MAPPING OF MONOGENIC AND QUANTITATIVE TRAIT LOCI USING A WHOLE GENOME SCAN APPROACH AND SINGLE NUCLEOTIDE POLYMORPHISM PLATFORMS

BY

ALYSTA D. MARKEY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Jonathan Beever, Chair, Director of Research
Professor Sandra Rodriguez-Zas
Clinical Associate Professor Clifford Shipley
Assistant Professor Anna Dilger

Abstract

The release of the bovine genome sequence in 2004 opened the door for the development of high density single nucleotide polymorphism (SNP) panels that can be used for linkage disequilibrium mapping of traits in cattle populations. The BovineSNP50 Beadchip, containing 54,001 SNP markers, was released in 2008 and the recently released BovineHD Beadchip contains 777,000 SNP markers. These SNP platforms were utilized to map both monogenic and quantitative traits to finite genomic regions in beef cattle populations.

The recessive defect, hypotrichosis is an autosomal recessive form of hairlessness that affects Hereford cattle. A whole-genome association analysis was conducted using BovineSNP50 BeadChip to map the hypotrichosis locus to a chromosomal region. Significant association was detected between the hypotrichosis phenotype and a locus on bovine chromosome 5 (BTA5) and homozygosity analysis localized the associated region to between 29 and 32 Mb. Sequencing of six candidate genes in the region revealed an eight base pair deletion in exon 1 of the bovine *KRT71* gene that was consistent with putative genotypic status within archived samples. The deletion mutation is predicted to result in a frameshift and early truncation of the K71 protein. A DNA based diagnostic was developed that will permit animals to be screened for the mutation and thus allow animal breeders to make informed mating decisions and decrease the incidence of hypotrichosis in the Hereford breed.

Genotype imputation from the BovineSNP50 to the high density BovineHD platform was conducted to provide a denser marker panel to carry out whole genome association mapping and allow for more refinement of genomic regions associated with traits. The imputed genotypes were used in a whole genome scan in order to discover quantitative trait loci that may be influencing growth, carcass and meat quality traits in an US Simmental-Angus population. Association was detected with the traits birth weight, back fat, yield

grade, ribeye area and marbling. There were 81 SNP associations with birth weight that were sequestered into 10 genomic regions. The carcass trait, back fat exhibited association with 209 SNPs in 46 genomic regions and yield grade displayed association with 172 SNPs in 69 genomic areas while there were four SNP associations with ribeye area, representing one genomic region. There were 127 SNP associations with the meat quality trait, marbling, that represented 32 genomic regions. These regions are ideal targets for development of markers for use in marker assisted selection and are also excellent regions to investigate with further fine mapping and discovery of causal variants underlying the quantitative trait loci. The utilization of genotype imputation from the BovineSNP50 to the high density 770K platform allowed localization of QTL to refined regions of the genome, often eliminating the need for further fine mapping efforts prior to targeted re-sequencing.

The whole genome scan using imputed genotypes revealed a locus on BTA 6 from 37 to 42 Mb that was associated with back fat, marbling and birth weight in an US Simmental-Angus population. The region on BTA6 was evaluated in an attempt to further refine the candidate locus interval by implementing haplotype analysis with the imputed SNP markers that were used the whole genome association analysis. Two regions within the 5 Mb interval were associated with the phenotypes; the first was near 38.83 Mb that was associated with both back fat and birth weight. The second region near 39.27 Mb was associated with back fat, marbling and birth weight. The first region, near 38.83 Mb, corresponds to the LCORL-NCAPG locus that has been implicated in body size in many species as well as contributing to growth and carcass phenotypes in cattle. Re-sequencing of the LCORL-NCAPG locus in animals of know haplotype status revealed several polymorphisms that were consistent with haplotype status. This research contributes to the mounting evidence implicating the LCORL-NCAPG region as influencing growth, carcass and meat quality traits in cattle.

Acknowledgements

I have been blessed to have a great group of people behind me during my entire academic career and I am very thankful for each and every one of you. First of all I would like to thank my parents for their love and support not only in my academic pursuits but throughout my life. You have always been there for me and pushed me to excel while never expecting too much or being unrealistic. You raised me to appreciate the little things, to always respect others, to work hard and always do my best and for that I will be forever grateful. And to Jason – thank you for your love and support throughout this entire endeavor. I could not have done this without your understanding and desire for me to finish although I know it took me longer than either of us would have preferred. And yes, thank you for asking about my day and listening patiently even though I could see your eyes glaze over as I talked. I really appreciate that you care enough to ask!

I would like to thank my lab-mates and good friends Erin Wagner, Brandy Marron, Kristen Walker and Stacey Meyers for everything they have done for me. I am truly blessed to have had such a great group of colleagues who all become more than just co-workers. I appreciate all the assistance with projects, moral support when things went awry, shared "wohoos" when things went right, and always having your support no matter what was happening. I am very happy that I have been able to work with group of people who had the ability to pull together and work as a cohesive team. I am so thankful to have had such a great work environment and without you all I am pretty certain I would have lost my sanity! We have shared many good times over the last few years and I will miss my daily interactions with each of you!

I would like to thank "boss" for everything; both things that fell under official duties as well as those things in the unofficial category. I am thankful for all the guidance in research and teaching I truly appreciate your patience, constant availability and the fact that you really do care about each and every one of us. Although we all seem to be

exasperated when you “show us how it’s done” it is nice to have a boss who can truly guide us, teach us, trouble shoot with us and who not only has the ability to work side by side with us but is always willing to do so. I am also grateful for all things you did that were not required of an advisor - lab lunches and birthday cakes, Christmas parties and cookouts, days on the Beever ranch working cows and the times that we simply sat around and had enjoyable conversations. All of these things are going to be hard to find again in future supervisors and I am truly thankful that I have had this good of an experience at least once!

And lastly I would like to thank my committee members Sandra Rodriguez-Zas, Clifford Shipley and Anna Dilger. I appreciate the guidance from each of you and the contributions that you made to my thesis research and academic career at the University of Illinois.

Table of Contents

List of Figures vii
List of Tables viii
Chapter 1. Literature Review 1
Introduction 1
Current status of Bovine Genome Technologies 2
Trait Mapping Approaches 4
Selected Review of Trait Mapping 9
Genotype Imputation 15
Chapter 2. A deletion mutation in KRT71 is associated with congenital hypotrichosis in Hereford cattle. 17
Abstract 17
Background 18
Results and discussion 20
Conclusion 29
Methods 30
Figures 35
Chapter 3. A genome wide association analysis for growth, carcass and meat quality traits in cattle 41
Abstract 41
Background 42
Results and discussion 43
Conclusion 53
Methods 54
Figures 57
Tables 60
Chapter 4. Evaluation of a quantitative trait locus on BTA6 influencing growth, carcass and meat quality traits in beef cattle 63
Abstract 63
Background 64
Results and discussion 65
Conclusion 71
Methods 71
Figures 76
Tables 85
References 94
Appendix 110

List of Figures

Figure 2.1 Manhattan plot of whole genome case/control association analysis.....35

Figure 2.2. Schematic representing the genotypes of affected calves on BTA5. 36

Figure 2.3. Alignment of genomic DNA sequences encompassing exon 1 of the bovine....37
KRT71 gene.

Figure 2.4. Predicted protein sequences of a normal and HY K71 protein. 38

Figure 2.5. The DNA sequences of a normal and HY allele as amplified by the DNA.....39
based diagnostic.

Figure 2.6. Tissue section of normal and Hypotrichosis affected animals.....40

Figure 3.1. Manhattan plots of single nucleotide polymorphism Bonferonni corrected 57
associations.

