouping

The Q-Q plot comparing the observed null distribution to the expected null distribution (Student's t-distribution) were presented as Figure 3 (Method I) and Figure 4 (Method II). The dashed lines on these Q-Q plots had an intercept of 0 and a slope of 1. In Figure 3, most of the points were located approximately along the dashed lines, which suggested that the observed empirical null distributions fitted the theoretical distribution well. In the tested range, N_t did not have any obvious impact on the distribution. When the sample number (n_s) was low, however, there are more outliers deviating from expectation in the observed dataset.

We also tested the null distribution of Method II using the simulated datasets (Figure 4). The observed null distribution did not fit Student's t-distribution as well as that of Method I, which is not surprising given that the theoretical distribution was derived based on Method I. From the Q-Q plots, we found that the approximate slope of the points was lower than expectation, which implied that the null distribution of Method II was wider than expected. Some of the null distributions were right-skewed, especially when sample number (n_s) was low.

Testing the existence of linkage between markers is not always necessary. For human as well as most widely-used species in genomic studies, there usually are reference maps with fairly good quality, in which case the linkage group (or chromosome) is mostly clear. However, the grouping of markers can be critical for *de novo* sequencing, where the reference maps for these

sequences may not be available. In addition, even with the help of a reference map, the testing of linkage may help with detecting and mapping chromosome structural variation, such as chromosomal translocations or crossovers, where the information provided by reference maps can be very limited. For Method I, the goodness of fit of our estimator to the theoretical distribution enabled a traditional parametric hypothesis testing method for marker grouping. For Method II, even though the t-test is not feasible, it may be possible to construct empirical null distribution and test the linkage with Monte-Carlo methods in practice (not tested in the current study).

5.3.3 Performance in genome mapping

The results of simulation experiment 4 showed that the increase of sample number significantly reduced the number of mismatches, the average number of SNPs within each mismatch block, and the average size of mismatch blocks (Figure 5 and 6). Especially, when 50 samples were used, 96.3% of the mismatch blocks were small misplacements between two adjacent SNPs. The 95th percentile of the size of the mismatch blocks was 27,450 base pairs, i.e., 95% of the mismatch blocks were no larger than 0.02745 centimorgan. The results confirmed that the estimation provided by our method is qualified for the construction of linkage maps.

5.3.4 Variance partitioning model

In the estimation described above it was assumed that allele dosage could be measured without error, which will not be realistic in practice. We also constructed a variance partitioning model for Method II, in order to model possible error introduced during DNA quantification. Under an assumption of variance homogeneity for the measurement error, we can estimate the variance of random measurement error (σ_e^2). Then we have:

$$\left(\sigma^2_{N'_{A1}-N'_{B1}}\right)_{observed} = \left(\sigma^2_{N'_{A1}-N'_{B1}}\right)_{true} + 2\sigma^2_e, \qquad \text{(Equation 5.11)}$$

which implies that the observed variance of $(N_{A1} - N_{B1})$ is always greater than the true value when there is measurement error.

This model may allow us to adjust the estimation for possible measurement error and leads to more robust estimation. However, the real distribution of the measurement error is usually unknown or varies greatly according to the DNA quantification technology used. Further optimization based on this model is possible but it is beyond the discussion of the method at this time.

5.3.5 From sperm cells to other haploids

Even though the proposed model is based on allele dosage data of sperm cells, the method can also apply to all gametes (e.g. eggs or pollen) of diploid organisms, if the gametes are available in sufficient numbers. The application of the proposed method is therefore not limited to male animals. The model can potentially be integrated into the HAPPY (HAPloid equivalents of DNA and the PolYmerase chain reaction) mapping strategy as well, which uses DNA samples containing approximately 1 haploid genome equivalent as an analogue of real haploids like gametes, and thereby constructing physical genome maps (Dear and Cook 1993). The proposed model is also able to model the data of such haploid analogues by replacing the "recombination rate" with "probability of breakage". As a result, a "multiple HAPPY mapping" method has the potential to expand applications to somatic cells including characterization of genome modification in tumors.

5.3.6 Future work: From *in silico* experiments to benchtop experiments

Given the evidence for the theoretical potential of the new approach, significant challenges remain to reduce the method to practice. This strategy can be tested in three stages in new studies: 1) develop an optimized protocol of low-cost sample preparation and data collection, 2) develop an integrated analysis toolset for data manipulation, statistical analysis and map construction, and 3) create high-resolution individualized genomic maps of a model animal. In the first two stages, we can evaluate to what degree current DNA quantification platforms are able to support our method, and illustrate the feasibility of the mapping strategy. Based on these results we may be able to generate high-resolution whole-genome genetic maps using model species in the third step. It may ultimately lead to personalized maps for human and livestock applications.

An important component in the application of this method is selecting the quantification platform, in order to improve the quality of allelic dosage data and accuracy of mapping. So far, potential platforms for DNA quantification includes 1) bead-based quantitative genotyping technology (e.g. Illumina PorcineSNP60v2+ Beadchip), 2) droplet digital PCR (ddPCR) technology (e.g. Raindance Raindrop system), 3) nCounter, another digital quantification platform, and 4) next-generation sequencing (see Section 1.1.3). The evaluation should include the accuracy of DNA quantification, throughput of the system, and the economic cost.

Further optimization of DNA quantification may consider the DNA preparation step. We can explore the best protocol for DNA preparation from sperm to achieve the best performance in quantification and mapping. Theoretically, the use of more sperm will lead to more accurate estimation and higher resolution of the resulting maps. However, this may not be realistic in practice due to the physical aspects of DNA preparation and other factors such as quantification. These experiments will help us to find the best parameters for the DNA extraction and

quantification, including the number of sperm or the amount sperm DNA within a sample that suits the specific DNA quantification platform.

The second stage can be the development of better tools for the analysis stage of the mapping. These can be designed in order to generate fast outputs from the data generated by multi-sperm typing. An integrated analysis toolset can provide the necessary tools for allelic dosage data cleaning and quality control, recombination rate estimation, SNP grouping, SNP ordering, and genetic map construction and visualization.

In terms of high-resolution whole-genome maps, the capability and feasibility of our proposed method can be demonstrated by creating a simplified regional genetic map for a well-characterized rearrangement in a mammalian model. Another excellent model is the genome of C. elegans. It has a relatively small and compact genome of 100Mb (hermaphrodite karyotype 5 autosomal pairs, XX) and the most complete genome assembly of any metazoan organism. Most importantly, many isogenic strains are available that have well characterized genetic rearrangements and breakpoints defined to the precise base pair, such as translocations, deletions, inversions and duplications (Genetic balancers). These models will allow the mapping approach to be rigorously tested and validated for all species.

5.4 Conclusion

We proposed a strategy ("multiple sperm typing") for inter-marker recombination rate estimation and genome linkage mapping using allele dosage data. The methods can serve as a generalization of the traditional single sperm typing method. The simulation experiments provided support for our hypothesis that the new methods can significantly improve the accuracy of inter-marker recombination rate estimation, and the estimates showed good statistical property for marker

grouping. Further extensions are possible in practice, such as optimization through adjusting random errors and application in other haploid data (e.g. eggs or pollen). Given that our methods may give better estimation of recombination rates, reduce the difficulty of data collection, and make better use of high-throughput and low-cost technologies, we believe that the methods may lead to high-resolution, low-cost linkage maps, and be applied in human medicine and animal genomics.

5.5 References

Agarwala, R. 2000. A Fast and Scalable Radiation Hybrid Map Construction and Integration Strategy. Genome Res. **10**: 350–364.

Arnheim, N., Li, H., and Cui, X. 1991. Genetic mapping by single sperm typing. Anim. Genet. **22**: 105–115.

Auton, A., and McVean, G. 2007. Recombination rate estimation in the presence of hotspots. Genome Res. **17**: 1219–27.

Cheema, J., and Dicks, J. 2009. Computational approaches and software tools for genetic linkage map estimation in plants. Brief. Bioinform. **10**: 595–608.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S., and Myers, R.M. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science **250**: 245–50.

Dear, P.H. 2001. Genome Mapping. *in* Encyclopedia of Life Sciences.

Dear, P.H., and Cook, P.R. 1993. Happy mapping: linkage mapping using a physical analogue of meiosis. Nucleic Acids Res. **21**: 13–20.

Decker, J.E. 2015. Agricultural Genomics: Commercial Applications Bring Increased Basic Research Power. PLOS Genet. **11**: e1005621.

Genetic balancers. [Online] Available: http://www.wormbook.org/chapters/www_geneticbalancers/geneticbalancers.html.

Everts-van der Wind, A., Larkin, D.M., Green, C. a, Elliott, J.S., Olmstead, C. a, Chiu, R., Schein, J.E., Marra, M. a, Womack, J.E., and Lewin, H. a 2005. A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. Proc. Natl. Acad. Sci. U. S. A. **102**: 18526–31.

Fearnhead, P., and Donnelly, P. 2001. Estimating recombination rates from population genetic data. Genetics **159**: 1299–318.

Fierst, J.L. 2015. Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. Front. Genet. **6**: 1–8.

Gayen, A.K. 1951. The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. Biometrika **38**: 219–47.

Giglio, S., Broman, K.W., Matsumoto, N., Calvari, V., Gimelli, G., Neumann, T., Ohashi, H., Voullaire, L., Larizza, D., Giorda, R., Weber, J.L., Ledbetter, D.H., and Zuffardi, O. 2001. Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am. J. Hum. Genet. **68**: 874–83.

Gonen, S., Lowe, N.R., Cezard, T., Gharbi, K., Bishop, S.C., and Houston, R.D. 2014. Linkage maps of the Atlantic salmon (Salmo salar) genome derived from RAD sequencing. BMC Genomics **15**: 166.

Groenen, M.A.M., Wahlberg, P., Foglio, M., Cheng, H.H., Megens, H.-J., Crooijmans, R.P.M.A., Besnier, F., Lathrop, M., Muir, W.M., Wong, G.K.-S., Gut, I., and Andersson, L. 2008. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res. **19**: 510–519.

Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.I., Hickenbotham, M., Huang, W., Magrini, V.J., Richt, R.J., Sander, S.N., Stewart, D. a, Stromberg, M., Tsung, E.F., Wylie, T., Schedl, T., Wilson, R.K., and Mardis, E.R. 2008. Whole-genome sequencing and variant discovery in C. elegans. Nat. Methods **5**: 183–188.

Jeffreys, A.J., Ritchie, A., and Neumann, R. 2000. High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. Hum. Mol. Genet. **9**: 725–

33.

Karp, R., Ruzzo, W., and Tompa, M. 1996. Algorithms in molecular biology (lecture notes).

Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A., Jonasdottir, A., Walters, G.B., Gylfason, A., Kristinsson, K.T., Gudjonsson, S.A., Frigge, M.L., Helgason, A., Thorsteinsdottir, U., and Stefansson, K. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. Nature **467**: 1099–1103.

Lander, E.S., and Botstein, S. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**: 185.

Liu, B.H. 1998. Statistical genomics: linkage, mapping, and QTL analysis.

Marques, E., de Givry, S., Stothard, P., Murdoch, B., Wang, Z., Womack, J., and Moore, S.S. 2007. A high resolution radiation hybrid map of bovine chromosome 14 identifies scaffold rearrangement in the latest bovine assembly. BMC Genomics **8**: 254.

Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R., Shimada, A., Wada, H., Kusakabe, T., Suzuki, N., Kinoshita, M., Kanamori, A., Terado, T., Kimura, H., Nonaka, M., and Shima, A. 2000. A detailed linkage map of medaka, Oryzias latipes: comparative genomics and genome evolution. Genetics **154**: 1773–84.

Olson, J.M., and Boehnke, M. 1990. Monte Carlo comparison of preliminary methods for ordering multiple genetic loci. Am. J. Hum. Genet. **47**: 470–82.

Pole, J.C.M., McCaughan, F., Newman, S., Howarth, K.D., Dear, P.H., and Edwards, P.A.W. 2011. Single-molecule analysis of genome rearrangements in cancer. Nucleic Acids Res. **39**: e85–e85.

Reinelt, G. 1991. TSPLIB—a traveling salesman problem library. ORSA J. Comput. **3**: 376–384.

Snelling, W.M., Chiu, R., Schein, J.E., Hobbs, M., Abbey, C.A., Adelson, D.L., Aerts, J., Bennett, G.L., Bosdet, I.E., Boussaha, M., Brauning, R., Caetano, A.R., Costa, M.M., Crawford, A.M., Dalrymple, B.P., Eggen, A., Everts-van der Wind, A., Floriot, S., Gautier, M., Gill, C.A., Green, R.D., Holt, R., Jann, O., Jones, S.J., Kappes, S.M., Keele, J.W., de Jong, P.J., Larkin, D.M., Lewin, H.A., McEwan, J.C., McKay, S., Marra, M.A., Mathewson, C.A., Matukumalli, L.K., Moore, S.S., Murdoch, B., Nicholas, F.W., Osoegawa, K., Roy, A., Salih, H., Schibler, L., Schnabel, R.D., Silveri, L., Skow, L.C., Smith, T.P., Sonstegard, T.S., Taylor, J.F., Tellam, R., Van Tassell, C.P., Williams, J.L., Womack, J.E., Wye, N.H., Yang, G., and Zhao, S. 2007. A physical map of the bovine genome. Genome Biol. **8**: R165.

Sturtevant, A.H. 1913. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. **14**: 43–59.

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Hsi-Yang Fritz, M., Konkel, M.K., Malhotra, A., Stütz, A.M., Shi, X., Paolo Casale, F., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., Malig, M., Chaisson, M.J.P., Walter, K., Meiers, S., Kashin, S., Garrison, E., Auton, A., Lam, H.Y.K., Jasmine Mu, X., Alkan, C., Antaki, D., Bae, T., Cerveira, E., Chines, P., Chong, Z., Clarke, L., Dal, E., Ding, L., Emery, S., Fan, X., Gujral, M., Kahveci, F., Kidd, J.M., Kong, Y., Lameijer, E.-W., McCarthy, S., Flicek, P., Gibbs, R.A., Marth, G., Mason, C.E., Menelaou, A., Muzny, D.M., Nelson, B.J., Noor, A., Parrish, N.F., Pendleton, M., Quitadamo, A., Raeder, B., Schadt, E.E., Romanovitch, M., Schlattl, A., Sebra, R., Shabalin, A.A., Untergasser, A., Walker, J.A., Wang, M., Yu, F., Zhang, C., Zhang, J., Zheng-Bradley, X., Zhou, W., Zichner, T., Sebat, J., Batzer, M.A., McCarroll, S.A., Mills, R.E., Gerstein, M.B., Bashir, A., Stegle, O., Devine, S.E., Lee, C., Eichler, E.E., and Korbel, J.O. 2015. An integrated map of structural

variation in 2,504 human genomes. Nature **526**: 75–81.

Walther, B.A., and Moore, J.L. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography (Cop.). **28**: 815–829.

Xu, S. 2013. Principles of Statistical Genomics.

Xu, S., Ackerman, M.S., Long, H., Bright, L., Spitze, K., Ramsdell, J.S., Thomas, W.K., and Lynch, M. 2015. A Male-Specific Genetic Map of the Microcrustacean Daphnia pulex Based on Single-Sperm Whole-Genome Sequencing. Genetics **201**: 31–38.

Table 5.1 Four different kinds of sperm cells in a two-locus model

#	Sperm Type[1]	Genotype	Proportion[2]
1	Parental	A1-B1	$P_{11} = \frac{1 - r_{AB}}{2}$
2	Parental	A2-B2	$P_{22} = \frac{1 - r_{AB}}{2}$
3	Recombinant	A1-B2	$P_{12} = \frac{r_{AB}}{2}$
4	Recombinant	A2-B1	$P_{21} = \frac{r_{AB}}{2}$

[1]Assuming that allele A1 and B1 are on the same haplotype.

[2]r_{AB} represents the recombination rate between SNP A and SNP B.

Table 5.2 Parameters used in the simulation experiment (Two-locus model)

Parameter	**Values**
Sample number $(\boldsymbol{n_s})$	10, 20, 40, 80, 160
Sperm cell number in each sample $(\boldsymbol{N_t})$	1, 10, 100, 1000, 10000
True value of the recombination rate $(\boldsymbol{r_{AB}})$	0.001, 0.01, 0.1

Figure 5.1 Performance of Method I (Cor_Method) and Method II (Var_Method) under simplified scenarios.

Higher value on y-axis means higher accuracy.

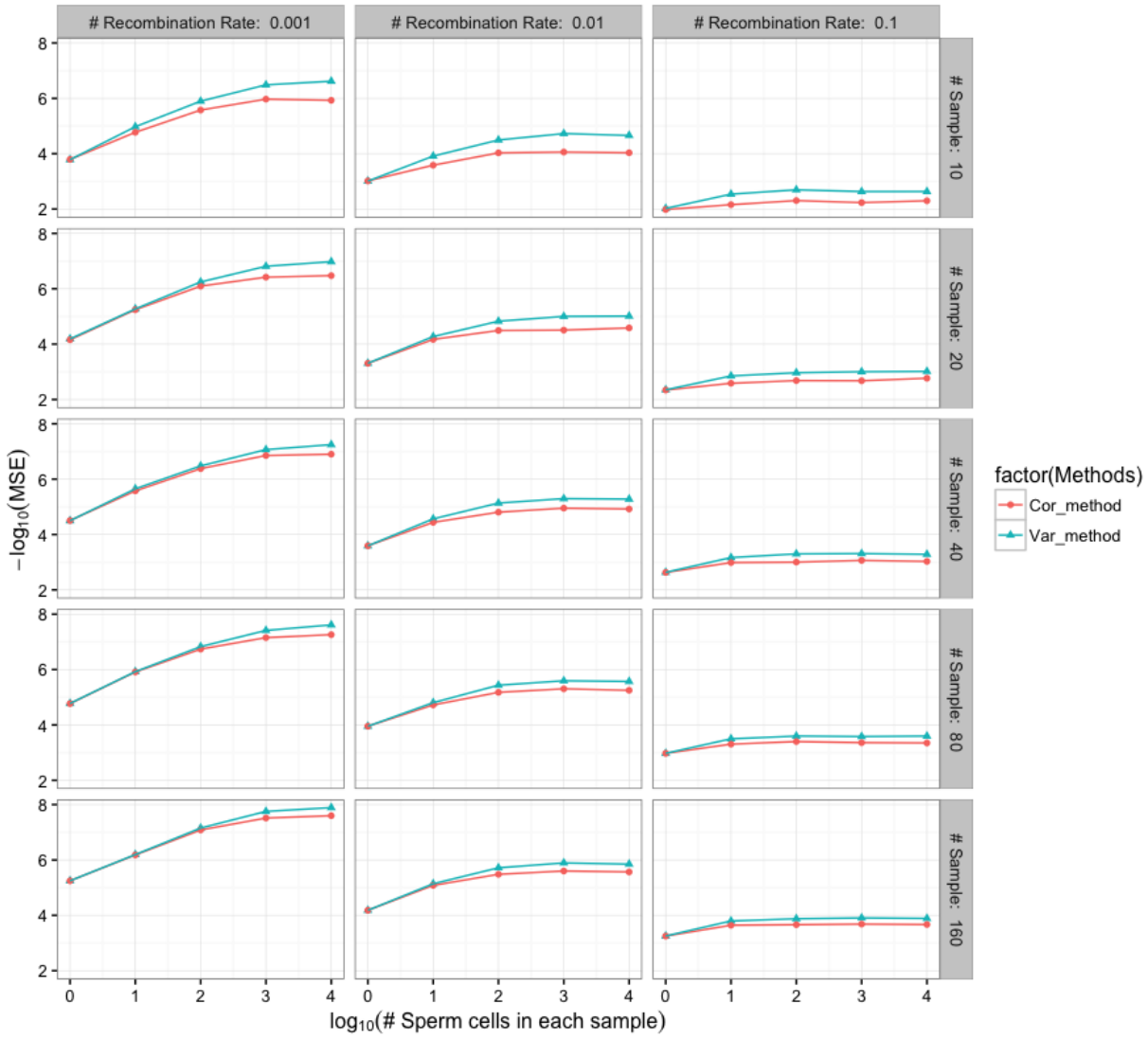


Figure 5.2 Performance of Method I (Cor_Method) and Method II (Var_Method) under more realistic scenarios.

Higher value on y-axis means higher accuracy.