Figure 4.1. Manhattan plots of single nucleotide polymorphisms Bonferonni corrected....76
associations.

Figure 4.2. Manhattan plots of haplotype association analysis for back fat, marbling.....78
and birth weight on BTA6.

Figure 4.3. DNA sequence from normal and NCAPGins animals. 80

Figure 4.4. Alignment of sequence generated from Sanger sequencing of normal and 81
NCAPGins alleles.

Figure 4.5. Pairwise linkage disequilibrium calculated as r^2 between all SNPs in BTA6 83
interval from 38.5-39.5 Mb.

Figure 4.6. Graphical representation of the relative position of the 20 SNP haplotypes ...84
in the region from 38.5 to 39.5 Mb on BTA6.

List of Tables

Table 3.1. Descriptive statistics of animals with both genotype and phenotype data for each trait. 60

Table 3.2. Selected single SNP associations with traits and genomic regions. 61

Table 4.1. Summary of selected single SNP associations with back fat, birth weight and marbling on BTA6. 85

Table 4.2a. Selected results from haplotype association analysis with back fat on BTA6. 88

Table 4.2b. Selected results from haplotype association analysis with back fat in region 2 on BTA6. 89

Table 4.2c. Selected results from haplotype association analysis with marbling on BTA6. 90

Table 4.2d. Selected results from haplotype association analysis with birth weight in region 1 on BTA6. 91

Table 4.2e. Selected results from haplotype association analysis with birth weight in region 2 on BTA6. 92

Table 4.3. Summary of polymorphisms detected by direct sequencing that were consistent with haplotype status for BW1 haplotype on BTA6. 93

Chapter 1. Literature Review

Introduction

Selection and breeding of livestock likely began shortly after domestication occurred between 7000-9000 BC. During early stages of domestication, selection was likely not a deliberate process but more for traits such as tractability. In comparison, current methods of selection are sophisticated in nature with phenotypic data being collected on a number of traits of economic importance. These phenotypic records are combined with pedigree information and are used to generate estimates of genetic merit. Indeed, the tremendous gains that have been made in production animal species over the past 60 years can be attributed to the use of quantitative theory and the use of breeding values. Even so, there are some traits that have not been as amenable to these selection tools, namely phenotypes that are difficult or too expensive to measure in many individuals. Therefore, the last 20 years has seen substantial investment in understanding the molecular architecture of these traits via the field of animal genomics. The identification of specific genes and how their associated polymorphisms change the phenotype of an individual may potentially accelerate progress in animal selection through the implementation of genomic tools.

Genomics can prove useful for selection of both simple and complex traits. Simple traits, such as deleterious recessive traits, may have a large impact, especially within a breed due to the widespread use of artificial insemination that allows germplasm to be rapidly disseminated. The use of genomics to develop DNA based diagnostics provides a tool to select against individuals carrying recessive defects. While recessive defects can have a large impact on one breed or population, there are ongoing challenges for all beef breeds to improve economically important traits. Unlike poultry and swine, overall advancements in beef populations are slower due to longer generation intervals; therefore genomic information about economically important traits is more crucial. Traits that are

considered important in livestock species are complex in nature and thus, are controlled by many genes or regions of the genome known as quantitative trait loci (QTL). Exploration of the genome for QTL is of continual interest, as much information can be gained concerning variation that exists within and between cattle populations as well as gene function. Investigations of the bovine genome for regions associated with economically important traits can be efficiently executed due to recent technological advances in bovine genomic resources.

Current status of bovine genome technologies

Genome Sequencing

As with any species, the ultimate physical genome map is the complete DNA sequence. The preliminary draft of the bovine sequence assembly was released in 2004 with 3X coverage while the most recent draft of the bovine assembly (Btau_4.2) was released in 2009 with 7.15X coverage [1]. Even though the cost of genome sequencing has become more affordable assembly of the sequence can still be problematic. A separate bovine genome assembly, completed by the University of Maryland (UMD), improved upon the previous errors seen in Btau_4.2 which included gaps in sequence, inversions and unplaced contiguous sequences [2]. The UMD assembly resulted in approximately 91% of the assembled sequence being anchored to a chromosome. Access to a fully sequenced and annotated genome will greatly assist in comparative genomics studies and candidate gene selection.

Comparative mapping between species allows full advantage to be taken of a more advanced genome sequence and annotation such as that of human genome. The evaluation of similarity between human and bovine genomes was undertaken by determining homologous synteny blocks and evolutionary break points between the two species [3]. Bovine genomic sequence and bacterial artificial chromosome (BAC) end sequences from

cattle radiation hybrid maps were examined for similarity to the human genome and over 200 homologous synteny blocks were discovered [4]. Comparative maps are useful to define evolutionary breakpoints between species, facilitate identification of positional candidate genes and allow information to be drawn from species with more complete genome sequence.

Single Nucleotide Polymorphism Genotyping Platforms

The release of the bovine genome sequence opened the door for the development of high density single nucleotide polymorphism (SNP) panels that can be used for linkage disequilibrium (LD) mapping. SNP discovery was completed through a collaboration of research between Illumina, USDA ARS, University of Missouri and the University of Alberta and resulted in the commercial release of the BovSNP50 BeadChip in 2008 [5]. This assay contains 54,609 informative SNP probes spanning the bovine genome with an average spacing of 49.4kb and an average minor allele frequency of 0.24 across all loci. Recently, a higher density SNP assay was released containing over with 777,000 total SNPs, 749,000 of which were validated across 23 breeds with a minor allele frequency greater than 0.05 and a mean spacing of 3.4kb. This assay, known as the BovineHD BeadChip, displayed an average minor allele frequency of 0.25 across all loci for Bos taurus taurus. The bovine SNP assays can be utilized for a variety of applications including mapping of monogenic and quantitative traits, genome-wide selection and evaluation of genetic merit [5].

The initial Bovine SNP array is somewhat biased in design as only SNPs with high minor allele frequencies were selected and only the mainstream breeds of cattle were used. The BovineHD assay examined a more diverse population of animals for SNPs, thus this assay should contain more rare alleles that may be useful in mapping rare phenotypic variants. The denser panel should also allow for more exact mapping of traits.

Trait mapping approaches

The whole genome scan approach has been utilized to locate QTL for production, carcass and efficiency traits in cattle with the eventual goal of elucidating a causative quantitative trait nucleotide underlying the QTL. Traditional QTL studies were undertaken with the expectation that a few QTL existed that accounted for large portions of the genetic variation. A QTL often has to account for a large amount of phenotypic variation in order to be detected in whole genome scans using relatively few markers spread evenly across the genome. QTL that do account for a large amount of variation are often already at or near fixation within a population thus there may be little room for genetic improvement through selection for the specific quantitative trait nucleotide (QTN).

Traditionally, QTL studies have involved experimental populations designed and bred specifically for mapping of QTL. These experimental crosses are typically between breeds that have been in existence for decades and that are also known to be divergent for specific phenotypes. A cross between two divergent breeds will create sufficient enough LD that QTL can be mapped using sparse marker panels with spacing of 15 to 50 cM [6]. Markers that are developed using these experimental mapping populations are often only of value within the families that were used in the discovery phase and application to commercial populations often requires further fine mapping and validation to determine if a QTL is segregating in the population. One goal of QTL mapping is to create a direct marker by elucidating the causative mutation underlying the QTL, but this is often a lengthy process with the substantiation of a causative mutation being difficult. Traditional QTL mapping studies require many years and a considerable amount of resources to develop and phenotype the mapping population as well as elucidate and prove functionality behind any putative causative mutations that are discovered.

Linkage and linkage disequilibrium methods

Phenotypic traits in livestock have been mapped to chromosomal locations through the use of linkage and association that both take advantage of correlations between an allele at a locus and a phenotype, but differ in respect to allelic association in families. Linkage pertains to correlations that exist within families such that different alleles may be positively correlated with the phenotype in different families. Under association (also known as linkage disequilibrium) mapping, a specific allele is correlated with a phenotype across multiple families. While both linkage and association mapping make use of recombination events in order to map phenotypes to specific chromosomal regions, linkage mapping only utilizes recombination events that have occurred within a pedigree to follow chromosomal segments back to a common ancestor within the pedigree. Association mapping capitalizes on historical recombination events that have occurred prior to the genotyped population such that a shared chromosomal segment from a common ancestor appears in multiple descendents within the genotyped population resulting in linkage disequilibrium or non-independence of alleles at two loci.

There are advantages and disadvantages to both linkage and association mapping. Mapping of QTL through the use of linkage allows small or rare QTL to be mapped as the population can be structured to maximize the existence of rare alleles. However, linkage mapping is limited to those QTL that existed in the founder animals when the population was developed, unlike association mapping populations that may contain more phenotypic diversity. Linkage mapping populations accumulate relatively few recombinations that may lead to a QTL being assigned to large chromosomal intervals of 20 cM or more [7], while association mapping typically results in a much more refined chromosomal interval. Linkage mapping necessitates a specific population structure resulting from controlled matings and for QTL mapping these crosses often take advantage of diverse breeds or lines with opposing phenotypes. The development of such a population is a costly process, whereas association mapping populations are often selected from existing populations. Association

mapping is dependent on the amount of linkage disequilibrium that exists within the population, which has been created through the presence of alleles that originated from common ancestors outside of the mapping population. The inherent ability of linkage disequilibrium mapping methods and association mapping populations to refine QTL to small intervals also requires much denser marker panels than those typically used with linkage mapping methods and populations but sufficient coverage and marker density should be provided through the thousands of SNP markers currently available.

Statistical analysis considerations

Although tens of thousands of markers are advantageous for mapping a trait to a small refined chromosomal interval the amount of data does pose statistical issues for those methods traditionally used to map QTL such as development of linkage maps, multiple comparisons, and population structure.

Traditional mapping efforts involve the creation of a linkage map that relates the markers used to each other and approximates distances and positions based on recombination frequencies seen within the pedigree. However, SNP panels will not yield themselves to the creation of linkage maps as the average spacing between SNPS on the Bov50k array is 50kb. If the rough estimate of 1Mb equals 1cM is used and recombination is very infrequent between markers less than 1cM apart it is evident that linkage maps will be a huge challenge to construct. Mapping methods that were designed with sparsely spaced markers such as those used in linkage mapping will not be suitable for the high marker densities seen with the SNP platforms. Association mapping is popular as it is less computationally intensive and as each marker is often tested individually the genomic marker map positions can be used thus alleviating the need for a linkage map.

The multiple comparison issue can be addressed by stringent statistical thresholds or Bayesian analysis methods that fit markers simultaneously. One method that is popular due

to its simplicity is the Bonferroni correction in which the p-value corresponding to the desired significance threshold is divided by the total number of tests or markers thus determining the adjusted threshold p-value [8]. While this protects against false positive errors, it comes at the cost of losing power by creating more false negative results. Another method is the false discovery rate (FDR) that addresses the proportion of false discoveries that are declared significant [9, 10]. This typically involves setting a threshold with the understanding that a certain number of the tests declared positive will actually be false positives. Both Bonferroni and FDR corrections are designed for independent tests however this is not the case with SNP panels as many of the markers will not be independent due to the existence of LD between some markers. Permutation testing, while computationally expensive, is considered the gold standard for multiple comparison correction [8]. Bayesian methods of analysis address the multiple comparison issue by testing all the markers simultaneously eliminating the need for statistical adjustments to the significance thresholds however, Bayesian methods require assumption about priors that are modified in the posterior analysis.

Ideally, association mapping methods would be used in completely unrelated populations as both known and unknown population structure may interfere with LD mapping, however completely unrelated populations do not exist in most livestock species thus methods to assess population structure must be addressed. Population stratification refers to differences in allele frequencies that exist in individuals from different populations which can be extremely problematic in association studies that look for differences in allele frequencies associated with a trait. Often affected individuals are more likely to be related than animals in the normal control group which can lead to spurious associations, thus family based tests are popular. Analyses designed specifically with a family structure in mind have been utilized to map both monogenic recessive traits and quantitative traits. Nuclear families with parental controls have been used to map qualitative traits via transmission disequilibrium tests (TDT) [11] and these methods have been modified to

handle quantitative traits through quantitative trait disequilibrium tests (QTDT) [12] in nuclear families as well as extended to handle general pedigrees [13], sibling pairs [14] and incorporate parental phenotypes [15]. The family based association tests are popular as they are robust to population stratification and since they use parental controls a matched control group is not required.

Unknown relatedness in a population can also lead to population stratification especially if the phenotype of interest varies with the allele frequencies of the subpopulations. In general there are two popular methods to control for unknown population structure, genomic control and structured association. Genomic control is a method wherein an estimate of inflation due to population stratification is derived and the test statistics are adjusted accordingly. This uniform adjustment may be imperfect as allele frequencies from some markers may differ more than others thus the standard adjustment may be too extreme in some instances or too relaxed in others. Structured association methods such as those implemented in the software STRUCTURE assign individuals to subpopulations and then test for association within subpopulation eliminating association due to population structure [16] however this method is computationally intensive and designed for unrelated individuals [7]. A third method, principal components analysis (PCA), uses multiple linear dimensions to infer continuous axes of genetic variation among individuals in a population [17]. The PCA method is implemented in the software EIGENSTRAT [18] that uses the top axes to adjust the genotypes and phenotypes by amounts ascribed to ancestry. PCA methods are fast and perform similarly to structured association methods [19] but both methods only asses variation along a few axes that may not or may not accurately portray the variation within the population. Alternatively, a pair wise relatedness matrix generated from either pedigree information or identity-by-state status at genomic loci can be used to capture complex relationships by estimating the relatedness between all individuals in the sample and the pair wise relatedness can be used to control for the effects of relatedness in population mapping.

Methods that combine both LD and linkage to fine map QTL have been developed [20] and utilized in commercial bovine and line cross porcine populations [21-23]. QTL are first mapped to a broad chromosomal location using a whole genome linkage analysis and then the QTL interval is refined through a denser marker panel and the linkage disequilibrium and linkage analysis (LDLA) process. A variance component framework is employed in LDLA that estimates variance components for the variances associated with the QTL and with the background genes. LDLA requires the inference of linkage phase of genotyped animals and prediction of IBD probabilities at putative QTL positions which can be computationally intense but is feasible with the small number of markers necessary to refine a QTL location.

Comparison of analysis methods was conducted to assess power, precision of location and error rates [24]. Methods that did not account for genetic background performed worse. The Bayesian approach performed best and did not require a correction for multiple testing as all markers were fitted at once. A close second was the mixed model analysis from the freely available DMU software package [25]. This method addresses genetic background through a polygenic genetic effect that is fitted as a random effect. Each SNP is successively tested thus a Bonferroni correction was applied. As the Bonferroni correction may be overly conservative, especially with increasing marker density [8] on a genome wide basis, significance can be assessed at the chromosome wide level which will not be as stringent due to fewer markers [24].