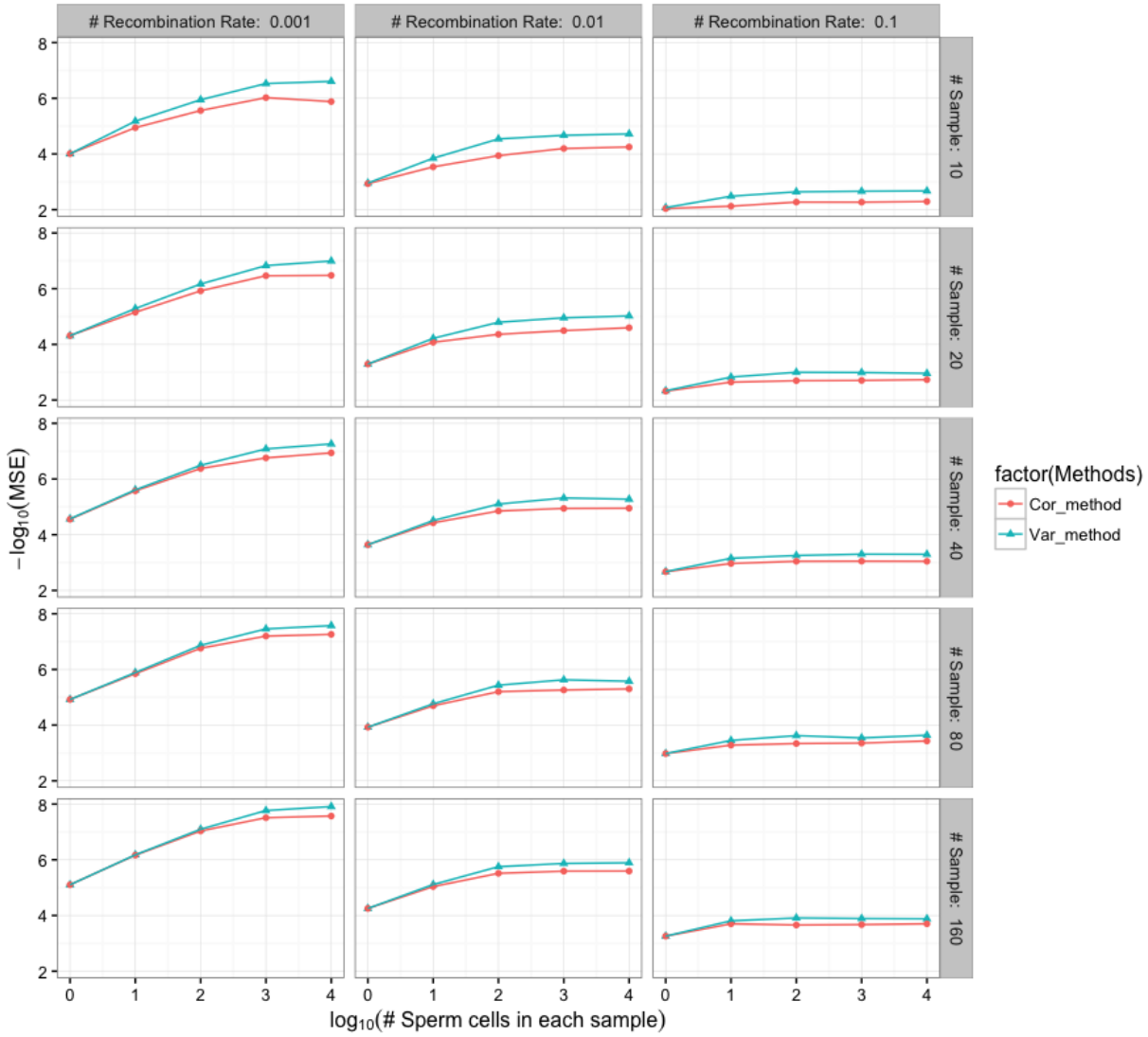


Figure 5.3 Q-Q plots comparing the observed null distribution to the expected null distribution (Student's t-distribution) for the estimation in Method I (Cor_Method).

Dashed lines have an intercept of 0 and a slope of 1.

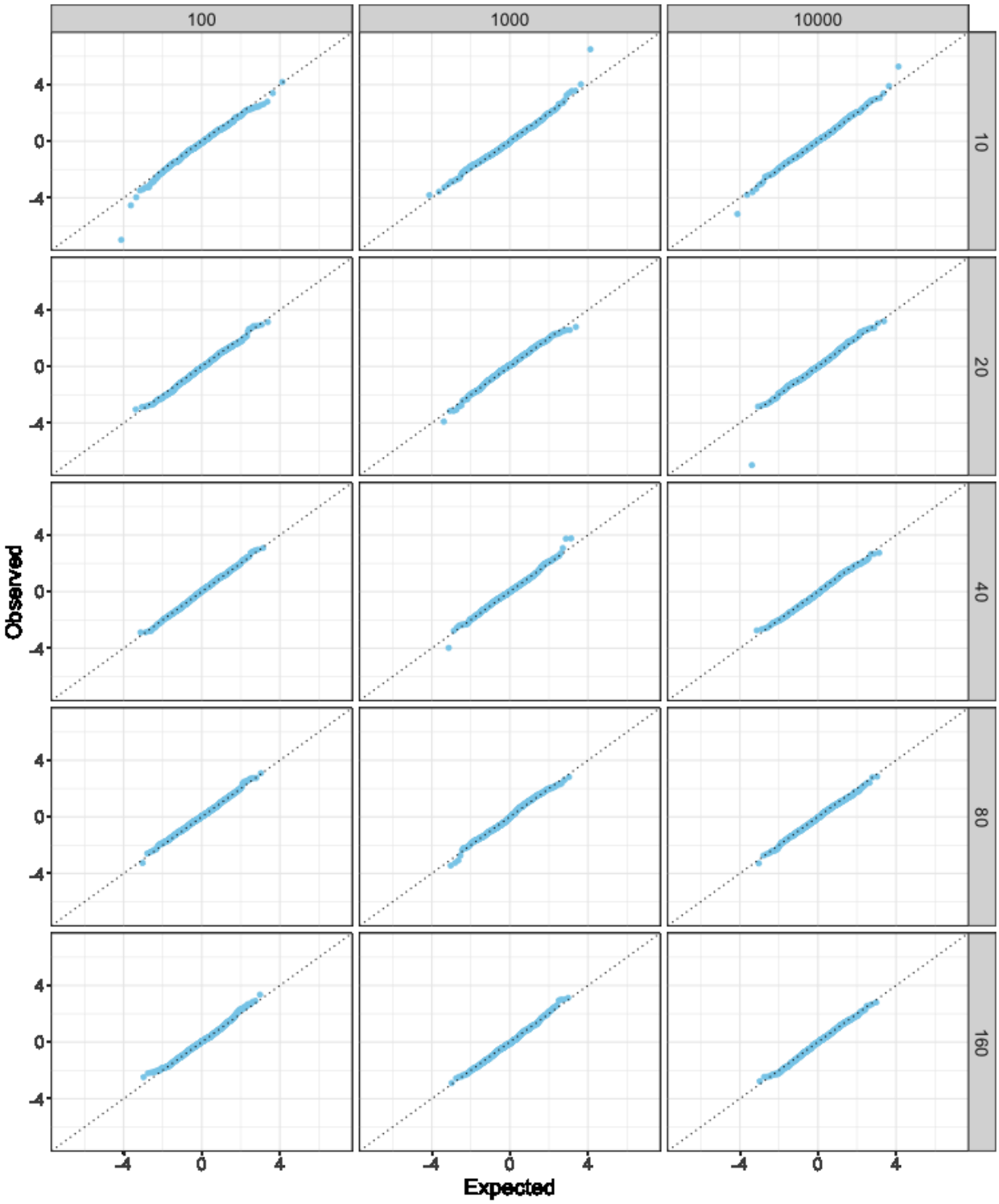


Figure 5.4 Q-Q plots comparing the observed null distribution to the expected null distribution (Student's t-distribution) for the estimation in Method II (Var_Method).

Dashed lines have an intercept of 0 and a slope of 1.

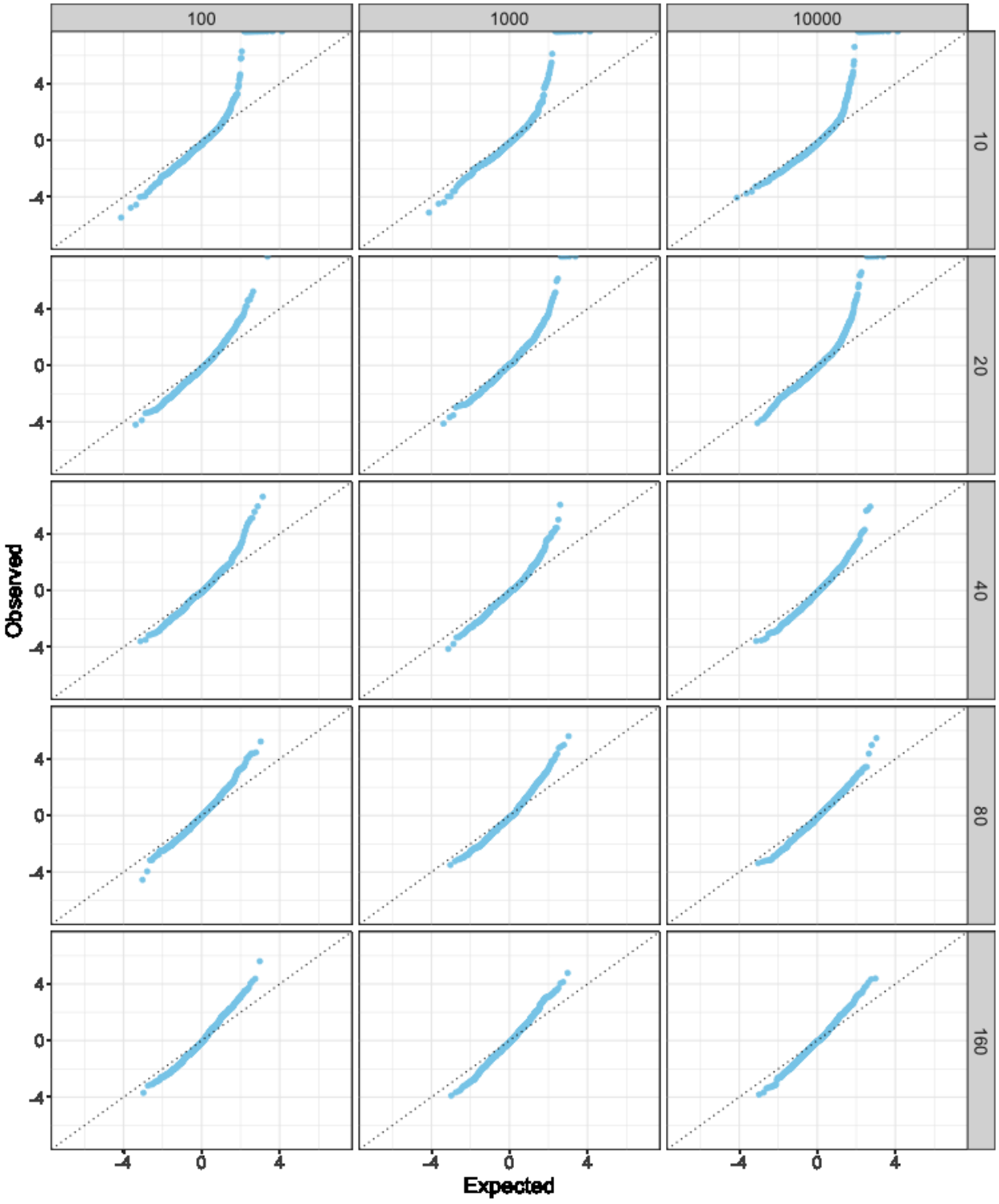


Figure 5.5 Distribution of the size of misplacements (in base pairs) in constructed maps of SSC1. Error bars represent one standard deviation.

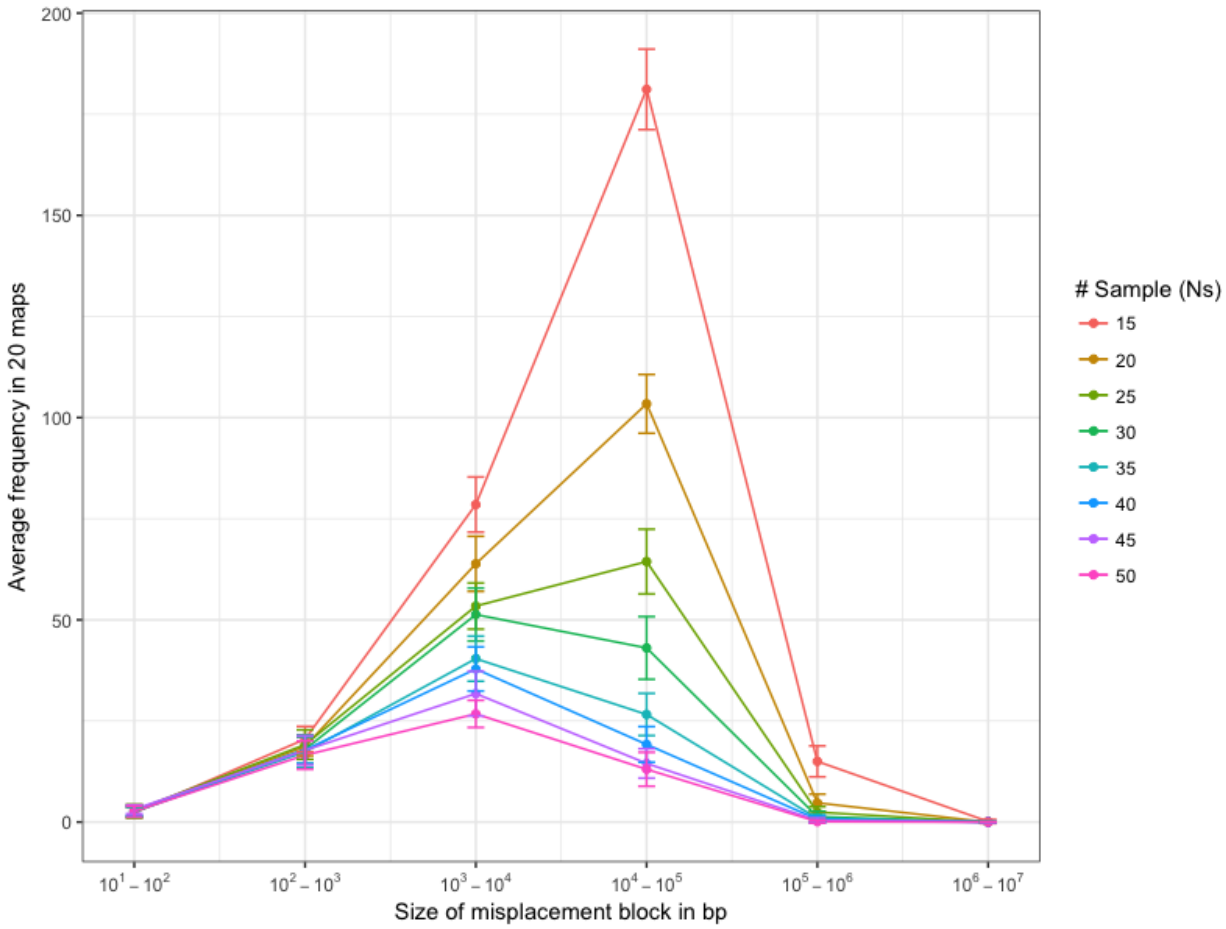
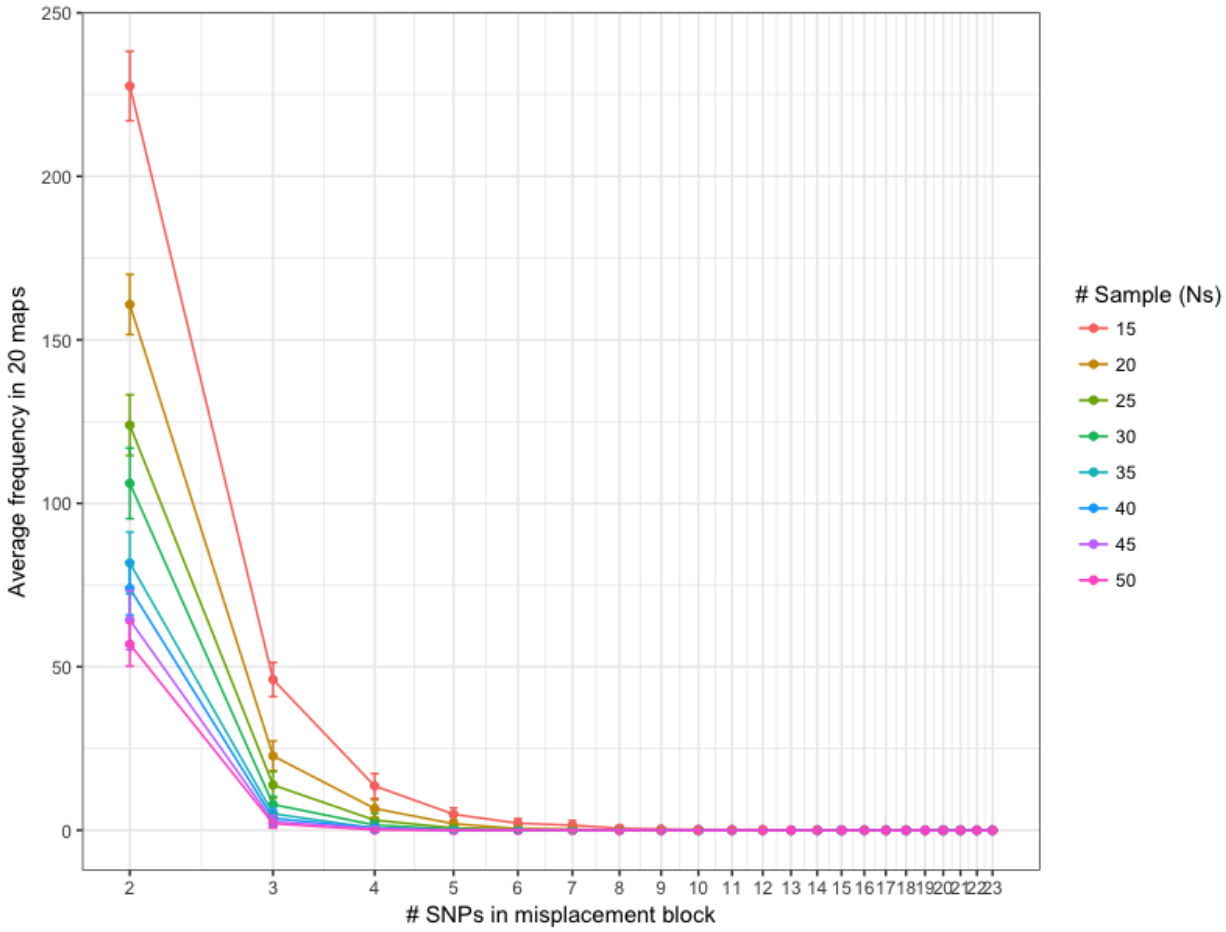


Figure 5.6 Distribution of the size of misplacements (in SNP number) in constructed maps of SSC1. Error bars represent one standard deviation.



5.6 Derivation of Method I (Cor_Method) and Method II (Var_Method)

5.6.1 Distribution of allelic dosage in the simplified scenario

Here we consider two SNPs in a two-locus model: SNP A (with possible allele A1 and A2) and SNP B (with possible allele B1 and B2). If all sperm samples have the same number of sperm cells (n_t), then the allele dosage of allele A1 follows a binomial distribution:

$$N_{A1} \sim \text{Binomial}\left(n_t, \frac{1}{2}\right) \quad \text{(Equation 5.11)}$$

Similarly, the allele dosage of allele B1 also follows a binomial distribution:

$$N_{B1} \sim \text{Binomial}\left(n_t, \frac{1}{2}\right) \quad \text{(Equation 5.12)}$$

In this two-locus model, there are four possible types of sperm: two with parental genotypes and two recombinants (Table 1). Assuming that allele A1 and B1 are on the same haplotype, i.e. parental genotypes are A1-B1 and A2-B2, the joint distribution of the number of these four different combinations ($N_{11}, N_{12}, N_{21}, N_{22}$) would follow a multinomial distribution:

$$N = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \text{Multinomial}(n_t, P). \quad \text{(Equation 5.13)}$$

, where $P = (p_{11}, p_{12}, p_{21}, p_{22})$ represents the expected proportion of the four kinds of sperm (Table 1). A variance-covariance matrix can then be calculated based on the property of multinomial distribution.

5.6.2 Method I and Method II in the simplified scenario

The covariance of N_{A1} and N_{B1} can be derived from the multinomial distribution:

$$\begin{aligned}\sigma_{n_{A1},n_{B1}} &= \sigma_{(n_{11}+n_{12}),(n_{11}+n_{21})} \\ &= \sigma_{n_{11}} + \sigma_{n_{11},n_{21}} + \sigma_{n_{11},n_{12}} + \sigma_{n_{12},n_{21}} \\ &= \frac{n_t}{4}(1-2r_{AB})\end{aligned} \qquad \text{(Equation 5.14)}$$

, and then we get:

$$\begin{aligned}\sigma^2_{N_{A1}} &= \sigma^2_{N_{B1}} = \frac{n_t}{4} \\ \sigma_{N_{A1},N_{B1}} &= \frac{n_t}{4}(1-2r_{AB})\end{aligned} \qquad \text{(Equation 5.15)}$$

When n_t is large enough, the binomial distribution (Eq. 5.13) can be approximated with a normal distribution:

$$\binom{N_{A1}}{N_{B1}} \sim \text{Normal}\left[\frac{1}{2}\binom{n_t}{n_t}, \frac{1}{4}n_t\begin{pmatrix}1 & 1-2r_{AB}\\ 1-2r_{AB} & 1\end{pmatrix}\right] \qquad \text{(Equation 5.16)}$$

Then the Pearson's correlation between N_{A1} and N_{B1} can be calculated as:

$$\rho_{N_{A1},N_{B1}} = \frac{\sigma_{N_{A1},N_{B1}}}{\sqrt{\sigma_{N_{A1}}\sigma_{N_{B1}}}} = 1-2r \qquad \text{(Equation 5.17)}$$

So the recombination rate can be estimated as:

$$r_{AB} = (1 - \rho_{N_{A1},N_{B1}})/2 \qquad \text{(Equation 5.18)}$$

, which is Method I (Cor_Method) under the assumption that all sperm samples have the same number of sperm cells.