Selected Review of Trait mapping

Bovine technologies such as linkage, physical and comparative maps as well as genome sequences and SNP arrays have been employed as they have been developed. These technologies have been instrumental to map both monogenic recessive and

quantitative traits to chromosomal locations and even discover the underlying genetic variant in some instances.

Mapping of Recessive Monogenic Traits

Some of the first mutations underlying recessive defects were discovered by taking advantage of homologous phenotypes with known causative mutations in other species to determine an appropriate candidate gene for resequencing [26, 27]. Once microsatellite marker panels had been developed and linkage maps created [28-33] whole genome scans to discover regions of the genome associated with recessive phenotypes began [34]. These whole genome linkage scans utilized multi-generational pedigrees comprised of 46 to 163 total animals descending from one proband and utilized 17 to 51 affected offspring to map the associated region to a chromosomal location and to trace the region back through the pedigree to the putative founder individual [34-36]. Further fine mapping, comparative mapping and candidate gene selection was then required before likely causative mutations were discovered [37-39].

Association mapping methods and the discovery of thousands of SNP markers has taken mapping of recessive diseases to a new level often eliminating the need for large groups of familial individuals with numerous affected offspring and decreasing the occurrence of fine mapping.

High density SNP platforms have been used successfully to map recessive diseases in livestock species. Five separate recessive diseases were mapped to a specific genomic region using either a 25K Affymetrix array or a 60K iSelect panel. Causative mutations were identified for three of the defects using between three and twelve affected individuals [40]. Homozygosity mapping was utilized to identify a critical region for each disease. Strong candidate genes were identified within the homozygous region and then sequenced to identify causative mutations. A causative mutation was identified for four of the diseases allowing breeders to screen animals and avoid at risk matings [40, 41]. The critical map

region was as small 0.87 Mb for renal lipofuscinosis (RL) and as large as 11.87 Mb for ichthyosis fetalis (IF). Three highly related animals were used to map IF while twelve animals were used to map RL. The smaller RL critical region alludes to the necessity of a less highly related as well as a larger disease mapping population. Many recessive traits have been successfully mapped through the use of SNP platforms with as few as three affected individuals [40, 42-45].

Mapping of Quantitative Traits

Quantitative traits are influenced by many genes and thus it is not as straightforward to uncover underlying genetic variants as with monogenic recessive traits. Initial searches for genomic regions underlying quantitative phenotypes in cattle utilized disease tagged QTL where a simple recessive disease is known to be linked with a quantitative phenotype [46]. Since the QTL and recessive disease are linked, mapping the recessive disease also by default maps the quantitative phenotype.

The development of microsatellite linkage maps [28, 33] allowed whole genome scans for quantitative traits influencing milk production to take place using linkage analysis in daughter and granddaughter pedigree designs [31, 47]. Whole genome scans for carcass and growth traits soon followed using linkage analysis in cross bred pedigrees [48, 49]. Searches for the quantitative trait nucleotide (QTN) or genetic variation underlying QTL were aided by across species comparisons incorporating known DNA sequences from murine or humans to position candidate genes within QTL regions as well as determine the bovine sequence underlying specific QTL [50].

The use of linkage mapping methods and microsatellite linkage maps remains a viable method to undertake whole genome QTL scans but association mapping methods employing thousands of SNPs have been explored as tool to map QTL. Moderate size SNP panels such as the Affeymetrix 10K SNP array have been used successfully to map QTL through the use of both linkage and linkage disequilibrium but agreement between the two

methods was limited which was attributed to the difference in the way LD information was treated between the analyses, the moderate density of the marker panel limited power and that linkage disequilibrium methods are capable of detecting regions that are not revealed by linkage methods [51, 52]. The larger 50K SNP panel has also been utilized to investigate economically important traits in dairy cattle through linkage and linkage disequilibrium [53] and greater agreement was noted between the two methods when a larger SNP assay was used. The bovine SNP50 panel has been used with linkage disequilibrium methods to map QTL for milk production [54] and fertility traits in dairy [55] and growth traits in crossbred beef cattle [56]. Linkage disequilibrium mapping methods have been credited for their lower computational demands, the ability to detect effects both of moderate size and of low heritability traits as well as the ability to define narrower genomic regions [51, 57] when compared to linkage analyses.

Quantitative Trait Nucleotide Discovery

While markers found to be in LD with a quantitative trait can be used in marker assisted selection programs the ultimate marker is the actual QTN or the DNA variant that is actually responsible for the phenotypic variance seen. As the search for a QTN is a lengthy process and the causality of a DNA variant is often difficult to prove there are very few cases of validated QTN in livestock species.

Typically, the search for a QTN begins with a whole genome scan followed by fine mapping efforts then positional cloning and candidate gene selection. The whole genome scan often utilizes linkage within specific family structures in order to create LD between QTL and sparsely spaced genetic markers. The family structures are often backcross or F2 populations created from crosses between inbred lines however, in dairy breeds the daughter and granddaughter designs are popular as most selection programs are based on within-breed selection rather than across breed selection. As the whole genome scans involving linkage mapping methods usually result in large confidence intervals fine mapping

is necessary to narrow the target region. Fine mapping efforts may involve increasing marker density in the region and using both linkage and linkage disequilibrium or haplotypes to define a more precise interval. The next step involves candidate gene selection or positional cloning if genome sequence assemblies are lacking in the target region. Finally, sequencing of candidate genes and discovery of polymorphisms is undertaken. Any variant that is revealed is evaluated for possible causality of the phenotype and may require functional studies to validate the possible QTN. The DNA variants in acylCoA:diacyglycerol acyltransferase (DGAT1), insulin-like growth factor 2 (IGF2) and growth differentiation factor 8 (GDF8) have been deemed validated QTN due to evidence provided from functional studies [58-61].

The search for the acylCoA:diacyglycerol acyltransferase (DGAT1) mutation began with several whole genome scans using linkage mapping methods and sparsely spaced microsatellite markers [47, 62, 63] and a locus on BTA14 was identified that had a major effect on milk percentage as well as milk yield. Fine mapping of the region was completed using a combined linkage and linkage disequilibrium method [64] while others used linkage disequilibrium to refine the region to a 3 cM interval. As bovine genome sequence was not available, positional cloning of candidate genes in the region was conducted and a polymorphism was identified in a strong candidate gene, DGAT1, that was shown to have a major effect on milk characteristics [50]. DGAT1 was an obvious candidate for milk yield and composition due to known physiology as mice carrying a knock-out mutation for DGAT1 do not lactate [65, 66]. The identified polymorphism resulted in a lysine to alanine substitution in exon 8 of DGAT1 and the wild type allele was found to be highly conserved in other mammalian species adding to the evidence of the mutation being the causal variant. A functional study using a baculovirus expression system to express both the normal K and mutant A alleles found a significant difference in accumulation of triglyceride between the two alleles [58].

The second validated QTN, IGF2, was uncovered through a whole genome scan revealing a paternally expressed QTL with moderate effect on muscle mass on pig chromosome 2 [67]. Comparative mapping efforts revealed a strong candidate gene with a known role in myogenesis, insulin-like growth factor 2 (IGF2), [68, 69] while fine mapping using a haplotype approach narrowed the confidence interval and also proposed IGF2 as a candidate [59]. A nucleotide substitution was found in intron 3 which is in a conserved CpG island that is hypermethylated in skeletal muscle. An *in vitro* functional assay revealed that the mutant allele does not allow interaction with a nuclear factor while an *in vivo* functional study showed that pigs inheriting the mutation from their sire had a threefold increase in IGF2 mRNA [59].

The search for the third validated QTN began with a whole genome scan in a Romanov Texel sheep F2 population revealing a QTL with a major effect on muscle mass on ovine chromosome 2. The interval was refined through fine mapping efforts and comparative mapping revealed a strong candidate gene, growth differentiation factor 8 (GDF8), in the region. Sequencing of GDF8 did not reveal any mutations within the exonic or intronic regions of the gene but a G to A transition was discovered in the 3' untranslated region. This polymorphism creates a miRNA target site for miR1 and miR206 which are both expressed in skeletal muscle. Functional analysis revealed a threefold difference in the ratio of GDF8 protein in serum of wild-type versus heterozygous (G/A) sheep and a 1.5 fold difference in relative abundance of G versus A transcripts in skeletal muscle of heterozygous sheep [60]. Further functional analysis confirmed that mRNA transcripts with the A allele more tightly associate with RNA induced silencing complex than wild type transcripts [61].

These validated QTN that have been identified in livestock species all shared similar paths to defining a QTN. All involved a whole genome scan to identify regions of the genome associated with a quantitative trait, fine mapping of the QTL region to define a smaller target region, comparative mapping with candidate gene selection based on known function and evaluation of DNA variants for causality of the phenotype.

Genotype imputation

Genotype imputation is the practice of predicting genotypes at markers that were not directly assayed. This allows an increase in marker density, reduction of genotyping costs or facilitates meta-analysis. Genotyping costs may be reduced by using a high density panel on select individuals to impute genotypes into a set of individuals that were genotyped with a subset of markers. Imputation may be used to increase the number of SNPs available in a region for fine-mapping purposes or across the genome for association analysis thus increasing the resolution of the study. Meta-analysis may also be facilitated through the use of imputation of markers that differ between marker panels used in different studies.

The algorithms behind imputation methods differ but all require haplotype phasing of each individual that was genotyped. A diploid individual carries two copies of each chromosome known as homologous chromosomes. Phasing of the genotypes determines the set of alleles that are present on each of the homologous chromosomes, also known as a haplotype. Most imputation methods require a phased reference panel of individuals that are completely genotyped on a denser marker panel. The phased haplotypes of the group to be imputed are compared to the phased haplotypes of the more dense reference panel in order to impute the untyped or missing genotypes by matching similar haplotypes between the groups.

Several haplotype phasing methods have developed for use with unrelated individuals that are based on the expected maximization (EM) algorithm, coalescence and hidden Markov model (HMM) as well as long range identity by descent. Phasing methods based on the EM algorithm, such as those used in Arlequin and PL-EM [70], are useful with a small number of polymorphisms within a short region but are not computationally feasible with large scale genome wide data. Approximate coalescent models and HMMs underlie many of the population based methods. The approximate coalescent model is a method by

which the ancestry of alleles converges when looking back in time. This model recognizes that mutation and recombination result in the derivation of new haplotypes from old haplotypes. A HMM is developed from the approximate coalescent model and the HMM parameters are estimated iteratively, often through a stochastic EM algorithm [71]. The combination of approximate coalescent models and HMMs are utilized in software such as PHASE [72], fastPHASE [73], MACH [74] and IMPUTE2 [75]. While PHASE is accurate it is only computationally feasible with small sample and marker sets. FastPHASE is capable of handling genome-wide datasets but is slightly less accurate than PHASE. Both MACH and IMPUTE v2 can handle large sample sizes and have better accuracy than fastPHASE [76]. While MACH and IMPUTE v2 have been used mostly for imputation of untyped markers they can be used for haplotype phase inference. The HMM algorithm also underlies the software Beagle [77] which is a user friendly software that is similar in accuracy as FastPHASE and IMPUTE v2 [78, 79].

Chapter 2. A deletion mutation in KRT71 is associated with congenital hypotrichosis in Hereford cattle.

Abstract

Background

Hypotrichosis, an autosomal recessive form of hairlessness, has been recognized within the Hereford breed of cattle for more than 30 years. Calves affected with hypotrichosis are born with partial absence of hair and a hair coat that is "curly or fuzzy" in texture. Curly hair, at birth, is most pronounced on the head and neck. Adult animals may periodically display a "patchy" hide where hair has been lost and re-growth is slow. As the hair coat is responsible for thermoregulation and protection from the environment, large areas of missing hair can be detrimental to animal welfare and fitness.

Results

A whole-genome association analysis was conducted using 42,585 SNP markers from the Illumina® Bovine SNP50 BeadChip. Significant association was detected between the hypotrichosis phenotype and a locus on bovine chromosome 5 (BTA5). Homozygosity analysis indicated the associated region was between 29 and 32 Mb. Six candidate genes with known hair or hair follicle function were sequenced in known normal, carrier and affected animals. An eight base pair deletion mutation was found in exon 1 of the bovine *KRT71* gene that was consistent with putative genotypic status within archived samples. The deletion mutation is predicted to result in a frameshift and early truncation of the K71 protein.

Conclusion

The K71 protein is a hair follicle specific protein that is thought to participate in the formation of structurally important keratin intermediate filaments within the inner root sheath of the hair follicle. These filaments are integral to support and shape the growing hair shaft. A loss-of-function mutation in bovine KRT71 is associated with congenital hypotrichosis in Hereford cattle. A DNA based diagnostic was developed that will permit animals to be screened for the mutation and thus allow animal breeders to make informed mating decisions and decrease the incidence of hypotrichosis in the Hereford breed.

Background

Partial or complete absence of hair coat at birth, known as congenital hypotrichosis, has been reported in many domestic species and the term is broadly applied to a variety of mutations resultant in a similar dermatologic phenotype [80]. Bovine hypotrichosis has been described singularly, as well as in conjunction with other phenotypes in cattle breeds such as Friesian, Holstein-Freisian, Guernsey, Hereford, Jersey, Pinzgauer and Charolais [80]. Hypotrichosis in cattle has been observed in association with anodontia [81, 82], abnormal hair pigmentation [82-84], abnormal thyroid [85], anemia [86] and abnormal body size or weight [87-89]. Furthermore, it has been described as both viable [80, 84, 87, 88, 90-93] and lethal [82, 94] and inherited as an autosomal recessive [83-85, 87, 88, 90, 91, 93-96], dominant [81] or sex-linked [81] condition.

Hereford animals affected with the hypotrichosis (HY) phenotype represented in this study are born with partial or complete absence of hair that later becomes “fuzzy or kinky” in appearance. Hair loss becomes most pronounced on the poll, stifle, brisket, shoulder and neck as a calf ages. Affected calves are more susceptible to sunburn, hypothermia and reportedly have more frequent dermatophyte infections. Affected adult animals have a hair coat that may be “patchy” in appearance. This Hereford phenotype appears to be limited to

the hair coat with microscopic analysis showing abnormal hair development. Characteristic large atypical trichohyalin granules form in the inner root sheath and are associated with premature breakup of the sheath and loss of the hair [80, 84, 86, 97]. Electron microscopic studies demonstrated absence of intermediate filaments in these trichohyaline granules [80]. Pedigree analysis indicates an autosomal recessive mode of inheritance [81, 84, 87, 88, 93].