We can also derive the Method II (Var_Method) from the multinomial distribution. Notice that:

$$N_{A1} - N_{B1} = N_{11} + N_{12} - (N_{11} + N_{21}) = N_{12} - N_{21} \qquad \text{(Equation 5.19)}$$

, we have:

$$\sigma^2_{\mathrm{N_{A1}-N_{B1}}} = \sigma^2_{\mathrm{N_{12}-N_{21}}} = \sigma^2_{N_{12}} + \sigma^2_{N_{21}} - 2\sigma_{N_{12},N_{21}} = n_t r_{AB} \qquad \text{(Equation 5.20)}$$

5.6.3 Estimation using standardized dosage data

After we standardize the allele dosage data with Eq. 5.5, we have

$$n'_{A1,\mathrm{i}} = (n_{A1,\mathrm{i}} - \frac{n_{t,i}}{2})/\sqrt{n_{t,i}} \qquad \text{(Equation 5.21)}$$

, where $\frac{n_{t,i}}{2}$ is the expectation of $n_{A1,i}$. Based on Eq. 5.16, the standardized allelic dosage for all different sperm samples follows the same normal distribution:

$$\binom{N'_{A1}}{N'_{B1}} \sim \mathrm{Normal}\left[0, \frac{1}{4}\begin{pmatrix} 1 & 1-2r_{AB} \\ 1-2r_{AB} & 1 \end{pmatrix}\right] \qquad \text{(Equation 5.22)}$$

The correlation can then be calculated as:

$$\rho_{N\prime_{A1},N\prime_{B1}} = \frac{\sigma_{N\prime_{A1},N\prime_{B1}}}{\sqrt{\sigma_{N\prime_{A1}}\sigma_{N\prime_{B1}}}} = 1 - 2r \qquad \text{(Equation 5.23)}$$

And we will get a generalized estimator that can be used even when the sperm samples we used have different amounts of sperm cells:

$$r_{AB} = (1 - \rho_{N\prime_{A1},N\prime_{B1}})/2 \qquad \text{(Equation 5.24)}$$

In this case, the estimator in Method II can be modified as:

$$\sigma^2_{N\prime_{A1}-N\prime_{B1}} = \sigma^2_{N\prime_{12}-N\prime_{21}} = \sigma^2_{N\prime_{12}} + \sigma^2_{N\prime_{21}} - 2\sigma_{N\prime_{12},N\prime_{21}} = r_{AB} \qquad \text{(Equation 5.25)}$$

5.6.4 Variance partitioning model

When measurement error is present, the observed allele dosage of allele A1 and B1 is no longer the true allele dosage:

$$\binom{N\prime_{A1}}{N\prime_{B1}}_{observed} = \binom{N\prime_{A1}}{N\prime_{B1}}_{true} + \binom{N_{e,A1}}{N_{e,B1}} \qquad \text{(Equation 5.26)}$$

, where $N_{e,A1}$ and $N_{e,B1}$ is the random error introduced during DNA quantification for allele A1 and B1, respectively. In this case, we can calculate the relationship between $\left(\sigma^2_{N'_{A1}-N'_{B1}}\right)_{observed}$ and $\left(\sigma^2_{N'_{A1}-N'_{B1}}\right)_{true}$:

$$\left(\sigma^2_{N'_{A1}-N'_{B1}}\right)_{observed} = \sigma^2_{N'_{A1}-N'_{B1}+N_{e,A1}+N_{e,B1}} = \sigma^2_{N'_{A1}-N'_{B1}} + \sigma^2_{N_{e,A1}} + \sigma^2_{N_{e,B1}} \quad \text{(Equation 5.27)}$$

If we assume that the variance of random measurement error is homogeneous and denoted as σ^2_e. Then we have this variance partitioning model:

$$\left(\sigma^2_{N'_{A1}-N'_{B1}}\right)_{observed} = \left(\sigma^2_{N'_{A1}-N'_{B1}}\right)_{true} + 2\sigma^2_e \quad \text{(Equation 5.28)}$$

5.7 Ordering method used in the simulation experiment

Based on the estimated pairwise inter-marker recombination rates in a linkage group, we can find the optimal order of these markers. This kind of problem, which tries to minimize the sum of additive distances, can easily be mathematically transformed into a well-studied optimization problem: the travelling salesman problem (TSP). There are proven algorithms and software that we can borrow and use in the ordering of markers (Karp et al. 1996). Previous studies show that this strategy is an efficient way to order markers (Agarwala 2000; Everts-van der Wind et al. 2005). In this section, we introduce the analysis conducted in our fourth simulation experiment, as well as some background information about TSP that we used in our analysis.

In this study, we use the sum of adjacent distance (SAD) as the criterion, or called scoring function, to evaluate SNP orders:

$$SAD = \sum_{i=1}^{n-1} d_{i,i+1} \qquad \text{(Equation 5.29)}$$

, where n is the number of SNPs, and $d_{i,i+1}$ is the inter-marker genetic distance between i^{th} and i+1th SNP. The sum of adjacent distance is a widely-used scoring function as it is "additive" (Xu 2013), which is necessary for the ordering algorithm we used here. For this scoring function, the order with a minimum SAD value is the best (Olson and Boehnke 1990). For a linkage map, the additive genetic distance between SNPs can be calculated with the recombination rates between them ($r_{i,i+1}$) using map functions. In the study, we choose Haldane's map function as we assume no interference between loci in our simulation studies (Liu 1998). So we have:

$$d_{i,i+1} = \frac{ln(1-2r_{i,i+1})}{2} \qquad \text{(Equation 5.30)}$$

After the inter-marker genetic distance was calculated, we transformed our ordering problem to TSP. TSP is usually described as a problem to find a round trip with a minimum total length visiting each of a set of cities exactly once (Reinelt 1991). If we set the inter-city distance in TSP as the inter-marker genetic distance in our analysis, the optimization process would be quite similar, except that the order of SNP is a linear path rather than a cycle. To overcome this difference, we introduced a dummy SNP. This is a common strategy in TSP-related problems, and has been applied in constructing RH maps (Karp et al. 1996). The distance between this dummy SNP and any other SNP was artificially set as 0, so that the dummy SNP can serve as a "bridge" to connect any other two SNPs. Then our ordering problem was equivalent to a TSP with $n+1$ cities.

Solution to the TSP was then found with CONCORDE, a very effective package for solving TSP (http://www.math.uwaterloo.ca/tsp/concorde/index.html).

Chapter 6 General conclusion

Nate Silver has emphasized two points in his book, *The Signal and the Noise*, for people mining gold from the increasing volume of data: 1) good interpretation of complex data requires sound understanding of the data and the context, as well as appropriate statistical methods; and 2) we make progress when we make more attempts and learn from the results (Silver 2012). In my thesis work, we aimed to better interpret available high-throughput genomic data into biologically meaningful conclusions, by integrating biological information from various sources (e.g. different types of data) and knowledge in different areas (e.g. genomics, statistics and animal science), and attempted to make better use of high-throughput genomic data in the genetic analysis in pigs. We worked toward this goal in the four studies in my thesis work:

1. For the studies about GWAS (Chapter 2-4), we discussed the issue about detection power and precision due to the high dimension of genomic data.
    a. In the study about the adaptive LASSO (Chapter 2), we assessed its performance and discussed how it may help to increase the detection power in GWAS.
    b. In the GWAS of fetal response to type 2 PRRSV challenge (Chapter 3), we made use of permutation to improve the precision of the results, and used transcriptomic data to further scan for candidate genes.
    c. In the TDT study (Chapter 4), we improved the precision of GWAS by integrating raw genotyping data (fluorescence intensity data) into the analysis.
2. In the study about multiple sperm typing (Chapter 5), we proposed a new model for the allele dosage data of haploids, in order to improve the efficiency of linkage map construction.

This chapter aims to summarize the four studies in my thesis work, and discuss how these studies may help toward the goal. In the following sections, I first give a brief summary of the four studies, as well as our effort to improve the analysis of genomic data. Then I discuss the significance of these studies in the context of genetic improvement, followed with some thoughts on additional outputs and future work.

6.1 Summary

Chapter 2 of the thesis discusses the application of the adaptive LASSO in GWAS, and its advantage in detection power. We first applied the adaptive LASSO to a public simulated dataset from QTLMAS 2010, and compared it with the methods previously published in QTLMAS 2010 (e.g. BayesC), in terms of detection power and precision. The adaptive LASSO outperformed other methods in detection power, though it had a lower precision. We then applied this strategy in a GWAS of meat colour in commercial crossbred pigs, as previous results using Bayesian approaches were disappointing in terms of the number of regions found to be associated with the traits of interest. A total of 20 genomic regions were detected and confirmed, with significant SNPs (FDR < 0.1) found in all the candidate regions. The results showed support for three candidate regions detected previously using the same population, while extending the number of candidate regions for meat colour in this dataset. Seventeen of the 20 regions overlapped with known QTL for meat colour in pigs. Twelve genes within the candidate regions were found to be functionally related to meat colour and close to the significant SNPs. Related biological pathways included muscle fibre type composition, glucose metabolism, and lipid metabolism.

The emergence of high-throughput genotyping technology significantly increases the density of genotype data. While high-density genotype data enables the analysis on a genome

scale, its high dimension causes new challenges. Here we discussed one of the challenges, how to find the optimal point in the trade-off between detection power and precision in GWAS. Our results show that the adaptive LASSO may be an option for GWAS when higher detection power is desired, such as searching for QTL with small-to-medium effects. In the GWAS of meat colour in pigs, we also tried to control false discoveries through an additional single marker association. Since meat colour traits are genetically correlated, our strategy of only focusing on candidate genomic regions associated with multiple traits may also help to filter out false discoveries. The analysis did extend the candidate genomic regions for meat colour in this dataset, and our results were supported by those obtained by previous studies (e.g. known QTL and candidate genes). We comment further on the LASSO below in relation to Chapter 3.

In Chapter 3, we introduce a GWAS using data from 928 fetuses from pregnant gilts experimentally challenged with type 2 PRRSV. Fetuses were assessed for four traits, viral load in thymus (VLT), viral load in endometrium (VLE), fetal death (FD) and fetal viability (FV). They were also genotyped at a medium density (61,565 SNPs). Collectively, 21 candidate genomic regions were found in association with these traits, seven of which overlap with previously reported QTLs for pig health and reproduction. A comparison with ongoing and related transcriptomic analyses of fetal response to PRRSV infection found differentially expressed genes within 17 of the candidate regions. Some of these genes have been reported to have immune system functions and contribute to host response to PRRSV infection in previous studies. The results provide new evidence about the genetic basis of fetal response to PRRSV challenge, and may ultimately lead to alternative control strategies to reduce the impact of reproductive PRRS.

Even though the adaptive LASSO showed improved detection power in chapter 2, it failed to provide a quantitative measurement (such as P-value) for the probability that the detected associations were actually true. Such measurement can help to determine the relative importance of multiple associations (i.e. whether some associations are more likely to be truly positive rather than others or not). In the study descripted in Chapter 3, we applied a permutation technique to correct the "overestimation" behaviour of the original LASSO. Permutation tests take into account the structure of datasets, and provide empirical significance test for association analysis, even when the dimension of genomic data is very high. Results of a related transcriptomic analysis were also integrated into the screening of candidate genes, which provides functional evidence for the detected associations in the GWAS.

In Chapter 4, we conducted a TDT of fetal autolysis using data obtained from the same project as Chapter 3 (Pregnant gilt model, PGM). Seven SNPs showed significant transmission disequilibrium and passed a pedigree-based genotype quality check. However, follow-up analysis using raw genotyping data suggested that five of the seven SNPs may be mis-clustered in genotyping the parents, which may result in false transmission disequilibrium for the five SNPs. The other two candidate SNPs showed transmission disequilibrium in the study may be evaluated for further confirmation.

In PGM, genomic data were unavailable at first for those autolyzed fetuses in the project, which were deemed to be most vulnerable to the challenge. The information about the phenotypes (autolysis) carried by these animals might be lost in GWAS if we simply removing them from the analysis. Through TDT, we gained ability to exploit these autolyzed fetuses. Pedigree information was integrated into the analysis and used for genotype checking.

Raw genotyping data (e.g. fluorescence intensity data) carry much information about genotyping quality, but this quality information is rarely used after genotypes are called. Sporadic erroneous genotypes will not be a problem in most genomic analysis. Some statistical tools such as TDT, however, are more vulnerable to erroneous genotypes. In the study descripted in Chapter 4, we used a likelihood-based method to analyse the raw genotyping data, and detected suspicious SNPs that might be mis-clustered during genotype calling. It is likely that some incorrect genotypes passed the pedigree-based genotype quality check in our analysis, and caused false positives in the TDT. Integrating raw genotyping data into association studies may help to assessing the quality of genotype data and to improve the precision of the results (i.e. reduce false discoveries). For example, one possible way is using a normal mixture model to describe the genotyping results instead of only genotypes (Hackett et al. 2013).

In Chapter 5, we proposed a method for inter-marker recombination rate estimation and genome linkage mapping using allele dosage data. The method can serve as a generalization of traditional single sperm typing method, while being able to work with a much higher throughput. Simulation experiments provide support for our hypothesis that this new method is able to significantly improve the accuracy of inter-marker recombination rate estimation. Further optimization through adjusting random errors is possible. This method was compared to other approaches such as radiation hybrid (RH) mapping and HAPPY mapping.

The allele dosage data of haploids (sperm cells in Chapter 5) have never been used to estimate inter-marker recombination rate, Possible reasons include the lack of high-throughput DNA quantification technology in the past, and that the statistical model for the allele dosage data is not as explicit as those models widely used before (e.g. single sperm typing method or pedigree-

based method). However, with the development in new biotechnologies (e.g. DNA microarray, next generation sequencing), allele dosage data are getting more accessible. The study described in Chapter 5 may help to make better use of the allele dosage data, and provide new opportunities in constructing high-resolution, low-cost genome maps with high-throughput, and may thereby lead to further application in genomic studies. These potential applications including genome-wide genotyping of CSVs, low-cost genome assembly and estimation of recombination rates on an individual level.

6.2 Significance in genetic improvement

One output of this thesis work is the potential of genomics in animal genetic improvement, especially for pigs in terms of health and product quality. The candidate genomic regions detected in the GWAS in Chapter 2 and Chapter 3 may provide new evidence about the genetic basis of the traits under investigations (meat colour and host response to PRRSV challenge) in pigs. Genetic markers related to the gene PRKAG3 and gene RYR1 (or CRC1) has been successfully applied in the animal selection for better meat quality, and our study may potentially provide more materials for marker-assisted selection in the future. For host response to PRRSV challenge, genomic tools can be an important alternative opportunity to control PRRS, as developing effective vaccines against a wide range of PRRSV strains is still a challenge at present. Our result may provide new genetic markers that can help select for animals resistant, or with reduced susceptibility to PRRSV. In addition, the genomic candidate regions detected in our analysis may be integrated into genomic selection in the future, through statistical tools that allow the integration of prior knowledge, such as information on causative mutations (e.g. BayesRC).

The statistical model proposed in Chapter 5 may serve as a new way to construct high-resolution genome maps, which may provide additional materials for genetic improvement. The benefit is twofold: 1) high-resolution genome maps may enable genome-wide genotyping of CSVs. Including these new genetic markers may help capture the phenotypic variations determined by them. It may help to discover the "missing heritability" due to not including CSVs; and 2) Linkage maps on an individual level enable the measurement of the recombination rate of single animals. As discussed in Chapter 1, that may help to determine animals tending to have more genetic recombination events during meiosis, which may potentially speed up the progress of genetic improvement.

6.3 Additional outputs and implications

1. The discussion in Chapter 2 may provide support for the application of the adaptive LASSO in GWAS, especially when high detection power is desired.
2. The results in Chapter 4 implies the possibility to integrate raw genotyping data into GWAS, in order to reduce false positives due to mis-clustering in genotyping for platforms like Illumina BeadChip.
3. The new linkage mapping method introduced in Chapter 5 may help construct linkage maps with desired cost, which may apply in human medicine and animal genomics.

6.4 Future work

1. The result in Chapter 2 shows relatively high detection power of the adaptive LASSO, while it suffered from a relatively low precision. Some extensions in the LASSO family has been reported to have better control of the precision issue, such as multi-step adaptive

elastic net (Xiao and Xu 2015). Assessing their performance in GWAS may help explore more statistical options in genomic data analysis.

2. The result in Chapter 3 shows new candidate genomic regions possibly associated with fetal response to type 2 PRRSV. These candidate regions may need to be validated in other populations. In addition, by sequencing the candidate regions and scanning new genetic variants, it is possible to discover causative mutations for the fetal response.
3. The raw genotyping data in the Illumina BeadChip platform may be directly integrated into GWAS through a normal mixture model (Hackett et al. 2013). It may help to taken the clustering quality into account in the analysis.
4. As discussed in Chapter 5, challenges remain in reducing the new statistical model for constructing linkage maps to practice. The development and optimization of experiment protocol as well as the development of analysis toolkit needs further studies in the future.

6.5 Prospects

Great changes are happening in both genomics and statistics, which show us brand new opportunities of applying genomic data in more and more areas, such as animal production and human medicine. In terms of genomics, several large projects, such as ENCODE (ENCODE Project Consortium 2012) and FAANG (Andersson et al. 2015), are developing deeper and deeper understanding about how the genome regulates biological processes in different species. In addition, more powerful tools for genome engineering are emerging, such as CRISPR (Jinek et al. 2012; Cong et al. 2013). In the meanwhile, even though the complexity of the data keeps increasing, fortunately, tools in statistics and data science are also developing rapidly to meet those new challenges in data analysis, with many of them already successfully applying in the analysis

of genomic data in animals (e.g. Bayesian methods), or being evaluated for possible application (González-Recio and Forni 2011).

The connection of the changes in both animal genomics and statistics may be the key to gold mining from animal genomic data. A sound understanding of questions in practical production has been driving the development of statistical tools (e.g. Student's t-test), while introducing or generalizing existing statistical tools invented for other areas for a new application may significantly improve the outcomes (e.g. the wide application of BLUP and solutions to "travelling salesman problem"). The emergence of high-throughput genomic data is challenging existing strategies of analysis in terms of both genomic and statistics. Potential solutions to the challenges may be involved in topics across an even more broader spectrum from biology (e.g. biological processes underlying a trait, collection of phenotypes, and the depth of phenotypes) to statistics (e.g. experimental design, statistical modelling, and statistical prediction methods) and informatics (e.g. storing and processing biological data and computational algorithms). My thesis work was directly involved in a few interesting statistical topics in the context of genomics, with some discussions related to biology, animal science and bioinformatics, and hopefully had some outputs in terms of animal production. In the meanwhile, my Ph.D. program has given me an appreciation of not only those topics involving my thesis work, but also knowledge across the broad spectrum from biology to statistics.

References

Andersson, L., Archibald, A.L., Bottema, C.D., Brauning, R., Burgess, S.C., Burt, D.W., Casas, E., Cheng, H.H., Clarke, L., Couldrey, C., Dalrymple, B.P., Elsik, C.G., Foissac, S., Giuffra,

E., Groenen, M. a, Hayes, B.J., Huang, L.S., Khatib, H., Kijas, J.W., Kim, H., Lunney, J.K., McCarthy, F.M., McEwan, J.C., Moore, S., Nanduri, B., Notredame, C., Palti, Y., Plastow, G.S., Reecy, J.M., Rohrer, G. a, Sarropoulou, E., Schmidt, C.J., Silverstein, J., Tellam, R.L., Tixier-Boichard, M., Tosser-Klopp, G., Tuggle, C.K., Vilkki, J., White, S.N., Zhao, S., Zhou, H., and FAANG Consortium 2015. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol. **16**: 57.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science **339**: 819–23.

ENCODE Project Consortium 2012. An integrated encyclopedia of DNA elements in the human genome. Nature **489**: 57–74.

González-Recio, O., and Forni, S. 2011. Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. Genet. Sel. Evol. **43**: 7.

Hackett, C.A., McLean, K., and Bryan, G.J. 2013. Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PLoS One **8**: e63939.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337**: 816–21.

Silver, N. 2012. The signal and the noise : why so many predictions fail--but some don’t. Penguin Press, New York, NY.

Bibliography

Aasted, B., Bach, P., Nielsen, J., and Lind, P. 2002. Cytokine profiles in peripheral blood mononuclear cells and lymph node cells from piglets infected in utero with porcine reproductive and respiratory syndrome virus. Clin. Diagn. Lab. Immunol. 9: 1229–34.