Animal productivity has increased greatly in the last 50 years due to the use of estimated progeny differences and rigorous selection criteria. The use of artificial insemination has allowed the genetics from popular sires to be distributed throughout the population. Strict selection criteria and artificial insemination allow for vast improvements within a breed but decreases in the effective population size are seen when a limited number of sires are relied upon heavily. A reduction in effective population size increases the amount of inbreeding within a population, thus providing an ideal environment for the emergence of recessive diseases. This has been demonstrated in the Holstein-Friesian breed that has millions of animals but an estimated effective population between 49 and 125 [98, 99]. The recessive disease bovine leukocyte adhesion deficiency (BLAD) has been estimated to be carried by 14% of Holstein-Friesian sires [100] while complex vertebral malformation (CVM) has been estimated to be carried by ~25% of Holstein-Friesian sires [101]. Both diseases are thought to be disseminated in the population by one popular founder sire [100]. In 1992, it was estimated that 0.2% of Holstein animals born in the US were affected with BLAD resulting in an annual loss of $5 million [100].

Marker assisted selection utilizes DNA based diagnostic tests to aid in selection of animals based on the association of a phenotype with a genetic variant. Marker assisted selection can be used to select towards or away from a particular phenotype through the use of either predictive or causative markers. Predictive markers are genetically linked to the mutation that is responsible for the phenotype. While predictive markers are easier to

develop there is opportunity for recombination to occur between the marker and the causative mutation that would eliminate the effectiveness of the marker as a diagnostic for the phenotype. Directly screening for the causative mutation is preferred as recombination events will not be an issue thus a consistent association between the genotype and phenotype will always be observed.

Marker assisted selection has been used successfully in a German Holstein population to select against sires carrying the recessive BLAD and CVM alleles. The frequency of BLAD carrier sires has decreased from 9.4% in 1997 to 0.3% in 2007 and CVM carrier sires has decreased from 8.3% in 2002 to 2.3% in 2007 [102]. This approach is considered a successful method to deal with autosomal recessive diseases [103].

Traditionally, mapping of recessively inherited diseases required extensive pedigrees and many affected offspring in order to accurately position a disease to a specific chromosomal location. However, the completion of the bovine genome sequence and the emergence of new high-density SNP platforms such as the Illumina® Bovine SNP50 platform (Illumina, Inc. San Diego, CA) allow disease mapping with as few as three affected individuals [40, 43, 44, 104-106].

The HY disease locus was mapped to a locus on BTA5 using Illumina® Bovine SNP50 platform. Candidate genes were sequenced and a deletion mutation consistent with expected genotypic status in archived samples was discovered. A DNA based diagnostic was developed that will allow animals to be screened for the mutation.

Results and Discussion

Hypotrichosis in Herefords

Hypotrichosis in Herefords has been has been characterized as a viable, recessively inherited, semi-hairless condition [88, 93] with affected individuals displaying a short, fine, curly hair coat at birth that became wiry with age [87]. The skin is described as thin and

easily pliable with a sparse hair coat on the lateral and ventral aspects of the body [80, 107]. Additionally, the Huxley's layers present with abnormally large trichohyalin granules that lack the normal associated micro and macro filaments. Hair from hypotrichotic Herefords is also shown to be more soluble than hair from normal Hereford or Holstein animals; when treated with 10% tetramethylammonium hydroxide (TMAH), hair from hyptrichotic individuals is more soluble than hair from normal animals [108]. As TMAH solubilizes some keratin proteins [109], this implies a structural difference in the cortical keratin proteins.

The mapping population in this study consisted of 18 affected calves with eight of those tracing back to one bull known to be a carrier of hypotrichosis. The remaining calves were from two separate groups with the first being comprised of three affected half siblings and two distantly related calves. The second group contained five affected calves from a single herd of unknown relatedness.

A whole genome association and homozygosity analysis [110] revealed association with the hypotrichosis phenotype on BTA5 (Figure 2.1A). Homozygosity analysis of the SNP haplotypes revealed extended regions of homozygosity on BTA5 in eight affected calves (Figure 2.2) with a consensus region from 29.41 Mb to 31.93 Mb (Btau4.0). Manual inspection of the SNP genotypes revealed a region of homozygosity from 29.77 Mb to 30.36 Mb (Btau4.0) in thirteen supposed affected calves (Figure 2.2) while microsatellite genotypes (Table S1, See supplementary file: Markey_thesis_supplement.xlsx) characterized the region as between 29.53 Mb and 30.45 Mb (Btau4.0). Four calves (Group C) were excluded from this inspection as they were heterozygous in this region (Table S1, See supplementary file: Markey_thesis_supplement.xlsx). Removal of these four calves from the WGA analysis resulted in greater significance and less background (Figure 2.1B) when compared to the WGA including all reported affected calves (Figure 2.1A). Two associations of the hypotrichosis phenotype with SNPs were found on BTA7 near 67 Mb but

further investigation revealed neither regions of homozygosity in the affected calves nor obvious candidate genes in the region.

Candidate genes were selected based on function and location relative to the homozygosity analysis of the affected calves (Figure 2.2). Six candidate genes were re-sequenced: *KRT72*, *KRT71*, *KRT74*, *KRT83*, *KRT86*, and *KRT81*. An eight base pair deletion (TGTGCCCA, Figure 2.3) of nucleotides 334 to 341 [Genbank: NM_001075970.1] was found in exon 1 of the *KRT71* gene which was consistent with expected genotype status. Small deletion mutations are thought to result from slipped mispairing during replication and the deleted segment is often flanked by direct repeats between 2 and 11 base pairs in length [111]. The HY deletion is flanked by a direct GTGCCCA repeat that may indicate that the mechanism underlying the deletion was due to slipped mispairing. At the level of translation the deletion mutation is predicted to result in a frameshift after AA 93 with a stop codon at AA 107, which is significantly shorter than the normal bovine K71 protein that is 525 AA in length (Figure 2.4). If the mutant mRNA transcript were not targeted for nonsense mediated decay [112] then the resultant truncated protein would not contain the alpha helical rod domain or the helix boundary motifs that are integral to heterodimer formation. This truncated protein would likely be unable to participate in filament formation resulting in instability in the cytoskeletal network and a lack of tissue integrity [113]. Electron microscopy of HY affected calves revealed large atypical trichohyaline granules in the inner root sheath confirming a lack of intermediate filaments (Figure 2.6B). This would result in a weakened attachment of the hair to the follicle, thus affected animals would be likely to exhibit progressive hair loss on high friction areas such as the neck, shoulder, stifle and poll regions.

All animals within this mapping population of Herefords genotyped as expected for genetic status with the exception of four purported affected calves and their direct relatives. The four calves (23, 110, 286 and 287) did not have a deleterious mutation consistent with

their reported genotypic status in the six keratin genes that were resequenced. It is probable that these four calves were misclassified based on phenotype and were not affected with the same type of congenital hypotrichosis as ultrastructural examination of granule for lack of intermediate keratin filaments was not part of initial phenotypic classification. Clinically similar but unique phenotypes of hypotrichosis occur in cattle and have been seen infrequently within the Hereford breed. In human disorders, there is a clinical overlap between autosomal recessively inherited hypotrichosis and monilethrix as both can vary in severity from small affected patches to total alopecia with hairs that are easily broken [112]. A similar situation may have occurred here especially if different veterinary practitioners were involved in the diagnosis of hypotrichosis.

Calf 101 was homozygous for the HY allele but did not exhibit an extended region of homozygosity surrounding *KRT71*. Further investigation revealed this calf was only homozygous for 4 SNPs (~300 Kb) around the *KRT71* gene, which attests to the age of the mutation as linkage disequilibrium in a region will decay with each subsequent generation thus resulting in the smaller homozygous region seen in calf 101. One group of affected calves descends from a bull born in 1982 while the other group descends from a bull born in 1970. Based on pedigree information, there are no common ancestors between the two bulls. The oldest carrier bull we found was born in 1967 indicating that the proband was likely born prior to 1967. As there are instances of similar hypotrichotic phenotypes described in Herefords in 1934 and 1950 [87, 88] the proband may be much older than the heterozygous HY sires we have discovered. A population of 286 Hereford sires was examined and the frequency of the HY allele was found to be 0.0829.

Hair follicle and hair formation

The hair follicle is responsible for the production of hair that occurs in a cyclical fashion undergoing three stages: the anagen phase of active growth, the catagen phase

with cessation of growth and the telogen phase, a dormant stage, where the hair is shed [114, 115]. Bovine hair follicles are first formed at around 77 days of gestation with hairs emerging around 203 days of gestation. The first hair cycle in the cow is about 200 days in length while the mouse is only approximately 25 days. Only one type of follicle is observed in cattle although there is some variation in follicle size. Each follicle is associated with a sweat gland, sebaceous gland and arrector pili muscle [116].

The hair follicle is composed of eight layers. The outermost layer is the connective tissue sheath moving medially is the outer root sheath, the companion layer and the inner root sheath (IRS) which is composed of the Henle's layer, Huxley's layer and IRS cuticle. The innermost layer, the hair shaft, is composed of the hair cuticle, cortex and medulla [113]. The companion, Henle's and Huxley's layers as well as the IRS cuticle are thought to envelope the hair shaft and function to support and mold the hair into the proper shape during anagen [113, 117-120].

Hair loss usually results from an anomaly of hair follicle cycling but congenital hair defects are generally caused by mutations in structural proteins such as keratins [121]. In humans, there are 54 keratin genes 21 of which are hair or hair follicle specific [112]. Autosomal recessive forms of hypotrichosis have been associated with mutations in the structural proteins CD5N, DSG4, and P2RY. Mutations within KRT81, KRT83 and KRT86 have been implicated in monilethrix, an autosomal dominant disorder of the hair cortex. Affected individuals have hairs that are beaded in appearance and break easily. There is a large variation in expressivity with some phenotypes being visible to the naked eye while others are almost undetectable under electron microscope. In humans, there is a clinical overlap between localized autosomal recessive hypotrichosis and monilethrix that may lead to misdiagnosis [112].

Structure and function of hair specific keratins

Keratins are the main structural component of the hair follicle and individual keratins are specific to each layer of the hair follicle [114, 122]. The keratin family consists of over 40 members divided into two main subcategories of epithelial or hair keratins. Further distinction is made according to charge properties, which provides classification of type I and type II keratins. Type I keratins have an acidic charge and are generally smaller while type II keratins, including K71, have a basic charge and are larger [114]. Keratin proteins are comprised of a head domain, an alpha-helical rod domain and tail domain.

Structural keratins of the hair follicle undergo obligate pairing between a type I and a type II keratin aligned in parallel to form a heterodimer. Heterodimer molecules align in an anti-parallel staggered fashion to form tetramers that bind together through the head and tail domains to form filament chains [123, 124]. These chains are known as keratin intermediate filaments (KIF) due to their intermediate size of 7 to 12 nm, which is smaller than microtubules but larger than microfilaments. Intermediate filaments are non-polar, thus are unlikely to participate in cellular transport but do serve as scaffolding for the cytoskeleton [125]. Assembly of the KIF may be severely affected by mutations in the keratin genes that change the AA sequence [126] and may result in failure of tissue integrity due to the lack of stable heterodimer formation of the mutated keratin with its respective partner [118].

Keratin heterodimer formation takes place through associations between the alpha-helix rod domains of a type I and type II keratin monomer. The boundaries of the alpha-helix rod domain contains helix initiation (HIP) and termination (HTP) peptides that are necessary for proper heterodimer formation [115]. The HIP and HTP motifs are host to mutations that result in many of the human keratin diseases [112] and filament elongation is disrupted by mutations within these peptides [112, 127].

The KIF are believed to associate with trichohyalin which aligns the filaments along the long axis of the IRS cells [128]. Mature IRS cells contain a layer of trichohyalin

surrounding individual KIF separating them from the cell wall [128, 129]. Trichohyalin cross-links and binds to the head and tail regions of the KIF to add support and promote the longitudinal alignment to the hair shaft. The KIF, containing K71, extend from the inner Huxley's layer through the outer Henle's layer where the filaments attach to the surrounding companion layer at desmosomes [118]. The interaction of desomosomes and cytoskeletal network is critical for preservation of tissue integrity [130].

The K71 protein is a type II, hair follicle specific, epithelial keratin [131] that is produced by keratinizing cells of the IRS in humans [113]. Cells of the Henle's and Huxley's layers as well as the IRS cuticle synthesize K71 in humans [123], but in mice the protein is expressed only in the Henle's and Huxley's layers [114]. Expression of keratins in bovine and human tissues is very similar with respect to charge, size and immunoreactivities [132]. The bovine K71 protein has three domains: amino acids 1-131 encompass the head domain; 132-441 the rod domain; 442-525 the tail domain. The human K71 protein forms a heterodimer with the IRS specific proteins K25, K27 or K28 [125]. A single keratin monomer that does not have an equimolar proportion of its heterodimer partner is usually degraded [133] however, it has been proposed that the association of type I keratins with the cellular membrane may be sufficient to prevent this degradation [134].

Keratin mutations

The location of mutations within keratin genes seems to play a role in the severity of the pathology as well as the mode of inheritance. The well studied epidermolysis bullosa simplex (EBS), a blistering disorder of the skin, can provide insights to the mechanism behind mode of inheritance and severity of keratin disorders. Severe dominant forms of EBS are usually associated with mutations within the essential HIP or HTP of either K5 or K14, while milder dominant forms of EBS are coupled with mutations outside of the helix boundary motifs, such as the 1A, L12 and central 2B sub-domains of the rod domain. These

dominant mutations do not prevent intermediate filament formation or elongation but are a source of structural weakness. While dominant forms of EBS involve the incorporation of both normal and mutant proteins in the intermediate filaments, recessive forms of EBS are unlikely to incorporate mutant proteins in the intermediate filaments. These recessive mutations seem to arise from premature termination codons that likely result in transcripts targeted for nonsense mediated mRNA decay and thus, no mutant protein will be present in the cells [112].

The theory of nonsense mediated mRNA decay is supported by the Y204X mutation within KRT14, which leads to a premature termination codon in sub-domain 1B of the rod domain. The presence of K14 was undetectable in the skin and likely resulted from nonsense-mediated mRNA decay of the premature termination codon containing K14 transcript. Overall, twelve other incidences of recessive forms of EBS have implicated premature termination codons resulting in no detectable levels of K14 in the skin [135].