Agace, W.W. 2006. Tissue-tropic effector T cells: generation and targeting opportunities. Nat. Rev. Immunol. 6: 682–692.

Agarwala, R. 2000. A Fast and Scalable Radiation Hybrid Map Construction and Integration Strategy. Genome Res. 10: 350–364.

Akira, S., and Takeda, K. 2004. Toll-like receptor signalling. Nat. Rev. Immunol. 4: 499–511.

Alamuru, N.P., Behera, S., Butchar, J.P., Tridandapani, S., Kaimal Suraj, S., Babu, P.P., Hasnain, S.E., Ehtesham, N.Z., and Parsa, K.V.L. 2014. A novel immunomodulatory function of PHLPP1: inhibition of iNOS via attenuation of STAT1 ser727 phosphorylation in mouse macrophages. J. Leukoc. Biol. 95: 775–783.

Aldhous, P. 1993. Managing the genome data deluge. Science 262: 502–3.

Alkan, C., Coe, B.P., and Eichler, E.E. 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12: 363–376.

Anderson, S.K., Gallinger, S., Roder, J., Frey, J., Young, H.A., and Ortaldo, J.R. 1993. A cyclophilin-related protein involved in the function of natural killer cells. Proc. Natl. Acad. Sci. U. S. A. 90: 542–546.

Andersson, L., Archibald, A.L., Bottema, C.D., Brauning, R., Burgess, S.C., Burt, D.W., Casas, E., Cheng, H.H., Clarke, L., Couldrey, C., Dalrymple, B.P., Elsik, C.G., Foissac, S., Giuffra, E., Groenen, M. a, Hayes, B.J., Huang, L.S., Khatib, H., Kijas, J.W., Kim, H., Lunney, J.K., McCarthy, F.M., McEwan, J.C., Moore, S., Nanduri, B., Notredame, C., Palti, Y., Plastow,

G.S., Reecy, J.M., Rohrer, G. a, Sarropoulou, E., Schmidt, C.J., Silverstein, J., Tellam, R.L., Tixier-Boichard, M., Tosser-Klopp, G., Tuggle, C.K., Vilkki, J., White, S.N., Zhao, S., Zhou, H., and FAANG Consortium 2015. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol. 16: 57.

Archibald, A.L., Bolund, L., Churcher, C., Fredholm, M., Groenen, M. a M., Harlizius, B., Lee, K.-T., Milan, D., Rogers, J., Rothschild, M.F., Uenishi, H., Wang, J., and Schook, L.B. 2010. Pig genome sequence--analysis and publication strategy. BMC Genomics 11: 438.

Archibald, A.L., Haley, C.S., Brown, J.F., Couperwhite, S., McQueen, H.A., Nicholson, D., Coppieters, W., Van de Weghe, A., Stratil, A., Winterø, A.K., Fredholm, M., Larsen, N.J., Nielsen, V.H., Milan, D., Woloszyn, N., Robic, A., Dalens, M., Riquet, J., Gellin, J., Caritez, J.C., Burgaud, G., Ollivier, L., Bidanel, J.P., Vaiman, M., Renard, C., Geldermann, H., Davoli, R., Ruyter, D., Verstege, E.J.M., Groenen, M.A.M., Davies, W., Høyheim, B., Keiserud, A., Andersson, L., Ellegren, H., Johansson, M., Marklund, L., Miller, J.R., Anderson Dear, D. V., Signer, E., Jeffreys, A.J., Moran, C., Le Tissier, P., Muladno, Rothschild, M.F., Tuggle, C.K., Vaske, D., Helm, J., Liu, H.C., Rahman, A., Yu, T.P., Larson, R.G., and Schmitz, C.B. 1995. The PiGMaP consortium linkage map of the pig (Sus scrofa). Mamm. Genome 6: 157–175.

Arnheim, N., Li, H., and Cui, X. 1991. Genetic mapping by single sperm typing. Anim. Genet. 22: 105–115.

Auton, A., and McVean, G. 2007. Recombination rate estimation in the presence of hotspots. Genome Res. 17: 1219–27.

Ayala, J.E., Bracy, D.P., James, F.D., Julien, B.M., Wasserman, D.H., and Drucker, D.J. 2009. The glucagon-like peptide-1 receptor regulates endogenous glucose production and muscle glucose uptake independent of its incretin action. Endocrinology 150: 1155–1164.

Ayers, K.L., and Cordell, H.J. 2010. SNP selection in genome-wide and candidate gene studies via penalized logistic regression. Genet. Epidemiol. 34: 879–891.

Ayling, C.M., Moreland, B.H., Zanelli, J.M., and Schulster, D. 1989. Human growth hormone treatment of hypophysectomized rats increases the proportion of type-1 fibres in skeletal muscle. J. Endocrinol. 123: 429–35.

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

Barral, S., Haynes, C., Levenstien, M. a, and Gordon, D. 2005. Precision and type I error rate in the presence of genotype errors and missing parental data: a comparison between the original transmission disequilibrium test (TDT) and TDTae statistics. BMC Genet. 6: S150.

Basu, S., Pan, W., Shen, X., and Oetting, W.S. 2011. Multilocus association testing with penalized regression. Genet. Epidemiol. 35: 755–765.

Bateson, W., Saunders, E.R., and Punnett, R.C. 1905. Experimental studies in the physiology of heredity. Reports to Evol. Comm. R. Soc. 2.

Battagin, M., Gorjanc, G., Faux, A.-M., Johnston, S.E., and Hickey, J.M. 2016. Effect of manipulating recombination rates on response to selection in livestock breeding programs. Genet. Sel. Evol. 48: 44.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B. Stat. Methodol. 57: 289–300.

Berk, R.A. 2008. Statistical learning from a regression perspective. Springer New York, New York, NY.

Bickhart, D.M., and Liu, G.E. 2014. The challenges and importance of structural variation detection in livestock. Front. Genet. 5: 1–14.

Bloch, D.B., Nakajima, A., Gulick, T., Chiche, J.D., Orth, D., de La Monte, S.M., and Bloch, K.D. 2000. Sp110 localizes to the PML-Sp100 nuclear body and may function as a nuclear hormone receptor transcriptional coactivator. Mol. Cell. Biol. 20: 6138–46.

Boddicker, N., Waide, E.H., Rowland, R.R.R., Lunney, J.K., Garrick, D.J., Reecy, J.M., and Dekkers, J.C.M. 2012. Evidence for a major QTL associated with host response to Porcine Reproductive and Respiratory Syndrome Virus challenge. J. Anim. Sci. 90: 1733–1746.

Boddicker, N.J., Bjorkquist, A., Rowland, R.R., Lunney, J.K., Reecy, J.M., and Dekkers, J.C. 2014. Genome-wide association and genomic prediction for host response to porcine reproductive and respiratory syndrome virus infection. Genet. Sel. Evol. 46: 18.

Bovine HapMap Consortium, Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse, W., Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M., Lien, S., Matukumalli, L.K., McEwan, J.C., Nazareth, L. V., Schnabel, R.D., Weinstock, G.M., Wheeler, D.A., Ajmone-Marsan, P., Boettcher, P.J., Caetano, A.R., Garcia, J.F., Hanotte, O., Mariani, P., Skow, L.C., Sonstegard, T.S., Williams, J.L., Diallo, B., Hailemariam, L., Martinez, M.L., Morris, C.A., Silva, L.O.C., Spelman, R.J., Mulatu, W., Zhao, K., Abbey, C.A., Agaba, M., Araujo, F.R., Bunch, R.J., Burton, J., Gorni, C., Olivier, H., Harrison, B.E., Luff, B., Machado, M.A., Mwakaya, J., Plastow, G., Sim, W., Smith, T., Thomas, M.B., Valentini, A., Williams, P.,

Womack, J., Woolliams, J.A., Liu, Y., Qin, X., Worley, K.C., Gao, C., Jiang, H., Moore, S.S., Ren, Y., Song, X.-Z., Bustamante, C.D., Hernandez, R.D., Muzny, D.M., Patil, S., San Lucas, A., Fu, Q., Kent, M.P., Vega, R., Matukumalli, A., McWilliam, S., Sclep, G., Bryc, K., Choi, J., Gao, H., Grefenstette, J.J., Murdoch, B., Stella, A., Villa-Angulo, R., Wright, M., Aerts, J., Jann, O., Negrini, R., Goddard, M.E., Hayes, B.J., Bradley, D.G., Barbosa da Silva, M., Lau, L.P.L., Liu, G.E., Lynn, D.J., Panzitta, F., and Dodds, K.G. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324: 528–32.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S.B., Krstic, P.S., Lindsay, S., Ling, X.S., Mastrangelo, C.H., Meller, A., Oliver, J.S., Pershin, Y. V, Ramsey, J.M., Riehn, R., Soni, G. V, Tabard-Cossa, V., Wanunu, M., Wiggin, M., and Schloss, J.A. 2008. The potential and challenges of nanopore sequencing. Nat. Biotechnol. 26: 1146–1153.

Bühlmann, P., and van de Geer, S. 2011. Statistics for High-Dimensional Data. Springer Berlin Heidelberg, Berlin, Heidelberg.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11: 94.

Burke, S.M., Issekutz, T.B., Mohan, K., Lee, P.W.K., Shmulevitz, M., and Marshall, J.S. 2008. Human mast cell activation with virus-associated stimuli leads to the selective chemotaxis of natural killer cells by a CXCL8-dependent mechanism. Blood 111: 5467–5476.

Butler, J.E., Lager, K.M., Golde, W., Faaberg, K.S., Sinkora, M., Loving, C., and Zhang, Y.I. 2014. Porcine reproductive and respiratory syndrome (PRRS): an immune dysregulatory pandemic. Immunol. Res. 59: 81–108.

Calzada-Nova, G., Schnitzlein, W., Husmann, R., and Zuckermann, F.A. 2010. Characterization of the cytokine and maturation responses of pure populations of porcine plasmacytoid dendritic cells to porcine viruses and toll-like receptor agonists. Vet. Immunol. Immunopathol. 135: 20–33.

Campbell, D.J., and Ziegler, S.F. 2007. FOXP3 modifies the phenotypic and functional properties of regulatory T cells. Nat. Rev. Immunol. 7: 305–310.

Chaisson, M.J.P., Wilson, R.K., and Eichler, E.E. 2015. Genetic variation and the de novo assembly of human genomes. Nat. Rev. Genet. 16: 627–640.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7.

Chang, K.C., da Costa, N., Blackley, R., Southwood, O., Evans, G., Plastow, G., Wood, J.D., and Richardson, R.I. 2003. Relationships of myosin heavy chain fibre types to meat quality traits in traditional and modern pigs. Meat Sci. 64: 93–103.

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J.S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.-C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B.R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R.J., and Georges, M. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. Nat. Genet. 40: 449–54.

Cheema, J., and Dicks, J. 2009. Computational approaches and software tools for genetic linkage map estimation in plants. Brief. Bioinform. 10: 595–608.

Chen, Y.-C., Wu, B.-K., Chu, C.-Y., Cheng, C.-H., Han, H.-W., Chen, G.-D., Lee, M.-T., Hwang, P.-P., Kawakami, K., Chang, C.-C., and Huang, C.-J. 2010. Identification and

characterization of alternative promoters of zebrafish Rtn-4/Nogo genes in cultured cells and zebrafish embryos. Nucleic Acids Res. 38: 4635–4650.

Choe, J.H., Choi, Y.M., Lee, S.H., Shin, H.G., Ryu, Y.C., Hong, K.C., and Kim, B.C. 2008. The relation between glycogen, lactate content and muscle fiber type composition, and their influence on postmortem glycolytic rate and pork quality. Meat Sci. 80: 355–362.

Chunyan Zhang, Heather Bruce, Tianfu Yang, Patrick Charagu, Robert Alan Kemp, Nicholas Boddicker, Younes Miar, Zhiquan Wang, and Graham Plastow. 2016. "Genome Wide Association Studies (GWAS) Identify QTL on SSC2 and SSC17 Affecting Loin Peak Shear Force in Crossbred Commercial Pigs." Edited by Jian-Feng Liu. PloS One 11 (2): e0145082.

Chunyan Zhang, Zhiquan Wang, Heather Bruce, Robert Alan Kemp, Patrick Charagu, Younes Miar, Tianfu Yang, and Graham Plastow. 2015. "Genome-Wide Association Studies (GWAS) Identify a QTL close to PRKAG3 Affecting Meat pH and Colour in Crossbred Commercial Pigs." BMC Genetics 16 (1): 33.

Cillari, E., Milano, S., and Pathology, G. 1992. Thymopentin reduces the susceptibility of aged mice to cutaneous leishmaniasis by modulating CD4 T-cell subsets. Immunology 76: 362–366.

Ciobanu, D., Bastiaansen, J., Malek, M., Helm, J., Woollard, J., Plastow, G., and Rothschild, M. 2001. Evidence for new alleles in the protein kinase adenosine monophosphate-activated gamma(3)-subunit gene associated with low glycogen content in pig skeletal muscle and improved meat quality. Genetics 159: 1151–62.

Clarke, N.F., Ilkovski, B., Cooper, S., Valova, V.A., Robinson, P.J., Nonaka, I., Feng, J.-J., Marston, S., and North, K. 2007. The pathogenesis of ACTA1-related congenital fiber type disproportion. Ann. Neurol. 61: 552–561.

Close, A.F., Rouillard, C., and Buteau, J. 2013. NR4A orphan nuclear receptors in glucose homeostasis: A minireview. Diabetes Metab. 39: 478–484.

Collins, F.S. 1999. Shattuck lecture--medical and societal consequences of the Human Genome Project. N. Engl. J. Med. 341: 28–37.

Collins, F.S., and McKusick, V.A. 2001. Implications of the Human Genome Project for medical science. JAMA 285: 540–4.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339: 819–23.

Cowan, C.M., Dentine, M.R., Ax, R.L., and Schuler, L.A. 1990. Structural variation around prolactin gene linked to quantitative traits in an elite Holstein sire family. : 577–582.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S., and Myers, R.M. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250: 245–50.

Crick, F.H.C., Barnett, L., Brenner, S., and Watts-Tobin, R.J. 1961. General nature of the genetic code for proteins. Nature 192: 1227–1232.

Crosslin, D.R., McDavid, A., Weston, N., Zheng, X., Hart, E., de Andrade, M., Kullo, I.J., McCarty, C.A., Doheny, K.F., Pugh, E., Kho, A., Hayes, M.G., Ritchie, M.D., Saip, A., Crawford, D.C., Crane, P.K., Newton, K., Carrell, D.S., Gallego, C.J., Nalls, M.A., Li, R., Mirel, D.B., Crenshaw, A., Couper, D.J., Tanaka, T., van Rooij, F.J.A., Chen, M.-H., Smith, A. V., Zakai, N.A., Yango, Q., Garcia, M., Liu, Y., Lumley, T., Folsom, A.R., Reiner, A.P., Felix, J.F., Dehghan, A., Wilson, J.G., Bis, J.C., Fox, C.S., Glazer, N.L., Cupples, L.A., Coresh, J., Eiriksdottir, G., Gudnason, V., Bandinelli, S., Frayling, T.M., Chakravarti, A., van

Duijn, C.M., Melzer, D., Levy, D., Boerwinkle, E., Singleton, A.B., Hernandez, D.G., Longo, D.L., Witteman, J.C.M., Psaty, B.M., Ferrucci, L., Harris, T.B., O'Donnell, C.J., Ganesh, S.K., Larson, E.B., Carlson, C.S., and Jarvik, G.P. 2013. Genetic variation associated with circulating monocyte count in the eMERGE Network. Hum. Mol. Genet. 22: 2119–2127.

Cui, X.F., Li, H.H., Goradia, T.M., Lange, K., Kazazian, H.H., Galas, D., and Arnheim, N. 1989. Single-sperm typing: determination of genetic distance between the G gamma-globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. Proc. Natl. Acad. Sci. U. S. A. 86: 9389–93.

de Bakker, P.I.W., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J., and Altshuler, D. 2005. Efficiency and power in genetic association studies. Nat. Genet. 37: 1217–1223.

De Donato, M., Peters, S.O., Mitchell, S.E., Hussain, T., and Imumorin, I.G. 2013. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PLoS One 8: e62137.

de Vries, A.G., Faucitano, L., Sosnicki, A., and Plastow, G.S. 2000. The use of gene technology for optimal development of pork meat quality. Food Chem. 69: 397–405.

Dear, P.H. 2001. Genome Mapping. in Encyclopedia of Life Sciences.

Dear, P.H., and Cook, P.R. 1993. Happy mapping: linkage mapping using a physical analogue of meiosis. Nucleic Acids Res. 21: 13–20.

Dear, P.H., Bankier, A.T., and Piper, M.B. 1998. A high-resolution metric HAPPY map of human chromosome 14. Genomics 48: 232–41.

Decker, J.E. 2015. Agricultural Genomics: Commercial Applications Bring Increased Basic Research Power. PLOS Genet. 11: e1005621.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. 82 E-Suppl: E313-328.

Dhanasekaran, D.N., and Reddy, E.P. 2008. JNK signaling in apoptosis. Oncogene 27: 6245–6251.

Dong, L., Meng, Y., Wang, J., and Liu, Y. 2014. Evaluation of droplet digital PCR for characterizing plasmid reference material used for quantifying ammonia oxidizers and denitrifiers. Anal. Bioanal. Chem. 406: 1701–1712.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J., and Wang, W. 2012. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat. Biotechnol. 31: 135–141.

Douglas, J.A., Skol, A.D., and Boehnke, M. 2002. Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. Am. J. Hum. Genet. 70: 487–95.

Doyle, M.E., and Egan, J.M. 2007. Mechanisms of action of glucagon-like peptide 1 in the pancreas. Pharmacol. Ther. 113: 546–593.

Duan, X., Nauwynck, H.J., and Pensaert, M.B. 1997. Effects of origin and state of differentiation and activation of monocytes/macrophages on their susceptibility to porcine reproductive and respiratory syndrome virus (PRRSV). Arch. Virol. 142: 2483–2497.

Duggal, P., Gillanders, E.M., Holmes, T.N., and Bailey-Wilson, J.E. 2008. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics 9: 516.

Dunne, P.G., Monahan, F.J., O'Mara, F.P., and Moloney, A.P. 2009. Colour of bovine subcutaneous adipose tissue: A review of contributory factors, associations with carcass and meat quality and its potential utility in authentication of dietary history. Meat Sci. 81: 28–45.

Dupuis, J., and O'Donnell, C.J. 2007. Interpreting results of large-scale genetic association studies. JAMA 297: 529.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21: 3439–3440.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4: 1184–1191.

Durkin, K., Coppieters, W., Drögemüller, C., Ahariz, N., Cambisano, N., Druet, T., Fasquelle, C., Haile, A., Horin, P., Huang, L., Kamatani, Y., Karim, L., Lathrop, M., Moser, S., Oldenbroek, K., Rieder, S., Sartelet, A., Sölkner, J., Stålhammar, H., Zelenika, D., Zhang, Z., Leeb, T., Georges, M., and Charlier, C. 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. Nature 482: 81–4.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D.P., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C.S., Bowman, R., SEARCH collaborators, Meyer, K.B., Haiman, C.A., Kolonel, L.K., Henderson, B.E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C.-Y., Wu, P.-E., Wang, H.-C., Eccles, D., Evans, D.G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M.R., Rahman, N., Chenevix-Trench, G., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska,

B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K.-Y., Noh, D.-Y., Ahn, S.-H., Hunter, D.J., Hankinson, S.E., Cox, D.G., Hall, P., Wedren, S., Liu, J., Low, Y.-L., Bogdanova, N., Schürmann, P., Dörk, T., Tollenaar, R.A.E.M., Jacobi, C.E., Devilee, P., Klijn, J.G.M., Sigurdson, A.J., Doody, M.M., Alexander, B.H., Zhang, J., Cox, A., Brock, I.W., MacPherson, G., Reed, M.W.R., Couch, F.J., Goode, E.L., Olson, J.E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J.L., McCredie, M., Southey, M., Giles, G.G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y.-D., Spurdle, A.B., Beesley, J., Chen, X., kConFab, AOCS Management Group, Mannermaa, A., Kosma, V.-M., Kataja, V., Hartikainen, J., Day, N.E., Cox, D.R., and Ponder, B.A.J. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087–93.

Eggert, J.M., Depreux, F.F.S., Schinckel, A.P., Grant, A.L., and Gerrard, D.E. 2002. Myosin heavy chain isoforms account for variation in pork quality. Meat Sci. 61: 117–126.

El-Metwally, S., Ouda, O.M., and Helmy, M. 2014. Next Generation Sequencing Technologies and Challenges in Sequence Assembly.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Muzny, D.M., Weinstock, G.M., Adelson, D.L., Eichler, E.E., Elnitski, L., Guigo, R., Hamernik, D.L., Kappes, S.M., Lewin, H.A., Lynn, D.J., Nicholas, F.W., Reymond, A., Rijnkels, M., Skow, L.C., Zdobnov, E.M., Schook, L., Womack, J., Alioto, T., Antonarakis, S.E., Astashyn, A., Chapple, C.E., Chen, H.-C., Chrast, J., Camara, F., Ermolaeva, O., Henrichsen, C.N., Hlavina, W., Kapustin, Y., Kiryutin,

B., Kitts, P., Kokocinski, F., Landrum, M., Maglott, D., Pruitt, K., Sapojnikov, V., Searle, S.M., Solovyev, V., Souvorov, A., Ucla, C., Wyss, C., Anzola, J.M., Gerlach, D., Elhaik, E., Graur, D., Reese, J.T., Edgar, R.C., McEwan, J.C., Payne, G.M., Raison, J.M., Junier, T., Kriventseva, E. V., Eyras, E., Plass, M., Donthu, R., Larkin, D.M., Reecy, J., Yang, M.Q., Chen, L., Cheng, Z., Chitko-McKown, C.G., Liu, G.E., Matukumalli, L.K., Song, J., Zhu, B., Bradley, D.G., Brinkman, F.S.L., Lau, L.P.L., Whiteside, M.D., Walker, A., Wheeler, T.T., Casey, T., German, J.B., Lemay, D.G., Maqbool, N.J., Molenaar, A.J., Seo, S., Stothard, P., Baldwin, C.L., Baxter, R., Brinkmeyer-Langford, C.L., Brown, W.C., Childers, C.P., Connelley, T., Ellis, S.A., Fritz, K., Glass, E.J., Herzig, C.T.A., Iivanainen, A., Lahmers, K.K., Bennett, A.K., Dickens, C.M., Gilbert, J.G.R., Hagen, D.E., Salih, H., Aerts, J., Caetano, A.R., Dalrymple, B., Garcia, J.F., Gill, C.A., Hiendleder, S.G., Memili, E., Spurlock, D., Williams, J.L., Alexander, L., Brownstein, M.J., Guan, L., Holt, R.A., Jones, S.J.M., Marra, M.A., Moore, R., Moore, S.S., Roberts, A., Taniguchi, M., Waterman, R.C., Chacko, J., Chandrabose, M.M., Cree, A., Dao, M.D., Dinh, H.H., Gabisi, R.A., Hines, S., Hume, J., Jhangiani, S.N., Joshi, V., Kovar, C.L., Lewis, L.R., Liu, Y. -s., Lopez, J., Morgan, M.B., Nguyen, N.B., Okwuonu, G.O., Ruiz, S.J., Santibanez, J., Wright, R.A., Buhay, C., Ding, Y., Dugan-Rocha, S., Herdandez, J., Holder, M., Sabo, A., Egan, A., Goodell, J., Wilczek-Boney, K., Fowler, G.R., Hitchens, M.E., Lozado, R.J., Moen, C., Steffen, D., Warren, J.T., Zhang, J., Chiu, R., Schein, J.E., Durbin, K.J., Havlak, P., Jiang, H., Liu, Y., Qin, X., Ren, Y., Shen, Y., Song, H., Bell, S.N., Davis, C., Johnson, A.J., Lee, S., Nazareth, L. V., Patel, B.M., Pu, L.-L., Vattathil, S., Williams, R.L., Curry, S., Hamilton, C., Sodergren, E., Wheeler, D.A., Barris, W., Bennett, G.L., Eggen, A., Green, R.D., Harhay, G.P., Hobbs, M., Jann, O., Keele, J.W., Kent, M.P., Lien, S., McKay, S.D., McWilliam, S., Ratnakumar,

A., Schnabel, R.D., Smith, T., Snelling, W.M., Sonstegard, T.S., Stone, R.T., Sugimoto, Y., Takasuga, A., Taylor, J.F., Van Tassell, C.P., MacNeil, M.D., Abatepaulo, A.R.R., Abbey, C.A., Ahola, V., Almeida, I.G., Amadio, A.F., Anatriello, E., Bahadue, S.M., Biase, F.H., Boldt, C.R., Carroll, J.A., Carvalho, W.A., Cervelatti, E.P., Chacko, E., Chapin, J.E., Cheng, Y., Choi, J., Colley, A.J., de Campos, T.A., De Donato, M., Santos, I.K.F. d. M., de Oliveira, C.J.F., Deobald, H., Devinoy, E., Donohue, K.E., Dovc, P., Eberlein, A., Fitzsimmons, C.J., Franzin, A.M., Garcia, G.R., Genini, S., Gladney, C.J., Grant, J.R., Greaser, M.L., Green, J.A., Hadsell, D.L., Hakimov, H.A., Halgren, R., Harrow, J.L., Hart, E.A., Hastings, N., Hernandez, M., Hu, Z.-L., Ingham, A., Iso-Touru, T., Jamis, C., Jensen, K., Kapetis, D., Kerr, T., Khalil, S.S., Khatib, H., Kolbehdari, D., Kumar, C.G., Kumar, D., Leach, R., Lee, J.C.-M., Li, C., Logan, K.M., Malinverni, R., Marques, E., Martin, W.F., Martins, N.F., Maruyama, S.R., Mazza, R., McLean, K.L., Medrano, J.F., Moreno, B.T., More, D.D., Muntean, C.T., Nandakumar, H.P., Nogueira, M.F.G., Olsaker, I., Pant, S.D., Panzitta, F., Pastor, R.C.P., Poli, M.A., Poslusny, N., Rachagani, S., Ranganathan, S., Razpet, A., Riggs, P.K., Rincon, G., Rodriguez-Osorio, N., Rodriguez-Zas, S.L., Romero, N.E., Rosenwald, A., Sando, L., Schmutz, S.M., Shen, L., Sherman, L., Southey, B.R., Lutzow, Y.S., Sweedler, J. V., Tammen, I., Telugu, B.P.V.L., Urbanski, J.M., Utsunomiya, Y.T., Verschoor, C.P., Waardenberg, A.J., Wang, Z., Ward, R., Weikard, R., Welsh, T.H., White, S.N., Wilming, L.G., Wunderlich, K.R., Yang, J., and Zhao, F.-Q. 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science. 324: 522–528.

ENCODE Project Consortium 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome J. 4: 250.

Estelle, J., Gil, F., Vazquez, J.M., Latorre, R., Ramirez, G., Barragan, M.C., Folch, J.M., Noguera, J.L., Toro, M.A., and Perez-Enciso, M. 2008. A quantitative trait locus genome scan for porcine muscle fiber traits reveals overdominance and epistasis. J. Anim. Sci. 86: 3290–3299.

Everts-van der Wind, A., Larkin, D.M., Green, C. a, Elliott, J.S., Olmstead, C. a, Chiu, R., Schein, J.E., Marra, M. a, Womack, J.E., and Lewin, H. a 2005. A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. Proc. Natl. Acad. Sci. U. S. A. 102: 18526–31.

Faaberg, K.S., Balasuriya, U.B., Brinton, M.A., Gorbalenya, A.E., Leung, F..-C., Nauwynck, H., Snijder, E.J., Stadejek, T., Yang, H., and Yoo, D. 2012. Arteriviridae. in A.M.Q. King, E. Lefkowitz, M.J. Adams, and E.B. Carstens, eds. Virus Taxonomy. Elsevier.

Falconer, D.S., and Mackay, T.F.C. 1996. Introduction to quantitative genetics. Fourth Edi.

Fan, B., Onteru, S.K., Du, Z.Q., Garrick, D.J., Stalder, K.J., and Rothschild, M.F. 2011. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. PLoS One 6: e14726.

Fearnhead, P., and Donnelly, P. 2001. Estimating recombination rates from population genetic data. Genetics 159: 1299–318.

Fernando, R.L., and Garrick, D. 2013. Bayesian Methods Applied to GWAS. in C. Gondro, J. van der Werf, and B. Hayes, eds. Genome-wide association studies and genomic prediction. Humana Press, Totowa, NJ.

Fierst, J.L. 2015. Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. Front. Genet. 6: 1–8.

Florini, J.R., Ewton, D.Z., and Coolican, S.A. 1996. Growth hormone and the insulin-like growth factor system in myogenesis. Endocr. Rev. 17: 481–517.

Ford, L.B., Cerovic, V., Milling, S.W.F., Graham, G.J., Hansell, C.A.H., and Nibbs, R.J.B. 2014. Characterization of conventional and atypical receptors for the chemokine CCL2 on mouse leukocytes. J. Immunol. 193: 400–411.

Frey, J.L., Bino, T., Kantor, R.R., Segal, D.M., Giardina, S.L., Roder, J., Anderson, S., and Ortaldo, J.R. 1991. Mechanism of target cell recognition by natural killer cells: characterization of a novel triggering molecule restricted to CD3- large granular lymphocytes. J. Exp. Med. 174: 1527–1536.

Frias-Staheli, N., Giannakopoulos, N. V, Kikkert, M., Taylor, S.L., Bridgen, A., Paragas, J., Richt, J.A., Rowland, R.R., Schmaljohn, C.S., Lenschow, D.J., Snijder, E.J., García-Sastre, A., and Virgin, H.W. 2007. Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. Cell Host Microbe 2: 404–16.

Friedman, J., Hastie, T., and Tibshirani, R. 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33: 124–31.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V., Weiler, J., O'Brien, P., and MacLennan, D. 1991. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science 253: 448–451.

Galina-Pantoja, L., Siggens, K., van Schriek, M.G.M., and Heuven, H.C.M. 2009. Mapping markers linked to porcine salmonellosis susceptibility. Anim. Genet. 40: 795–803.

Gao, H., Wu, Y., Li, J., Li, H., Li, J., and Yang, R. 2014. Forward LASSO analysis for high-order interactions in genome-wide association study. Brief. Bioinform. 15: 552–561.

Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D., and Province, M.A. 2009. Avoiding the high Bonferroni penalty in genome-wide association studies. Genet. Epidemiol. 34: 100–105.

Gately, M.K., Renzetti, L.M., Magram, J., Stern, A.S., Adorini, L., Gubler, U., and Presky, D.H. 1998. The interleukin-12/interleukin-12-receptor system: role in normal and pathologic immune responses. Annu. Rev. Immunol. 16: 495–521.

Gayen, A.K. 1951. The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. Biometrika 38: 219–47.

Genetic balancers. [Online] Available: http://www.wormbook.org/chapters/www_geneticbalancers/geneticbalancers.html.

Gentry, J.G., McGlone, J.J., Miller, M.F., and Blanton, J.R. 2004. Environmental effects on pig performance, meat quality, and muscle characteristics. J. Anim. Sci. 82: 209.

Giglio, S., Broman, K.W., Matsumoto, N., Calvari, V., Gimelli, G., Neumann, T., Ohashi, H., Voullaire, L., Larizza, D., Giorda, R., Weber, J.L., Ledbetter, D.H., and Zuffardi, O. 2001. Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am. J. Hum. Genet. 68: 874–83.

Gjerlaug-Enger, E., Kongsro, J., Aass, L., Odegård, J., and Vangen, O. 2011. Prediction of fat quality in pig carcasses by near-infrared spectroscopy. Animal 5: 1829–41.

Gocayne, J., Robinson, D.A., FitzGerald, M.G., Chung, F.Z., Kerlavage, A.R., Lentes, K.U., Lai, J., Wang, C.D., Fraser, C.M., and Venter, J.C. 1987. Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic receptors obtained by automated DNA sequence analysis: further evidence for a multigene family. Proc. Natl. Acad. Sci. U. S. A. 84: 8296–300.

Goddard, M.E., and Hayes, B.J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10: 381–391.

Gonen, S., Lowe, N.R., Cezard, T., Gharbi, K., Bishop, S.C., and Houston, R.D. 2014. Linkage maps of the Atlantic salmon (Salmo salar) genome derived from RAD sequencing. BMC Genomics 15: 166.

González-Recio, O., and Forni, S. 2011. Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. Genet. Sel. Evol. 43: 7.

Gordon, D., Haynes, C., Johnnidis, C., Patel, S.B., Bowcock, A.M., and Ott, J. 2004. A transmission disequilibrium test for general pedigrees that is robust to the presence of random genotyping errors and any number of untyped parents. Eur. J. Hum. Genet. 12: 752–761.

Gordon, D., Heath, S.C., Liu, X., and Ott, J. 2001. A transmission/disequilibrium test that allows for genotyping errors in the analysis of single-nucleotide polymorphism data. Am. J. Hum. Genet. 69: 371–80.

Griffith, M., Griffith, O.L., Mwenifumbo, J., Goya, R., Morrissy, a S., Morin, R.D., Corbett, R., Tang, M.J., Hou, Y., Pugh, T.J., Robertson, G., Chittaranjan, S., Ally, A., Asano, J.K., Chan, S.Y., Li, H.I., McDonald, H., Teague, K., Zhao, Y., Zeng, T., Delaney, A., Hirst, M., Morin, G.B., Jones, S.J.M., Tai, I.T., and Marra, M.A. 2010. Alternative expression analysis by RNA sequencing. Nat. Methods 7: 843–847.

Groenen, M.A., Cheng, H.H., Bumstead, N., Benkel, B.F., Briles, W.E., Burke, T., Burt, D.W., Crittenden, L.B., Dodgson, J., Hillel, J., Lamont, S., de Leon, A.P., Soller, M., Takahashi, H., and Vignal, A. 2000. A consensus linkage map of the chicken genome. Genome Res. 10: 137–47.

Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A.F., Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R.C., Clee, C., Crooijmans, R.P.M.A., Dawson, H.D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G.J., Fowler, K.E., Fredholm, M., Fritz, E., Gilbert, J.G.R., Giuffra, E., Gorodkin, J., Griffin, D.K., Harrow, J.L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S.J., Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H.A., Li, Y., Liu, W., Loveland, J.E., Lu, Y., Lunney, J.K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M.P., Narayan, J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J.M., Rodriguez-Zas, S., Rohrer, G.A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J.G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B.R., Sperber, G., Stadler, P., Sweedler, J. V., Tafer, H., Thomsen, B., Wali, R., Wang, J., Wang, J., White, S., Xu, X., Yerle, M., Zhang, G., Zhang, J., Zhang, J., Zhao, S., Rogers, J., Churcher, C., and Schook, L.B. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491: 393–398. Nature Publishing Group.

Groenen, M.A.M., Wahlberg, P., Foglio, M., Cheng, H.H., Megens, H.-J., Crooijmans, R.P.M.A., Besnier, F., Lathrop, M., Muir, W.M., Wong, G.K.-S., Gut, I., and Andersson, L. 2008. A

high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res. 19: 510–519.

Guàrdia, M.D., Estany, J., Balasch, S., Oliver, M.A., Gispert, M., and Diestre, A. 2009. Risk assessment of skin damage due to pre-slaughter conditions and RYR1 gene in pigs. Meat Sci. 81: 745–751.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J.T., Agnarsson, B. a, Baker, A., Sigurdsson, A., Benediktsdottir, K.R., Jakobsdottir, M., Xu, J., Blondal, T., Kostic, J., Sun, J., Ghosh, S., Stacey, S.N., Mouy, M., Saemundsdottir, J., Backman, V.M., Kristjansson, K., Tres, A., Partin, A.W., Albers-Akkers, M.T., Godino-Ivan Marcos, J., Walsh, P.C., Swinkels, D.W., Navarrete, S., Isaacs, S.D., Aben, K.K., Graif, T., Cashy, J., Ruiz-Echarri, M., Wiley, K.E., Suarez, B.K., Witjes, J.A., Frigge, M., Ober, C., Jonsson, E., Einarsson, G. V, Mayordomo, J.I., Kiemeney, L. a, Isaacs, W.B., Catalona, W.J., Barkardottir, R.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., and Stefansson, K. 2007. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat. Genet. 39: 631–637.

Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G., and Chee, M.S. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. Nat. Genet. 37: 549–554.

Guttridge, D.C. 2000. NF-kappa B-Induced loss of MyoD messenger RNA: possible role in muscle decay and cachexia. Science 289: 2363–2366.

Hackett, C.A., McLean, K., and Bryan, G.J. 2013. Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PLoS One 8: e63939.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning.

Hayes, B. 2013. Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). in C.Gondro; J. van der Werf; B. Hayes, ed. Genome-wide association studies and genomic prediction. Humana Press, Totowa, NJ.

Hayes, B., and Goddard, M.E. 2003. Evaluation of marker assisted selection in pig enterprises. Livest. Prod. Sci. 81: 197–211.

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

Heid, I.M., and Winkler, T.W. 2016. A multitrait GWAS sheds light on insulin resistance. Nat. Genet. 49: 7–8.

Hennebry, A., Berry, C., Siriett, V., O'Callaghan, P., Chau, L., Watson, T., Sharma, M., and Kambadur, R. 2008. Myostatin regulates fiber-type composition of skeletal muscle by regulating MEF2 and MyoD gene expression. AJP Cell Physiol. 296: C525–C534.

Hernández-Sánchez, J., Visscher, P., Plastow, G., and Haley, C. 2003. Candidate gene analysis for quantitative traits using the transmission disequilibrium test: the example of the melanocortin 4-receptor in pigs. Genetics 164: 637–44.

Hickey, J.M. 2013. Sequencing millions of animals for genomic selection 2.0. J. Anim. Breed. Genet. 130: 331–332.

Hill, W.G. 2000. Maintenance of quantitative genetic variation in animal breeding programmes. Livest. Prod. Sci. 63: 99–109.

Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.I., Hickenbotham, M., Huang, W., Magrini, V.J., Richt, R.J., Sander, S.N., Stewart, D. a, Stromberg, M., Tsung, E.F., Wylie, T., Schedl, T., Wilson, R.K., and Mardis, E.R. 2008. Whole-genome sequencing and variant discovery in C. elegans. Nat. Methods 5: 183–188.

Hindson, C.M., Chevillet, J.R., Briggs, H. a, Gallichotte, E.N., Ruf, I.K., Hindson, B.J., Vessella, R.L., and Tewari, M. 2013. Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat. Methods 10: 1003–1005.

Hiscox, J.A. 2002. The nucleolus – a gateway to viral infection? Arch. Virol. 147: 1077–1089.

Hoh, J., and Ott, J. 2005. Genetic linkage mapping. in Encyclopedia of Life Sciences.

Holtkamp, D.J., Kliebenstein, J.B., Neumann, E.J., Zimmerman, J.J., Rotto, H.F., Yoder, T.K., Wang, C., Yeske, P.E., Mowrer, C.L., and Haley, C. a. 2013. Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. J. Swine Heal. Prod. 21: 72–84.

Horvath, S., and Laird, N.M. 1998. A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am. J. Hum. Genet. 63: 1886–97.

Hou, J., Wang, L., Quan, R., Fu, Y., Zhang, H., and Feng, W. 2012. Induction of interleukin-10 is dependent on p38 mitogen-activated protein kinase pathway in macrophages infected with porcine reproductive and respiratory syndrome virus. Virol. J. 9: 165.

Hu, Z.-L., Park, C.A., Wu, X.-L., and Reecy, J.M. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. 41: D871–D879.

Huang, W.H., Ma, Z.X., Xu, Z.Y., Xiong, Y.Z., and Zuo, B. 2011. Detection of novel SNPs and mapping of the fatness QTL on pig chromosome 7q1.1-1.4 region. Genet. Mol. Res. 10: 3090–3097.

Hughes, S.M., Koishi, K., Rudnicki, M., and Maggs, A.M. 1997. MyoD protein is differentially accumulated in fast and slow skeletal muscle fibres and required for normal fibre type balance in rodents. Mech. Dev. 61: 151–163.

Hume, D.A., Pavli, P., Donahue, R.E., and Fidler, I.J. 1988. The effect of human recombinant macrophage colony-stimulating factor (CSF-1) on the murine mononuclear phagocyte system in vivo. J. Immunol. 141: 3405–9.

Huo, Y., Fan, L., Yin, S., Dong, Y., Guo, X., Yang, H., and Hu, H. 2013. Involvement of unfolded protein response, p53 and Akt in modulation of porcine reproductive and respiratory syndrome virus-mediated JNK activation. Virology 444: 233–240.

International HapMap Consortium, Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.-F., Galver, L.M., Fan, J.-B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.-F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Kwok, P.-Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Song, Y.Q., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I.W., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er,

I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Altshuler, D., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Tsunoda, T., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Gibbs, R.A., Belmont, J.W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Wheeler, D.A., Yakub, I., Gabriel, S.B., Onofrio, R.C., Richter, D.J., Ziaugra, L., Birren, B.W., Daly, M.J., Altshuler, D., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., and Stewart, J. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851–61.

Islam, Z.U., Bishop, S.C., Savill, N.J., Rowland, R.R.R., Lunney, J.K., Trible, B., and Doeschl-Wilson, A.B. 2013. Quantitative analysis of porcine reproductive and respiratory syndrome

(PRRS) viremia profiles from experimental infection: a statistical modelling approach. PLoS One 8: e83567.

James, G., Witten, D., Hastie, T., and Tibshirani, R. 2013. An Introduction to Statistical Learning. Springer New York, New York, NY.

Jeffreys, A.J., Ritchie, A., and Neumann, R. 2000. High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. Hum. Mol. Genet. 9: 725–33.

Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985. Hypervariable "minisatellite" regions in human DNA. Nature 314: 67–73.

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y., and Zhang, Q. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5: e13661.

Jiang, X., and Neapolitan, R.E. 2012. Mining pure, strict epistatic interactions from high-dimensional datasets: ameliorating the curse of dimensionality. PLoS One 7: e46771.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337: 816–21.

Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D., Winkler, C.A., and O'Brien, S.J. 2010. Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics 11: 724.

Jozsef, L., Tashiro, K., Kuo, A., Park, E.J., Skoura, A., Albinsson, S., Rivera-Molina, F., Harrison, K.D., Iwakiri, Y., Toomre, D., and Sessa, W.C. 2014. Reticulon 4 is necessary for endoplasmic reticulum tubulation, STIM1-Orai1 coupling, and store-operated calcium entry. J. Biol. Chem. 289: 9380–95.

Kan, Y.W., and Dozy, A.M. 1978. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. Proc. Natl. Acad. Sci. U. S. A. 75: 5631–5.

Karniychuk, U.U., and Nauwynck, H.J. 2013. Pathogenesis and prevention of placental and transplacental porcine reproductive and respiratory syndrome virus infection. Vet. Res. 44: 95.

Karniychuk, U.U., Saha, D., Geldhof, M., Vanhee, M., Cornillie, P., Van den Broeck, W., and Nauwynck, H.J. 2011. Porcine reproductive and respiratory syndrome virus (PRRSV) causes apoptosis during its replication in fetal implantation sites. Microb. Pathog. 51: 194–202.

Karniychuk, U.U., Saha, D., Vanhee, M., Geldhof, M., Cornillie, P., Caij, A.B., De Regge, N., and Nauwynck, H.J. 2012. Impact of a novel inactivated PRRS virus vaccine on virus replication and virus-induced pathology in fetal implantation sites and fetuses upon challenge. Theriogenology 78: 1527–1537.

Karp, R., Ruzzo, W., and Tompa, M. 1996. Algorithms in molecular biology (lecture notes).

Kawaguchi, T., Itou, M., Taniguchi, E., and Sata, M. 2014. Exendin‐4, a glucagon‐like peptide‐1 receptor agonist, modulates hepatic fatty acid composition and Δ‐5‐desaturase index in a murine model of non‐alcoholic steatohepatitis. Int. J. Mol. Med. 34: 782–787.

Kennedy, D., Jäger, R., Mosser, D.D., and Samali, A. 2014. Regulation of apoptosis by heat shock proteins. IUBMB Life 66: 327–338.

Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., Liu, W., Yang, G., Di, X., Ryder, T., He, Z., Surti, U., Phillips, M.S., Boyce-Jacino, M.T., Fodor, S.P., and Jones, K.W. 2003. Large-scale genotyping of complex DNA. Nat. Biotechnol. 21: 1233–1237.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, a. D. 2002. The Human Genome Browser at UCSC. Genome Res. 12: 996–1006.

Kerem, B.S., Buchanan, J.A., Durie, P., Corey, M.L., Levison, H., Rommens, J.M., Buchwald, M., and Tsui, L.C. 1989. DNA marker haplotype association with pancreatic sufficiency in cystic fibrosis. Am. J. Hum. Genet. 44: 827–34.

Kirkpatrick, B., Reshef, Y., Finucane, H., Jiang, H., Zhu, B., and Karp, R.M. 2012. Comparing pedigree graphs. J. Comput. Biol. 19: 998–1014.

Klont, R.E., Brocks, L., and Eikelenboom, G. 1998. Muscle fibre type and meat quality. Meat Sci. 49: S219–S229.

Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A., Jonasdottir, A., Walters, G.B., Gylfason, A., Kristinsson, K.T., Gudjonsson, S.A., Frigge, M.L., Helgason, A., Thorsteinsdottir, U., and Stefansson, K. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. Nature 467: 1099–1103.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z., Tanzer, A., Saunders, a C.E., Chi, J., Yang, F., Carter, N.P., Hurles, M.E., Weissman, S.M., Harkins, T.T., Gerstein, M.B., Egholm, M., and Snyder, M. 2009. Paired-end mapping reveals extensive structural variation in the human genome. Science 318: 420–6.

Krämer, N., Schäfer, J., and Boulesteix, A.-L. 2009. Regularized estimation of large-scale gene association networks using graphical Gaussian models. BMC Bioinformatics 10: 384.

Ladinig, A., Ashley, C., Detmer, S.E., Wilkinson, J.M., Lunney, J.K., Plastow, G., and Harding, J.C. 2015. Maternal and fetal predictors of fetal viral load and death in third trimester, type 2

porcine reproductive and respiratory syndrome virus infected pregnant gilts. Vet. Res. 46: 107.

Ladinig, A., Foxcroft, G., Ashley, C., Lunney, J.K., Plastow, G., and Harding, J.C.S. 2014a. Birth weight, intrauterine growth retardation and fetal susceptibility to porcine reproductive and respiratory syndrome virus. PLoS One 9: e109541.

Ladinig, A., Gerner, W., Saalmüller, A., Lunney, J.K., Ashley, C., and Harding, J.C. 2014b. Changes in leukocyte subsets of pregnant gilts experimentally infected with porcine reproductive and respiratory syndrome virus and relationships with viral load and fetal outcome. Vet. Res. 45: 128.

Ladinig, A., Lunney, J.K., Souza, C.J., Ashley, C., Plastow, G., Harding, J.C., Gerner, W., Saalmüller, A., Lunney, J.K., Ashley, C., and Harding, J.C. 2014c. Cytokine profiles in pregnant gilts experimentally infected with porcine reproductive and respiratory syndrome virus and relationships with viral load and fetal outcome. Vet. Res. 45: 113.

Ladinig, A., Wilkinson, J., Ashley, C., Detmer, S.E., Lunney, J.K., Plastow, G., and Harding, J.C.S. 2014d. Variation in fetal outcome, viral load and ORF5 sequence mutations in a large scale study of phenotypic responses to late gestation exposure to type 2 porcine reproductive and respiratory syndrome virus. PLoS One 9: e96104.

Laing, N.G., Clarke, N.F., Dye, D.E., Liyanage, K., Walker, K.R., Kobayashi, Y., Shimakawa, S., Hagiwara, T., Ouvrier, R., Sparrow, J.C., Nishino, I., North, K.N., and Nonaka, I. 2004. Actin mutations are one cause of congenital fibre type disproportion. Ann. Neurol. 56: 689–694.

Lander, E.S., and Botstein, S. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M.,

Grimwood, J., Cox, D.R., Olson, M. V, Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E. V, Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J. V, Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., and International Human Genome Sequencing Consortium 2001. Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Langen, R.C.J., Van Der Velden, J.L.J., Schols, A.M.W.J., Kelders, M.C.J.M., Wouters, E.F.M., and Janssen-Heininger, Y.M.W. 2004. Tumor necrosis factor-alpha inhibits myogenic differentiation through MyoD protein destabilization. FASEB J. 18: 227–37.

Lazzeroni, L.C., Arnheim, N., Schmitt, K., and Lange, K. 1994. Multipoint mapping calculations for sperm-typing data. Am. J. Hum. Genet. 55: 431–6.

Lee, E.A., Kim, J.M., Lim, K.S., Ryu, Y.C., Jeon, W.M., and Hong, K.C. 2012. Effects of variation in porcine MYOD1 gene on muscle fiber characteristics, lean meat production, and meat quality traits. Meat Sci. 92: 36–43.

Lee, K.M., Danuser, R., Stein, J. V, Graham, D., Nibbs, R.J., and Graham, G.J. 2014. The chemokine receptors ACKR2 and CCR2 reciprocally regulate lymphatic vessel density. EMBO J. 33: 2564–2580.

Lee, S.-M., and Kleiboeker, S.B. 2007. Porcine reproductive and respiratory syndrome virus induces apoptosis through a mitochondria-mediated pathway. Virology 365: 419–34.

Leng, C., Tran, M.N., and Nott, D. 2014. Bayesian adaptive Lasso. Ann. Inst. Stat. Math. 66: 221–244.

Li, C., Liu, D., Zhou, G., Xu, X., Qi, J., Shi, P., and Xia, T. 2012. Meat quality and cooking attributes of thawed pork with different low field NMR T21. Meat Sci. 92: 79–83.

Li, H.H., Gyllensten, U.B., Cui, X.F., Saiki, R.K., Erlich, H.A., and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335: 414–7.

Li, J., and Ma, S. 2013. Survival Analysis in Medicine and Genetics. CRC Press, Boca Raton, FL.

Li, L., and Yin, X. 2008. Sliced inverse regression with regularizations. Biometrics 64: 124–131.

Li, M.-X., Yeung, J.M.Y., Cherny, S.S., and Sham, P.C. 2012. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum. Genet. 131: 747–56.

Li, T., Sham, P.C., Vallada, H., Xie, T., Tang, X., Murray, R.M., Liu, X., and Collier, D.A. 1996. Preferential transmission of the high activity allele of COMT in schizophrenia. Psychiatr. Genet. 6: 131–3.

Lillehammer, M., Meuwissen, T.H.E., and Sonesson, A.K. 2011. Genomic selection for maternal traits in pigs. J. Anim. Sci. 89: 3908–3916.

Lim, A., Dimalanta, E.T., Potamousis, K.D., Yen, G., Apodoca, J., Tao, C., Lin, J., Qi, R., Skiadas, J., Ramanathan, A., Perna, N.T., Plunkett, G., Burland, V., Mau, B., Hackett, J., Blattner, F.R., Anantharaman, T.S., Mishra, B., and Schwartz, D.C. 2001. Shotgun optical maps of the whole Escherichia coli O157:H7 genome. Genome Res. 11: 1584–93.

Lindahl, G., Enfält, A.-C., Seth, G. von, Joseli, Å., Hedebro-Velander, I., Andersen, H.J., Braunschweig, M., Andersson, L., and Lundström, K. 2004. A second mutant allele (V199I) at the PRKAG3 (RN) locus—II. Effect on colour characteristics of pork loin. Meat Sci. 66: 621–627.

Litt, M., and Luty, J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44: 397–401.

Liu, B.H. 1998. Statistical genomics: linkage, mapping, and QTL analysis.

Liu, J., Chen, J., and Ye, J. 2009. Large-scale sparse logistic regression. in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’09.

Long, X., Creemers, E.E., Wang, D.-Z., Olson, E.N., and Miano, J.M. 2007. Myocardin is a bifunctional switch for smooth versus skeletal muscle differentiation. Proc. Natl. Acad. Sci. 104: 16570–16575.

Loughna, P.T., and Bates, P.C. 1994. Interactions between growth hormone and nutrition in hypophysectomized rats: skeletal muscle myosin heavy chain mRNA levels. Biochem. Biophys. Res. Commun. 198: 97–102.

Lourenco, D.A.L., Tsuruta, S., Fragomeni, B.O., Masuda, Y., Aguilar, I., Legarra, A., Bertrand, J.K., Amen, T.S., Wang, L., Moser, D.W., and Misztal, I. 2015. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. J. Anim. Sci. 93: 2653.

Lu, J., Tan, J., Shatadal, P., and Gerrard, D.E. 2000. Evaluation of pork color by using computer vision. Meat Sci. 56: 57–60.

Lu, S., Zong, C., Fan, W., Yang, M., Li, J., Chapman, A.R., Zhu, P., Hu, X., Xu, L., Yan, L., Bai, F., Qiao, J., Tang, F., Li, R., and Xie, X.S. 2012. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. Science 338: 1627–30.

Lu, X., Gong, Y.F., Liu, J.F., Wang, Z.P., Hu, F., Qiu, X.T., Luo, Y.R., and Zhang, Q. 2011a. Mapping quantitative trait loci for cytokines in the pig. Anim. Genet. 42: 1–5.

Lu, X., Liu, J.-F., Gong, Y.-F., Wang, Z.-P., Liu, Y., and Zhang, Q. 2011b. Mapping quantitative trait loci for T lymphocyte subpopulations in peripheral blood in swine. BMC Genet. 12: 79.

Lunney, J.K., and Chen, H. 2010. Genetic control of host resistance to porcine reproductive and respiratory syndrome virus (PRRSV) infection. Virus Res. 154: 161–169.

Lunney, J.K., Fritz, E.R., Reecy, J.M., Kuhar, D., Prucnal, E., Molina, R., Christopher-Hennings, J., Zimmerman, J., and Rowland, R.R.R. 2010. Interleukin-8, interleukin-1β, and interferon-γ levels are linked to PRRS virus clearance. Viral Immunol. 23: 127–134.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K.T. 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 7: e30087.

Lutz, M.A., and Correll, P.H. 2003. Activation of CR3-mediated phagocytosis by MSP requires the RON receptor, tyrosine kinase activity, phosphatidylinositol 3-kinase, and protein kinase C zeta. J. Leukoc. Biol. 73: 802–14.

MacLeod, I.M., Bowman, P.J., Vander Jagt, C.J., Haile-Mariam, M., Kemper, K.E., Chamberlain, A.J., Schrooten, C., Hayes, B.J., and Goddard, M.E. 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics 17: 144.

Madani, N., Millette, R., Platt, E.J., Marin, M., Kozak, S.L., Bloch, D.B., and Kabat, D. 2002. Implication of the Lymphocyte-Specific Nuclear Body Protein Sp140 in an Innate Response to Human Immunodeficiency Virus Type 1. J. Virol. 76: 11133–11138.

Mancini, R.A., and Hunt, M.C. 2005. Current research in meat color. Meat Sci. 71: 100–121.

Manickam, C., Dwivedi, V., Patterson, R., Papenfuss, T., and Renukaradhya, G.J. 2013. Porcine reproductive and respiratory syndrome virus induces pronounced immune modulatory responses at mucosal tissues in the parental vaccine strain VR2332 infected pigs. Vet. Microbiol. 162: 68–77.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., and Visscher, P.M. 2009. Finding the missing heritability of complex diseases. Nature 461: 747–53.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18: 1509–1517.

Marques, E., de Givry, S., Stothard, P., Murdoch, B., Wang, Z., Womack, J., and Moore, S.S. 2007. A high resolution radiation hybrid map of bovine chromosome 14 identifies scaffold rearrangement in the latest bovine assembly. BMC Genomics 8: 254.

Maxam, A.M., and Gilbert, W. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U. S. A. 74: 560–4.

Meuwissen, T., Hayes, B., and Goddard, M. 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front. 6: 6.

Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–29.

Meuwissen, T.H.E., and Goddard, M.E. 1996. The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. 28: 161.

Miar, Y., Plastow, G.S., Moore, S.S., Manafiazar, G., Charagu, P., Kemp, R.A., Van Haandel, B., Huisman, A.E., Zhang, C.Y., McKay, R.M., Bruce, H.L., and Wang, Z. 2014. Genetic and phenotypic parameters for carcass and meat quality traits in commercial crossbred pigs. J. Anim. Sci. 92: 2869–2884.

Milan, D. 2000. A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science 288: 1248–1251.

Min Jou, W., Haegeman, G., Ysebaert, M., and Fiers, W. 1972. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature 237: 82–8.

Mitchell, A.A., Cutler, D.J., and Chakravarti, A. 2003. Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. Am. J. Hum. Genet. 72: 598–610.

Morgan, T.H. 1911. Random segregation versus coupling in Mendelian inheritance. Science 34: 384.

Morrison, A.C., Wilson, C.B., Ray, M., and Correll, P.H. 2004. Macrophage-stimulating protein, the ligand for the stem cell-derived tyrosine kinase/RON receptor tyrosine kinase, inhibits IL-12 production by primary peritoneal macrophages stimulated with IFN-gamma and lipopolysaccharide. J. Immunol. 172: 1825–32.

Mucha, A., Ropka-Molik, K., Piórkowska, K., Tyra, M., and Oczkowicz, M. 2013. Effect of EGF, AREG and LIF genes polymorphisms on reproductive traits in pigs. Anim. Reprod. Sci. 137: 88–92.

Mucha, S., Pszczoła, M., Strabel, T., Wolc, A., Paczyńska, P., and Szydlowski, M. 2011. Comparison of analyses of the QTLMAS XIV common dataset. II: QTL analysis. BMC Proc. 5: S2.

Murray, R., Lee, F., and Chiu, C.P. 1990. The genes for leukemia inhibitory factor and interleukin-6 are expressed in mouse blastocysts prior to the onset of hemopoiesis. Mol. Cell. Biol. 10: 4953–6.

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., and Kanaya, S. 2011. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 39: e90–e90.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., and Kumlin, E. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235: 1616–22.

Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R., Shimada, A., Wada, H., Kusakabe, T., Suzuki, N., Kinoshita, M., Kanamori, A., Terado, T., Kimura, H., Nonaka, M., and Shima, A. 2000. A detailed linkage map of medaka, Oryzias latipes: comparative genomics and genome evolution. Genetics 154: 1773–84.

Neilan, J.G., Lu, Z., Afonso, C.L., Kutish, G.F., Sussman, M.D., and Rock, D.L. 1993. An African swine fever virus gene with similarity to the proto-oncogene bcl-2 and the Epstein-Barr virus gene BHRF1. J. Virol. 67: 4391–4.

Neumann, E.J., Kliebenstein, J.B., Johnson, C.D., Mabry, J.W., Bush, E.J., Seitzinger, A.H., Green, A.L., and Zimmerman, J.J. 2005. Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States. J. Am. Vet. Med. Assoc. 227: 385–92.

Nielsen, J., Bøtner, A., Tingstedt, J.-E., Aasted, B., Johnsen, C.., Riber, U., and Lind, P. 2003. In utero infection with porcine reproductive and respiratory syndrome virus modulates leukocyte subpopulations in peripheral blood and bronchoalveolar fluid of surviving piglets. Vet. Immunol. Immunopathol. 93: 135–151.

Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. 2011. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12: 443–451.

Nielsen, V.H., Bendixen, C., Arnbjerg, J., Sørensen, C.M., Jensen, H.E., Shukri, N.M., and Thomsen, B. 2000. Abnormal growth plate function in pigs carrying a dominant mutation in type X collagen. Mamm. Genome 11: 1087–1092.

Nogueiras, R., Perez-Tilve, D., Veyrat-Durebex, C., Morgan, D. a, Varela, L., Haynes, W.G., Patterson, J.T., Disse, E., Pfluger, P.T., Lopez, M., Woods, S.C., DiMarchi, R., Dieguez, C., Rahmouni, K., Rohner-Jeanrenaud, F., and Tschop, M.H. 2009. Direct control of peripheral lipid deposition by CNS GLP-1 receptor signaling Is mediated by the sympathetic nervous system and blunted in diet-Induced obesity. J. Neurosci. 29: 5916–5925.

Oksbjerg, N., Gondret, F., and Vestergaard, M. 2004. Basic principles of muscle development and growth in meat-producing mammals as affected by the insulin-like growth factor (IGF) system. Domest. Anim. Endocrinol. 27: 219–240.

Olson, J.M., and Boehnke, M. 1990. Monte Carlo comparison of preliminary methods for ordering multiple genetic loci. Am. J. Hum. Genet. 47: 470–82.

Onteru, S.K., Fan, B., Du, Z.Q., Garrick, D.J., Stalder, K.J., and Rothschild, M.F. 2012. A whole-genome association study for pig reproductive traits. Anim. Genet. 43: 18–26.

Ott, J., Kamatani, Y., and Lathrop, M. 2011. Family-based designs for genome-wide association studies. Nat. Rev. Genet. 12: 465–474.

Ott, J., Wang, J., and Leal, S.M. 2015. Genetic linkage analysis in the age of whole-genome sequencing. Nat. Rev. Genet. 16: 275–284.

Otto, G., Roehe, R., Looft, H., Thoelking, L., Knap, P.W., Rothschild, M.F., Plastow, G.S., and Kalm, E. 2007. Associations of DNA markers with meat quality traits in pigs with emphasis on drip loss. Meat Sci. 75: 185–195.

Paul, C., Simon, S., Gibert, B., Virot, S., Manero, F., and Arrigo, A.-P. 2010. Dynamic processes that reflect anti-apoptotic strategies set up by HspB1 (Hsp27). Exp. Cell Res. 316: 1535–1552.

Pavšič, J., Žel, J., and Milavec, M. 2016. Assessment of the real-time PCR and different digital PCR platforms for DNA quantification. Anal. Bioanal. Chem. 408: 107–121.

Pearen, M.A., Eriksson, N.A., Fitzsimmons, R.L., Goode, J.M., Martel, N., Andrikopoulos, S., and Muscat, G.E.O. 2012. The nuclear receptor, Nor-1, markedly increases type II oxidative muscle fibers and resistance to fatigue. Mol. Endocrinol. 26: 372–84.

Pearen, M.A., Goode, J.M., Fitzsimmons, R.L., Eriksson, N.A., Thomas, G.P., Cowin, G.J., Wang, S.-C.M., Tuong, Z.K., and Muscat, G.E.O. 2013. Transgenic muscle-specific Nor-1 expression regulates multiple pathways that effect adiposity, metabolism, and endurance. Mol. Endocrinol. 27: 1897–917.

Pearen, M.A., Myers, S.A., Raichur, S., Ryall, J.G., Lynch, G.S., and Muscat, G.E.O. 2008. The orphan nuclear receptor, NOR-1, a target of beta-adrenergic signaling, regulates gene expression that controls oxidative metabolism in skeletal muscle. Endocrinology 149: 2853–65.

Pearen, M.A., Ryall, J.G., Maxwell, M.A., Ohkura, N., Lynch, G.S., and Muscat, G.E.O. 2006. The orphan nuclear receptor, NOR-1, is a target of β-adrenergic signaling in skeletal muscle. Endocrinology 147: 5217–5227.

Pekin, D., Skhiri, Y., Baret, J.-C., Le Corre, D., Mazutis, L., Ben Salem, C., Millot, F., El Harrak, A., Hutchison, J.B., Larson, J.W., Link, D.R., Laurent-Puig, P., Griffiths, A.D., and Taly, V. 2011. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab Chip 11: 2156.

Peracino, B., Wagner, C., Balest, A., Balbo, A., Pergolizzi, B., Noegel, A.A., Steinert, M., and Bozzaro, S. 2006. Function and mechanism of action of Dictyostelium Nramp1 (Slc11a1) in bacterial infection. Traffic 7: 22–38.

Phatsara, C., Jennen, D.G.J., Ponsuksili, S., Murani, E., Tesfaye, D., Schellander, K., and Wimmers, K. 2007. Molecular genetic analysis of porcine mannose-binding lectin genes, MBL1 and MBL2, and their association with complement activity. Int. J. Immunogenet. 34: 55–63.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., and Albertson, D.G. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. 20: 207–11.

Pixley, F.J., and Stanley, E.R. 2004. CSF-1 regulation of the wandering macrophage: complexity in action. Trends Cell Biol. 14: 628–638.

Pole, J.C.M., McCaughan, F., Newman, S., Howarth, K.D., Dear, P.H., and Edwards, P.A.W. 2011. Single-molecule analysis of genome rearrangements in cancer. Nucleic Acids Res. 39: e85–e85.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35: D61–D65.

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

Qiao, M., Wang, Y., Xu, X., Lu, J., Dong, Y., Tao, W., Stein, J., Stein, G.S., Iglehart, J.D., Shi, Q., and Pardee, A.B. 2010. Mst1 is an interacting protein that mediates PHLPPs' induced apoptosis. Mol. Cell 38: 512–23.

Quinn, G.P., and Keough, M.J. 2002. Experimental Design and Data Analysis for Biologists.

Ramayo-Caldas, Y., Mach, N., Esteve-Codina, A., Corominas, J., Castelló, A., Ballester, M., Estellé, J., Ibáñez-Escriche, N., Fernández, A.I., Pérez-Enciso, M., and Folch, J.M. 2012. Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. BMC Genomics 13: 547.

Reich, D.E., and Goldstein, D.B. 2001. Detecting association in a case-control study while correcting for population stratification. Genet. Epidemiol. 20: 4–16.

Reinelt, G. 1991. TSPLIB—a traveling salesman problem library. ORSA J. Comput. 3: 376–384.

Rhodes, J., Beale, M.A., and Fisher, M.C. 2014. Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of Cryptococcus neoformans using Illumina HiSeq. PLoS One 9: e113501.

Rigby, R.E., Webb, L.M., Mackenzie, K.J., Li, Y., Leitch, A., Reijns, M. a M., Lundie, R.J., Revuelta, A., Davidson, D.J., Diebold, S., Modis, Y., MacDonald, A.S., and Jackson, A.P. 2014. RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. EMBO J. 33: 542–558.

Robinson, G.K.G. 1991. That BLUP is a good thing: The estimation of random effects. Stat. Sci. 6: 15–32.

Rohrer, G.A., Alexander, L.J., Keele, J.W., Smith, T.P., and Beattie, C.W. 1994. A microsatellite linkage map of the porcine genome. Genetics 136: 231–45.

Rohrer, G.A., Nonneman, D.J., Miller, R.K., Zerby, H., and Moeller, S.J. 2012. Association of single nucleotide polymorphism (SNP) markers in candidate genes and QTL regions with pork quality traits in commercial pigs. Meat Sci. 92: 511–518.

Ronaghi, M. 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res. 11: 3–11.

Rosendo, A., Iannuccelli, N., Gilbert, H., Riquet, J., Billon, Y., Amigues, Y., Milan, D., and Bidanel, J.P. 2012. Microsatellite mapping of quantitative trait loci affecting female reproductive tract characteristics in Meishan x Large White F2 pigs. J. Anim. Sci. 90: 37–44.

Rosenvold, K., Lærke, H.N., Jensen, S.K., Karlsson, A.H., Lundström, K., and Andersen, H.J. 2002. Manipulation of critical quality indicators and attributes in pork through vitamin E supplementation, muscle glycogen reducing finishing feeding and pre-slaughter stress. Meat Sci. 62: 485–496.

Rossow, K.D. 1998. Porcine reproductive and respiratory syndrome. Vet. Pathol. 35: 1–20.

Rothschild, M.F., Hu, Z.L., and Jiang, Z. 2007. Advances in QTL mapping in pigs. Int. J. Biol. Sci. 3: 192–197.

Ruiter, A., Jongbloed, A.W., van Gent, C.M., Danse, L.H., and Metz, S.H. 1978. The influence of dietary mackerel oil on the condition of organs and on blood lipid composition in the young growing pig. Am. J. Clin. Nutr. 31: 2159–2166.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., Altshuler, D.,

and International SNP Map Working Group 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928–33.

Sanger, F., and Coulson, A.R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 94: 441–448.

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., and Petersen, G.B. 1982. Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. 162: 729–773.

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

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467–70.

Schiavo, G., Dolezal, M.A., Scotti, E., Bertolini, F., Calò, D.G., Galimberti, G., Russo, V., and Fontanesi, L. 2014. Copy number variants in Italian Large White pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. Anim. Genet. 45: 745–749.

Schmid, S., Sachs, D., and TenOever, B.R. 2014. Mitogen-activated protein kinase-mediated licensing of interferon regulatory factor 3/7 reinforces the cell response to virus. J. Biol. Chem. 289: 299–311.

Schuenke, M.D., Kopchick, J.J., Hikida, R.S., Kraemer, W.J., and Staron, R.S. 2008. Effects of growth hormone overexpression vs. growth hormone receptor gene disruption on mouse hindlimb muscle fiber type composition. Growth Horm. IGF Res. 18: 479–486.

Schwartz, D.C., Li, X., Hernandez, L.I., Ramnarain, S.P., Huff, E.J., and Wang, Y.K. 1993. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science 262: 110–4.

Seideman, S.C., Cross, H.R., Smith, G.C., and Durland, P.R. 1984. Factors associated with fresh meat color: a review. J. Food Qual. 6: 211–237.

Serão, N.V.L., Kemp, R.A., Mote, B.E., Harding, J.C.S., Willson, P., Bishop, S.C., Plastow, G.S., and Dekkers, J.C.M. 2014. Whole-genome scan and validation of regions previously associated with PRRS antibody response and growth rate using gilts under health challenge in commercial settings. in Proceedings of 10th World Congress of Genetics Applied to Livestock Production. ASAS, Vancouver, British Columbia, Canada.

Serao, N.V.L., Matika, O., Kemp, R.A., Harding, J.C.S., Bishop, S.C., Plastow, G.S., and Dekkers, J.C.M. 2014. Genetic analysis of reproductive traits and antibody response in a PRRS outbreak herd. J. Anim. Sci. 92: 2905–2921.

Shi, G., Boerwinkle, E., Morrison, A.C., Gu C., C.C., Chakravarti, A., and Rao, D.C. 2011. Mining gold dust under the genome wide significance level: A two-stage approach to analysis of GWAS. Genet. Epidemiol. 35: 111–118.

Silver, N. 2012. The signal and the noise : why so many predictions fail--but some don’t. Penguin Press, New York, NY.

Škrlep, M., Kavar, T., Santé-Lhoutellier, V., and Čandek-Potokar, M. 2009. Effect of I199V polymorphism on PRKAG3 gene on carcass and meat quality traits in Slovenian commercial pigs. J. Muscle Foods 20: 367–376.

Smith, C. 1967. Improvement of metric traits through specific genetic loci. Anim. Prod. 9: 349–358.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., and Hood, L.E. 1986. Fluorescence detection in automated DNA sequence analysis. Nature 321: 674–9.

Snelling, W.M., Chiu, R., Schein, J.E., Hobbs, M., Abbey, C.A., Adelson, D.L., Aerts, J., Bennett, G.L., Bosdet, I.E., Boussaha, M., Brauning, R., Caetano, A.R., Costa, M.M., Crawford, A.M., Dalrymple, B.P., Eggen, A., Everts-van der Wind, A., Floriot, S., Gautier, M., Gill, C.A., Green, R.D., Holt, R., Jann, O., Jones, S.J., Kappes, S.M., Keele, J.W., de Jong, P.J., Larkin, D.M., Lewin, H.A., McEwan, J.C., McKay, S., Marra, M.A., Mathewson, C.A., Matukumalli, L.K., Moore, S.S., Murdoch, B., Nicholas, F.W., Osoegawa, K., Roy, A., Salih, H., Schibler, L., Schnabel, R.D., Silveri, L., Skow, L.C., Smith, T.P., Sonstegard, T.S., Taylor, J.F., Tellam, R., Van Tassell, C.P., Williams, J.L., Womack, J.E., Wye, N.H., Yang, G., and Zhao, S. 2007. A physical map of the bovine genome. Genome Biol. 8: R165.

Snelling, W.M., Chiu, R., Schein, J.E., Hobbs, M., Abbey, C.A., Adelson, D.L., Aerts, J., Bennett, G.L., Bosdet, I.E., Boussaha, M., Brauning, R., Caetano, A.R., Costa, M.M., Crawford, A.M., Dalrymple, B.P., Eggen, A., Everts-van der Wind, A., Floriot, S., Gautier, M., Gill, C.A., Green, R.D., Holt, R., Jann, O., Jones, S.J., Kappes, S.M., Keele, J.W., de Jong, P.J., Larkin, D.M., Lewin, H.A., McEwan, J.C., McKay, S., Marra, M.A., Mathewson, C.A., Matukumalli, L.K., Moore, S.S., Murdoch, B., Nicholas, F.W., Osoegawa, K., Roy, A., Salih, H., Schibler, L., Schnabel, R.D., Silveri, L., Skow, L.C., Smith, T.P., Sonstegard, T.S., Taylor, J.F., Tellam, R., Van Tassell, C.P., Williams, J.L., Womack, J.E., Wye, N.H., Yang, G., and Zhao, S. 2007. A physical map of the bovine genome. Genome Biol. 8: R165.

Sobel, E., Papp, J.C., and Lange, K. 2002. Detection and integration of genotyping errors in statistical genetics. Am. J. Hum. Genet. 70: 496–508.

Sonesson, A.K., de Greef, K.H., and Meuwissen, T.H.E. 1998. Genetic parameters and trends of meat quality, carcass composition and performance traits in two selected lines of large white pigs. Livest. Prod. Sci. 57: 23–32.

Spielman, R.S., and Ewens, W.J. 1998. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am. J. Hum. Genet. 62: 450–8.

Spielman, R.S., McGinnis, R.E., and Ewens, W.J. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. 52: 506–16.

Spielman, R.S., McGinnis, R.E., and Ewens, W.J. 1994. The transmission/disequilibrium test detects cosegregation and linkage. Am. J. Hum. Genet. 54: 559-60–3.

Staaf, J., Vallon-Christersson, J., Lindgren, D., Juliusson, G., Rosenquist, R., Höglund, M., Borg, A., and Ringnér, M. 2008. Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. BMC Bioinformatics 9: 409.

Steemers, F.J., Chang, W., Lee, G., Barker, D.L., Shen, R., and Gunderson, K.L. 2006. Whole-genome genotyping with the single-base extension assay. Nat. Methods 3: 31–33.

Stewart, C.L. 1994. Leukaemia inhibitory factor and the regulation of pre-implantation development of the mammalian embryo. Mol. Reprod. Dev. 39: 233–8.

Strain, M.C., Lada, S.M., Luong, T., Rought, S.E., Gianella, S., Terry, V.H., Spina, C.A., Woelk, C.H., and Richman, D.D. 2013. Highly precise measurement of HIV DNA by droplet digital PCR. PLoS One 8: e55943.

Sturtevant, A.H. 1913. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. 14: 43–59.

Suárez, P., Díaz-Guerra, M., Prieto, C., Esteban, M., Castro, J.M., Nieto, A., and Ortín, J. 1996. Open reading frame 5 of porcine reproductive and respiratory syndrome virus as a cause of virus-induced apoptosis. J. Virol. 70: 2876–82.

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Hsi-Yang Fritz, M., Konkel, M.K., Malhotra, A., Stütz, A.M., Shi, X., Paolo Casale, F., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., Malig, M., Chaisson, M.J.P., Walter, K., Meiers, S., Kashin, S., Garrison, E., Auton, A., Lam, H.Y.K., Jasmine Mu, X., Alkan, C., Antaki, D., Bae, T., Cerveira, E., Chines, P., Chong, Z., Clarke, L., Dal, E., Ding, L., Emery, S., Fan, X., Gujral, M., Kahveci, F., Kidd, J.M., Kong, Y., Lameijer, E.-W., McCarthy, S., Flicek, P., Gibbs, R.A., Marth, G., Mason, C.E., Menelaou, A., Muzny, D.M., Nelson, B.J., Noor, A., Parrish, N.F., Pendleton, M., Quitadamo, A., Raeder, B., Schadt, E.E., Romanovitch, M., Schlattl, A., Sebra, R., Shabalin, A.A., Untergasser, A., Walker, J.A., Wang, M., Yu, F., Zhang, C., Zhang, J., Zheng-Bradley, X., Zhou, W., Zichner, T., Sebat, J., Batzer, M.A., McCarroll, S.A., Mills, R.E., Gerstein, M.B., Bashir, A., Stegle, O., Devine, S.E., Lee, C., Eichler, E.E., and Korbel, J.O. 2015. An integrated map of structural variation in 2,504 human genomes. Nature 526: 75–81.

Sullivan, P.F., Lin, D., Tzeng, J.-Y., van den Oord, E., Perkins, D., Stroup, T.S., Wagner, M., Lee, S., Wright, F.A., Zou, F., Liu, W., Downing, A.M., Lieberman, J., and Close, S.L. 2008. Genomewide association for schizophrenia in the CATIE study: results of stage 1. Mol. Psychiatry 13: 570–584.

Sun, W., Ibrahim, J.G., and Zou, F. 2010. Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. Genetics 185: 349–59.

Sun, Y., Han, M., Kim, C., Calvert, J.G., and Yoo, D. 2012. Interplay between Interferon-Mediated Innate Immunity and porcine reproductive and respiratory syndrome virus. Viruses 4: 424–446.

Sutendra, G., Dromparis, P., Wright, P., Bonnet, S., Haromy, A., Hao, Z., McMurtry, M.S., Michalak, M., Vance, J.E., Sessa, W.C., and Michelakis, E.D. 2011. The role of Nogo and the mitochondria-endoplasmic reticulum unit in pulmonary hypertension. Sci. Transl. Med. 3: 88ra55.

Svegliati-Baroni, G., Saccomanno, S., Rychlicki, C., Agostinelli, L., De Minicis, S., Candelaresi, C., Faraci, G., Pacetti, D., Vivarelli, M., Nicolini, D., Garelli, P., Casini, A., Manco, M., Mingrone, G., Risaliti, A., Frega, G.N., Benedetti, A., and Gastaldelli, A. 2011. Glucagon-like peptide-1 receptor activation stimulates hepatic lipid oxidation and restores hepatic signalling alteration induced by a high-fat diet in nonalcoholic steatohepatitis. Liver Int. 31: 1285–1297.

Szydlowski, M., and Paczyńska, P. 2011. QTLMAS 2010: simulated dataset. BMC Proc. 5: S3.

Tang, Z., Li, Y., Wan, P., Li, X., Zhao, S., Liu, B., Fan, B., Zhu, M., Yu, M., and Li, K. 2007. LongSAGE analysis of skeletal muscle at three prenatal stages in Tongcheng and Landrace pigs. Genome Biol. 8: R115.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. 2011. Differential expression in RNA-seq: A matter of depth. Genome Res. 21: 2213–2223.

te Pas, M.F.W., Kruijt, L., Pierzchala, M., Crump, R.E., Boeren, S., Keuning, E., Hoving-Bolink, R., Hortós, M., Gispert, M., Arnau, J., Diestre, A., and Mulder, H.A. 2013. Identification of proteomic biomarkers in M. Longissimus dorsi as potential predictors of pork quality. Meat Sci. 95: 679–687.

Tianfu Yang, James Wilkinson, Zhiquan Wang, Andrea Ladinig, John Harding, and Graham Plastow. 2016. “A Genome-Wide Association Study of Fetal Response to Type 2 Porcine

Reproductive and Respiratory Syndrome Virus Challenge." Scientific Reports, 6 (February): 20305.

Tianfu Yang, Zhiquan Wang, Zhiqiu Hu and Graham Plastow, 2014. A New Method to Estimate Recombination Rate Based on SNP Allelic Dosage Data. in Proceedings of 10th World Congress of Genetics Applied to Livestock Production. Asas.

Tsika, R.W., Schramm, C., Simmer, G., Fitzsimons, D.P., Moss, R.L., and Ji, J. 2008. Overexpression of TEAD-1 in transgenic mouse striated muscles produces a slower skeletal muscle contractile phenotype. J. Biol. Chem. 283: 36154–36167.

Usai, M.G., Goddard, M.E., and Hayes, B.J. 2009. LASSO with cross-validation for genomic selection. Genet. Res. (Camb). 91: 427–436.

van der Steen, H.A.M., Prall, G.F.W., and Plastow, G.S. 2005. Application of genomics to the pork industry. J. Anim. Sci. 83: E1–E8.

VanRaden, P.M., Olson, K.M., Null, D.J., and Hutchison, J.L. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy Sci. 94: 6153–6161.

Veldman-Jones, M.H., Brant, R., Rooney, C., Geh, C., Emery, H., Harbron, C.G., Wappett, M., Sharpe, A., Dymond, M., Barrett, J.C., Harrington, E.A., and Marshall, G. 2015. Evaluating robustness and sensitivity of the NanoString technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. Cancer Res. 75: 2587–2593.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M.,

Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V, Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K. V, Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert,

R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. 2001. The sequence of the human genome. Science 291: 1304–51.

Veum, V.L., Dankel, S.N., Gjerde, J., Nielsen, H.J., Solsvik, M.H., Haugen, C., Christensen, B.J., Hoang, T., Fadnes, D.J., Busch, C., Våge, V., Sagen, J. V, and Mellgren, G. 2012. The nuclear receptors NUR77, NURR1 and NOR1 in obesity and during fat loss. Int. J. Obes. 36: 1195–1202.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. Van De, Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407–4414.

Waldmann, P., Mészáros, G., Gredler, B., Fuerst, C., and Sölkner, J. 2013. Evaluation of the lasso and the elastic net in genome-wide association studies. Front. Genet. 4: 270.

Waldmann, T.A., and Tagaya, Y. 1999. The multifaceted regulation of interleukin-15 expression and the role of this cytokine in NK cell differentiation and host response to intracellular pathogens. Annu. Rev. Immunol. 17: 19–49.

Walther, B.A., and Moore, J.L. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography (Cop.). 28: 815–829.

Wang, L., Bai, B., Liu, P., Huang, S.Q., Wan, Z.Y., Chua, E., Ye, B., and Yue, G.H. 2017. Construction of high-resolution recombination maps in Asian seabass. BMC Genomics 18: 63.

Wang, W., Meng, Q., Hu, X., and Li, N. 2009. Genetic variation and association of insulin-like growth factor binding protein-3 with performance in Swine. Biochem. Genet. 47: 315–21.

Wang, X., Zhang, H., Abel, A.M., Young, A.J., Xie, L., and Xie, Z. 2014. Role of phosphatidylinositol 3-kinase (PI3K) and Akt1 kinase in porcine reproductive and respiratory syndrome virus (PRRSV) replication. Arch. Virol. 159: 2091–2096.

Warr, A., Hall, R., Kim, K.S., Elizabeth, T., Koren, S., Phillippy, A.M., Bickhart, D.M., Rosen, B.D., Schroeder, S.G., Hume, D.A., Talbot, R., Rund, L., Schook, L.B., Chow, W., Howe, K., Nonneman, D., Rohrer, G.A., Putnam, N., Green, E., Watson, M., Smith, T.P.L., and Archibald, A.L. 2017. Exploiting Long Read Sequencing Technologies to Establish High Quality Highly Contiguous Pig Reference Genome Assemblies. in Plant and Animal Genome Conference XXIV.

Warriss, P.D., Brown, S.N., and Paściak, P. 2006. The colour of the adductor muscle as a predictor of pork quality in the loin. Meat Sci. 73: 565–569.

Watson, J.D., and Crick, F.H.C. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171: 737–738.

Weber, J.L., and May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388–96.

Weng, Z.-Q., Saatchi, M., Schnabel, R.D., Taylor, J.F., and Garrick, D.J. 2014. Recombination locations and rates in beef cattle assessed from parent-offspring pairs. Genet. Sel. Evol. 46: 34.

Westfall, P.H., and Young, S.S. 1993. Resampling-based multiple testing: Examples and methods for p-value adjustment.

Westfall, P.H., Young, S.S., and Wright, S.P. 1993. On adjusting P-values for multiplicity. Biometrics 49: 941.

Wiggans, G.R., VanRaden, P.M., and Cooper, T.A. 2011. The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. 94: 3202–3211.

Wilborn, B.S., Kerth, C.R., Owsley, W.F., Jones, W.R., and Frobish, L.T. 2004. Improving pork quality by feeding supranutritional concentrations of vitamin D. J. Anim. Sci. 82: 218.

Wilkinson, J.M., Bao, H., Ladinig, A., Hong, L., Stothard, P., Lunney, J.K., Plastow, G.S., and Harding, J.C.S. 2016. Genome-wide analysis of the transcriptional response to porcine reproductive and respiratory syndrome virus infection at the maternal/fetal interface and in the fetus. BMC Genomics 17: 383.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S. V 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18: 6531–6535.

Wood, J.D. 1984. Fat deposition and the quality of fat tissue in meat animals. in J. Wiseman, ed. Fats in Animal Nutrition. Butterworth-Heinemann, London, UK.

Wood, J.D., Richardson, R.I., Nute, G.R., Fisher, A.V., Campo, M.M., Kasapidou, E., Sheard, P.R., and Enser, M. 2004. Effects of fatty acids on meat quality: a review. Meat Sci. 66: 21–32.

Wu, T.T., Chen, Y.F., Hastie, T., Sobel, E., and Lange, K. 2009. Genome-wide association analysis by lasso penalized logistic regression. Bioinformatics 25: 714–721.

Xing, J., Gao, H., Wu, Y., Wu, Y., Li, H., and Yang, R. 2014. Generalized linear model for mapping discrete trait loci implemented with LASSO algorithm. PLoS One 9: e106985.

Xu, S. 2013. Principles of Statistical Genomics.

Xu, S., Ackerman, M.S., Long, H., Bright, L., Spitze, K., Ramsdell, J.S., Thomas, W.K., and Lynch, M. 2015. A Male-Specific Genetic Map of the Microcrustacean Daphnia pulex Based on Single-Sperm Whole-Genome Sequencing. Genetics 201: 31–38.

Xu, S., Zhu, D., and Zhang, Q. 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. 111: 12456–12461.

Yan, N., Regalado-Magdos, A.D., Stiggelbout, B., Lee-Kirsch, M.A., and Lieberman, J. 2010. The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. Nat. Immunol. 11: 1005–1013.

Yang, H., Wang, G., Sun, H., Shu, R., Liu, T., Wang, C., Liu, Z., Zhao, Y., Zhao, B., Ouyang, Z., Yang, D., Huang, J., Zhou, Y., Li, S., Jiang, X., Xiao, Z., Li, X.-J., and Lai, L. 2014. Species-dependent neuropathology in transgenic SOD1 pigs. Cell Res. 24: 464–481.

Yao, H.-P., Zhou, Y.-Q., Zhang, R., and Wang, M.-H. 2013. MSP–RON signalling in cancer: pathogenesis and therapeutic potential. Nat. Rev. Cancer 13: 466–481. Nature Publishing Group.

Yekutieli, D., and Benjamini, Y. 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29: 1165–1188.

Yin, S., Huo, Y., Dong, Y., Fan, L., Yang, H., Wang, L., Ning, Y., and Hu, H. 2012. Activation of c-Jun NH(2)-terminal kinase is required for porcine reproductive and respiratory syndrome virus-induced apoptosis but not for virus replication. Virus Res. 166: 103–108.

Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T., Bannai, M., Tokunaga, K., Sonoda, S., Komaki, R., Ihara, M., Saenko, V.A., Alipov, G.K., Sekine, I., Komatsu, K., Takahashi, H., Nakashima, M., Sosonkina, N., Mapendano, C.K., Ghadami, M., Nomura, M., Liang, D.-S., Miwa, N., Kim, D.-K., Garidkhuu, A., Natsume, N., Ohta, T., Tomita, H., Kaneko, A., Kikuchi, M., Russomando, G., Hirayama, K., Ishibashi, M., Takahashi, A., Saitou, N., Murray, J.C., Saito, S., Nakamura, Y., and Niikawa, N. 2006. A SNP in the ABCC11 gene is the determinant of human earwax type. Nat. Genet. 38: 324–330.

Zhan, H., and Xu, S. 2012. Adaptive ridge regression for rare variant detection. PLoS One 7: e44173.

Zhang, C., Wang, Z., Bruce, H., Kemp, R.A., Charagu, P., Miar, Y., Yang, T., and Plastow, G. 2015. Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs. BMC Genet. 16: 33.

Zhang, D., Wang, X., Li, Y., Zhao, L., Lu, M., Yao, X., Xia, H., Wang, Y., Liu, M.-F., Jiang, J., Li, X., and Ying, H. 2014. Thyroid hormone regulates muscle fiber type conversion via miR-133a1. J. Cell Biol. 207: 753–766.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. 1992. Whole genome amplification from a single cell: implications for genetic analysis. Proc. Natl. Acad. Sci. U. S. A. 89: 5847–51.

Zheng, G.X.Y., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M., Mudivarti, P.A., Wyatt, P.W., Bharadwaj, R., Makarewicz, A.J., Li, Y., Belgrader, P., Price, A.D., Lowe, A.J., Marks, P., Vurens, G.M., Hardenbol, P., Montesclaros, L., Luo, M., Greenfield, L., Wong, A., Birch, D.E., Short, S.W., Bjornson, K.P., Patel, P., Hopmans, E.S., Wood, C., Kaur, S., Lockwood, G.K., Stafford, D., Delaney, J.P., Wu, I., Ordonez, H.S., Grimes, S.M., Greer, S., Lee, J.Y., Belhocine, K., Giorda, K.M., Heaton, W.H., McDermott, G.P., Bent, Z.W., Meschi, F., Kondov, N.O., Wilson, R., Bernate, J.A., Gauby, S., Kindwall, A., Bermejo, C., Fehr, A.N., Chan, A., Saxonov, S., Ness, K.D., Hindson, B.J., and Ji, H.P. 2016. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat. Biotechnol. 34: 303–311.

Zhou, S., Wei, F., Nguyen, J., Bechner, M., Potamousis, K., Goldstein, S., Pape, L., Mehan, M.R., Churas, C., Pasternak, S., Forrest, D.K., Wise, R., Ware, D., Wing, R.A., Waterman, M.S., Livny, M., and Schwartz, D.C. 2009. A single molecule scaffold for the maize genome. PLoS Genet. 5: e1000711.

Zhu, P., and Craighead, H.G. 2012. Zero-mode waveguides for single-molecule analysis. Annu. Rev. Biophys. 41: 269–93.

Zou, H. 2006. The adaptive LASSO and its oracle properties. J. Am. Stat. Assoc. 101: 1418–1429.

Appendices

Appendix 1 Status of manuscripts related to the thesis (as of May 20, 2017)

1. Based on Chapter 2:

   Tianfu Yang, Zhiquan Wang, Younes Miar, Heather Bruce, Chunyan Zhang, and Graham Plastow, "A Genome-wide Association Study of Meat Colour in Commercial Crossbred Pigs." Accepted by *Canadian Journal of Animal Science*

2. Based on Chapter 3:

   Tianfu Yang, James Wilkinson, Zhiquan Wang, Andrea Ladinig, John Harding, and Graham Plastow. 2016. "A Genome-Wide Association Study of Fetal Response to Type 2 Porcine Reproductive and Respiratory Syndrome Virus Challenge." *Scientific Reports*, 6 (February): 20305.

3. Based on Chapter 5:

   Tianfu Yang, Zhiquan Wang, Zhiqiu Hu and Graham Plastow, 2014. A New Method to Estimate Recombination Rate Based on SNP Allelic Dosage Data. in *Proceedings of 10th World Congress of Genetics Applied to Livestock Production*. Asas.

   Tianfu Yang, Zhiquan Wang, Zhiqiu Hu and Graham Plastow, "A Statistical Model for Allele Dosage Data of Sperm and its Application in Constructing Individualized Linkage Maps". (Prepared for submission)

4. Other related manuscripts:

Chunyan Zhang, Zhiquan Wang, Heather Bruce, Robert Alan Kemp, Patrick Charagu, Younes Miar, Tianfu Yang, and Graham Plastow. 2015. "Genome-Wide Association Studies (GWAS) Identify a QTL close to PRKAG3 Affecting Meat pH and Colour in Crossbred Commercial Pigs." *BMC Genetics* 16 (1): 33.

Chunyan Zhang, Heather Bruce, Tianfu Yang, Patrick Charagu, Robert Alan Kemp, Nicholas Boddicker, Younes Miar, Zhiquan Wang, and Graham Plastow. 2016. "Genome Wide Association Studies (GWAS) Identify QTL on SSC2 and SSC17 Affecting Loin Peak Shear Force in Crossbred Commercial Pigs." Edited by Jian-Feng Liu. *PloS One* 11 (2): e0145082.

John C.S. Harding, Andrea Ladinig, Predrag Novakovic, Susan E. Detmer, Jamie M. Wilkinson, Tianfu Yang, Joan K. Lunney, and Graham S. Plastow. 2017. "Novel Insights into Host Responses and Reproductive Pathophysiology of Porcine Reproductive and Respiratory Syndrome Caused by PRRSV-2." (in press)

Xingjie Hao, Graham Plastow, Chunyan Zhang, Sutong Xu, Zhiqiu Hu, Tianfu Yang, Kai Wang, Zhiquan Wang, Shujun Zhang. "Genome-Wide Association Study Identifies Candidate Genes for Piglet Splay Leg Syndrome in Different Populations." Submitted to *BMC genetics* (under first revision)

Appendix 2

I also contributed to the analysis and helped co-author two publications related to the GWAS introduced in Chapter 2. One was published in *BMC Genetics* as "Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs". The second was published in PLoS One as "Genome Wide Association Studies (GWAS) Identify QTL on SSC2 and SSC17 Affecting Loin Peak Shear Force in Crossbred Commercial Pigs".

The studies reported in the two publications laid the groundwork for the GWAS of meat colour described in Chapter 2 (Section 2.2). Especially, the results of those two studies aroused our curiosity about improving detection power in the analysis. Given the fairly large pig population (n = 1977) and the wide range of phenotypic records, we sought to explore the possibilities of more candidate regions in addition to the major ones reported in the two publications. The idea directly led to our attempt to apply the adaptive LASSO in the dataset of meat colour, as the methods showed good detection power in our simulation analysis (Section 2.1).

The abstract of the two publications:

1.

Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs[1]

[1] Chunyan Zhang, Zhiquan Wang, Heather Bruce, Robert Alan Kemp, Patrick Charagu, Younes Miar, Tianfu Yang, and Graham Plastow. 2015. "Genome-Wide Association Studies (GWAS) Identify a QTL close to PRKAG3 Affecting Meat pH and Colour in Crossbred Commercial Pigs." BMC Genetics 16 (1): 33.

Chunyan Zhang[1], Zhiquan Wang[1], Heather Bruce[1], Robert Alan Kemp[2], Patrick Charagu[3], Younes Miar[1], Tianfu Yang[1] and Graham Plastow[1]

[1]Department of Agricultural, Food & Nutritional Sciences, University of Alberta, Edmonton, AB T6G 2P5, Canada.

[2]Genesus Inc, Oakville, MB, Canada.

[3]Hypor Inc, Regina, SK, Canada.

BACKGROUND Improving meat quality is a high priority for the pork industry to satisfy consumers' preferences. GWAS have become a state-of-the-art approach to genetically improve economically important traits. However, GWAS focused on pork quality are still relatively rare.

RESULTS Six genomic regions were shown to affect loin pH and Minolta colour a* and b* on both loin and ham through GWAS in 1943 crossbred commercial pigs. Five of them, located on Sus scrofa chromosome (SSC) 1, SSC5, SSC9, SSC16 and SSCX, were associated with meat colour. However, the most promising region was detected on SSC15 spanning 133-134 Mb which explained 3.51% - 17.06% of genetic variance for five measurements of pH and colour. Three SNPs (ASGA0070625, MARC0083357 and MARC0039273) in very strong LD were considered most likely to account for the effects in this region. ASGA0070625 is located in intron 2 of ZNF142, and the other two markers are close to PRKAG3, STK36, TTLL7 and CDK5R2. After fitting MARC0083357 (the closest SNP to PRKAG3) as a fixed factor, six SNPs still remained significant for at least one trait. Four of them are intragenic with ARPC2, TMBIM1, NRAMP1 and VIL1, while the remaining two are close to RUFY4 and CDK5R2. The gene network constructed demonstrated strong connections of these genes with two major hubs of PRKAG3 and UBC in the super-pathways of cell-to-cell signaling and interaction, cellular function and

maintenance. All these pathways play important roles in maintaining the integral architecture and functionality of muscle cells facing the dramatic changes that occur after exsanguination, which is in agreement with the GWAS results found in this study.

CONCLUSIONS There may be other markers and/or genes in this region besides PRKAG3 that have an important effect on pH and colour. The potential markers and their interactions with PRKAG3 require further investigation.

2.

Genome Wide Association Studies (GWAS) Identify QTL on SSC2 and SSC17 Affecting Loin Peak Shear Force in Crossbred Commercial Pigs[1]

Chunyan Zhang[1], Heather Bruce[1], Tianfu Yang[1], Patrick Charagu[2], Robert Alan Kemp[3], Nicholas Boddicker[3], Younes Miar[1], Zhiquan Wang[1], Graham Plastow[1]

[1]Department of Agricultural, Food & Nutritional Sciences, University of Alberta, Edmonton, AB T6G 2P5, Canada.

[2]Hypor Inc, Regina, SK, Canada.

[3]Genesus Inc, Oakville, MB, Canada.

Of all the meat quality traits, tenderness is considered the most important with regard to eating quality and market value. In this study we have utilised genome wide association studies (GWAS) for peak shear force (PSF) of loin muscle as a measure of tenderness for 1,976 crossbred

[1] Chunyan Zhang, Heather Bruce, Tianfu Yang, Patrick Charagu, Robert Alan Kemp, Nicholas Boddicker, Younes Miar, Zhiquan Wang, and Graham Plastow. 2016. "Genome Wide Association Studies (GWAS) Identify QTL on SSC2 and SSC17 Affecting Loin Peak Shear Force in Crossbred Commercial Pigs." Edited by Jian-Feng Liu. PloS One 11 (2): e0145082.

commercial pigs, genotyped for 42,721 informative SNPs using the Illumina PorcineSNP60 Beadchip. Four 1 Mb genomic regions, three on SSC2 (at 4 Mb, 5 Mb and 109 Mb) and one on SSC17 (at 20 Mb), were detected which collectively explained about 15.30% and 3.07% of the total genetic and phenotypic variance for PSF respectively. Markers ASGA0008566, ASGA0008695, DRGA0003285 and ASGA0075615 in the four regions were strongly associated with the effects. Analysis of the reference genome sequence in the region with the most important SNPs for SSC2_5 identified FRMD8, SLC25A45 and LTBP3 as potential candidate genes for meat tenderness on the basis of functional annotation of these genes. The region SSC2_109 was close to a previously reported candidate gene CAST; however, the very weak LD between DRGA0003285 (the best marker representing region SSC2_109) and CAST indicated the potential for additional genes which are distinct from, or interact with, CAST to affect meat tenderness. Limited information of known genes in regions SSC2_109 and SSC17_20 restricts further analysis. Re-sequencing of these regions for informative animals may help to resolve the molecular architecture and identify new candidate genes and causative mutations affecting this trait. These findings contribute significantly to our knowledge of the genomic regions affecting pork shear force and will potentially lead to new insights into the molecular mechanisms regulating meat tenderness.

Appendix 3 Two other studies are related to the GWAS introduced in Chapter 4

Two other studies are related to the GWAS introduced in Chapter 4. One was presented at the Canadian Meat Science Association (CMSA) 96th Annual Conference as "A genome-wide association study of dark cutting in beef cattle", and the other one was presented at Plant and Animal Genomics Conference XXV as "A Genomic Investigation of Porcine Periweaning Failure to Thrive Syndrome (PFTS)". I contributed to parts of the analysis and co-authored the two abstracts.

In Chapter 4, we described an alternative contingency-table test (see Section 4.3.6) to check if the candidate SNPs found in TDT also showed association in a direct comparison of viable and autolyzed fetuses. In addition to the analysis, I also contributed to two studies (Appendix 3) that aimed to detect candidate genomic regions associated with binary traits, using the contingency table test (see Section 1.2.1.4.1) and data from populations with relatively small size. For the contingency-table test described in Section 4.3.6, 280 animals were used in the GWAS of fetal autolysis in PRRSV challenge; Sixty-four beef cattle were used in the GWAS of dark cutting; and 107 pigs were used in the GWAS about PFTS. Moderate suggestive associations were found for dark cutting and PFTS vulnerability, however, we did not find major candidate regions for the three traits of interest.

The results of these studies may be a sign of the low detection power due to limited sample size and the high-dimension of genomic data, as briefly introduced in Section 1.4.1.4.2. Further investigation may aim to improve the power of detection. For example, in the GWAS of dark cutting in beef cattle, we also attempt to pre-screen the SNPs and only select those SNPs close to 40 previously identified candidate genes, in order to reduce the negative effect of multiple testing

on detection power. The pre-screening did result in three more suggestive associations, though still not significant (FDR > 0.1). Other possible strategies may be applied to improve the results, such as through haplotype-based association.

Study 1. Abstract presented at CMSA 96th Annual Conference

A genome-wide association study of dark cutting in beef cattle

Huaigang Lei[1], Tianfu Yang[1], Shahid Mahmood[1], Bimol Roy[1], Changxi Li[1,2], Graham S. Plastow[1], Heather L. Bruce[1]

[1]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada

[2]Agriculture and Agri-Food Canada, Lacombe Research and development Centre, 6000 C&E Trail, Lacombe, AB, Canada T4L

Dark cutting is detrimental to carcass grading because dark cutting carcasses are graded Canada B4 in the Canadian Beef Grading System, resulting in economic loss for beef producers. Dark cutting beef is caused by depletion of muscle glycogen before slaughtering, which may be affected by animal genetics. The current study aimed to identify single nucleotide polymorphisms (SNPs) associated with dark cutting through a genome wide association study (GWAS). In total, 64 beef cattle, of which 24 were graded Canada AA (normal) and 40 graded Canada B4 (dark), were genotyped using GeneSeek Genomic Profiler for Beef Cattle-HD (GGP-HD) featuring 76783 SNPs. All SNPs with a call rate lower than 90% or a minor allele frequency (MAF) lower than 5%

were removed in quality control, resulting in 70675 SNPs for the further association analyses. In addition, we tested SNPs in proximity to 40 previously identified candidate regions. Single-marker Fisher's exact test was conducted using PLINK 1.9 to compare the difference in allele frequencies between AA (normal) and B4 (dark cutters) carcasses. False discovery rate (FDR) was calculated toadjust for multiple testing using R. In total, we detected five SNPs on BTA3, 5, 16, and 28 showing suggestive association with FDR lower than 0.20. Although the detected SNP associations require validation in a larger population than studied, the results suggested the possibility of marker-assisted selection in beef cattle to reduce dark cutting.

Keywords: Genome-wide association study; Dark cutters; Beef cattle

Study 2. Abstract presented at Plant and Animal Genomics Conference XXV

A Genomic Investigation of Porcine Periweaning Failure to Thrive Syndrome (PFTS)

Francesca Bertolini[1], Tianfu Yang[2], Yanyun Huang[3], John Harding[3], Max F. Rothschild[1], Graham S Plastow[2]

[1]Department of Animal Science, Iowa State University, Ames, IA, USA

[2]Livestock Gentec, Department of Agricultural, Food & Nutritional Sciences, University of Alberta, Edmonton, AB, Canada

[3]Department of Large Animal Clinical Sciences, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, SK, Canada

Porcine Periweaning Failure to Thrive Syndrome (PFTS) affects young pigs around 2-3 weeks post-weaning and it is characterized by anorexia, lethargy and progressive debilitation that frequently leads to death. A genetic basis of this syndrome has been recently hypothesized but until now only one GWAS analysis has been reported, with a moderate suggestive association with the disease. In this experiment, 107 pigs were collected in North America within 2-3 weeks after weaning from commercial farms.

A total of 70 met the PFTS case definition (cases) while 37 were aged-matched penmates (control). Cases and controls were balanced across the farms using a 2:1 ratio. All animals were genotyped with the 80K porcine SNP chip and 53,810 filtered autosomal SNPs were considered for the analyses. Population structure analysis was performed, showing that the samples have a similar genetic background even between cases and controls. Then, three analyses were performed comparing cases and controls. These included Fst analysis, considering single SNPs and the average Fst calculated in 500Kb overlapping windows and logistic regression analysis. Only moderate suggestive associations were found.

A total of four regions not identified previously on chromosome 1, 3 and 11 were concordant for at least two types of analyses and contain some genes that are involved in behavior, energy homeostasis, caloric intake and growth and mature size. With these limited analyses, we could not confirm major regions of difference for PFTS but we identified the presence of novel genomic regions that may be moderately associated with this syndrome.