Mutations within KRT71 influencing hair coat have been discovered in canine, rattine, feline and murine species [134, 136-138] and exhibit both an autosomal dominant [134, 136, 137] and an autosomal recessive [134, 138] mode of inheritance. Wavy hair coat in dogs is inherited in an autosomal dominant manner and is controlled by a nonsynonymous mutation in KRT71 that alters one AA in the rod sub-domain 1A of the K71 protein possibly affecting either a coiled-coil domain or a pre-foldin domain. Dog breeds that are fixed for curly hair were homozygous for the mutation while breeds that segregate the curly hair phenotype exhibited all three genotypic states [136]. In rats, the Rex phenotype is an autosomal dominant trait with heterozygous animals displaying a wavy hair coat while homozygous animals exhibit complete loss of body hair and as adults display a sparse hair coat. The rat Rex mutation is a 7 bp deletion in KRT71 encompassing the 5’ splice site of exon 2 causing a deletion of six amino acids from the rod sub-domain 1A of the K71 protein [137]. Both the dog and rat mutations lie outside of the helix boundary motifs but are within the second half of the rod sub-domain 1A that has been associated with milder forms

of EBS in humans [112]. Wavy hair in mice is also controlled in an autosomal dominant manner with several mutations underlying the phenotype. The mouse KRT71 mutations Ca-J (also known as Ca-10J, Ca-Rin and Ca-9J), Rco12 and Rco13 involve a single AA in the HIP motif while the dominant Ca and Rgsc689 mutations involve a single AA in the HTP motif [139]. These dominant K71 mouse mutations have a similar position to those underlying mild forms of EBS in humans that are located in the helix boundary motifs of K5 and K14.

Mutations in KRT71 have been implicated in hair coat phenotypes that are inherited in an autosomal recessive manner in felids. The Devon Rex breed of cat is known for its curly coat that is often easily broken and frequently lacks guard hairs. The mutation underlying the Devon Rex phenotype is an 81bp deletion resulting in a K71 protein lacking 35 AA from the central portion of the rod sub-domain 2B [138]. Although the Devon Rex mutation does alter the K71 protein it is thought that residual activity of the protein still exists however, the mutated K71 filaments are structurally weakened allowing them to be easily broken [112]. The Sphynx breed of cat is hairless but sometimes displays fine down on the body or a wavy coat on the nose, tail and toes. A single base substitution was discovered in KRT71 that alters the 5' splice site of exon 4 resulting in adaption of an alternative splice cite downstream and a premature termination codon in the K71 protein [138], hence showing similarity to many recessive EBS phenotypes that arise from PTC [135]. The resultant K71 protein is 281 AA in length, thus the majority of the alpha-helical rod domain is absent, including the HTP motif that is essential for keratin type I and type II binding. Two separate genotype states underlie the phenotype in the Sphynx animals investigated. The first is the Sphynx allele in a homozygous state and the second is a compound heterozygote with both the Sphynx and Devon Rex alleles [138]. As the Sphynx originated from the Devon Rex it is should not be surprising that the Devon Rex mutation is present in the Sphynx population.

A KRT71 mutation, Rco3, has been discovered to also have a recessive mode of inheritance in murine species and exhibits many similarities to the HY mutation. The Rco3 mutation is a 10 base pair deletion resulting in a frameshift after AA residue 58 and a protein that is shortened to 134 AA in length due to incorporation of a PTC. The AA residues 59-134 show no similarity to any known or predicted protein. This is similar to the HY allele that is predicted to result in truncation of the K71 protein after AA residue 107 with residues 94 -107 showing no similarity to keratin proteins. Both truncated proteins would lack the rod and tail domains as well as the helix boundary motifs that are critical for heterodimer formation. Thus, if a truncated K71 protein is produced, it is unlikely that the mutant protein would participate in correct heterodimer formation. As there were no detectable levels of K71 in the IRS of the Rco3 mice [134], it is likely that the K71 transcript was targeted for nonsense mediated mRNA decay [112].

Further similarities between the Rco3 and HY mutations are seen when the phenotypes are examined. The Rco3 mutation is also inherited in an autosomal recessive manner and affects the skin, coat and nails as well as the touch and vibrissae systems. The first body coat of Rco3 homozygous mice was curly in nature while the second body coat displayed progressive alopecia that was severe and patchy. Hair texture was abnormal with the cortex displaying frequent kinks and twists. The homozygous Rco3 mice showed defective keratinization of the Henle's and Huxley's layers of the IRS. The Henle's layer of Rco3 mutants displayed accumulation of electron-dense material as well as lack of normal filament bundles that is likely due to the inability of the truncated K71 protein to form heterodimers and intermediate filaments [134].

Conclusion

A small deletion mutation was found in the bovine *KRT71* gene that is associated with the congenital hypotrichosis phenotype in Hereford cattle. The mutation is predicted to

result in a premature stop codon that will produce either an mRNA transcript that is degraded or a truncated protein that is unable to participate in proper heterodimer formation in the inner root sheath of the hair follicle. It is expected that formation of the keratin intermediate filaments necessary for structural support and hair shaft molding will be prevented. A DNA-based diagnostic was developed that will allow breeders to select against this recessive defect in the Hereford population while retaining valuable breeding stock that may otherwise be culled due to suspect pedigrees.

Methods

Samples

Samples of affected and known carrier animals were collected over a four year period from various herds. DNA was isolated from blood and semen samples using a simple salting out procedure [140] and stored at -20°C.

Eighteen affected animals and 54 known normal animals were genotyped using the Illumina® Bovine SNP50 platform (Illumina, Inc. San Diego, CA) at the University of Missouri. Results from one affected calf were removed due to poor genotyping quality. Microsatellite mapping was completed with 18 affected and 22 suspected carrier animals. Resequencing of candidate genes was completed for two normal, three suspected carrier and three affected animals.

Mapping of disease locus

Whole genome association and homozygosity analysis was completed using PLINK [110]. The "--assoc" and "--mperm" commands were used for the whole genome association analysis and to obtain the corrected empirical P-value (max(T)) with 10,000 permutations. The commands "--homozyg-group" and "--homozyg-verbose" were used to group the pools of overlapping segments and display the genotypes for each pool. The

command "--mind 0.15" allowed 15% missing genotypes per individual while the command "—maf 0.01" set the minor allele frequency to 1%. The command "--cow" set the chromosome codes for the cow. The command "--allow-no-sex" was used to allow ambiguously-sexed individuals. The command "--homozyg-density 100" was used to allow 1 snp per 100 kb.

Microsatellite marker development and fragment analysis

Microsatellites were selected with the simple sequence repeat identification tool (SSRIT) [141], sequences were masked using RepeatMasker [142] and primers were designed using Primer Designer v 2.0 (Scientific and Educational Software). PCR products were fluorescently tagged via an M13 protocol [143], amplified in multiplex PCR and fragment analysis performed using an ABI Prism 3730xl Analyzer (Applied Biosystems, Foster City, CA). Microsatellite genotypes were analyzed with GeneMarker™ (Softgenetics®, LLC, State College, PA).

Microsatellite multiplexes were amplified in 10 µl reactions using ~20-50 ng genomic DNA with final concentrations of 1x PCR buffer (containing 1.5 mM $MgCl_2$; Qiagen, Valencia, CA), 200 µM each dNTP (Fermentas), 0.0125 µM each M13 tailed primer, 0.25 µM each standard primer, 0.15-0.25 µM each fluorescently labeled M13 primer (Applied Biosystems, Foster City, CA) and 0.35 units HotStarTaq *Plus* DNA Polymerase (Qiagen, Valencia, CA). Cycling parameters varied but typically included initial activation of 5 min at 95°C followed by 35 cycles of 60s at 94°C, 90s at 60°C and 90s at 72°C followed by a final extension for 1 hour at 72°C then 5m at 10°C.

Multiplexed products were combined and purified. Four µl of each PCR reaction was mixed with five times the volume of isopropanol, applied to Promega Wizard® SV96 Binding Plates and submitted to vacuum until dry. Products were washed four times with 200 µl of 80% Ethanol and vacuumed until dry after each wash. After a final drying period

of four minutes products were eluted with 80 µl Optima water and 10 µl of elute was dehydrated and submitted to fragment analysis.

Primer design and PCR

All primers were designed using Primer Designer v 2.0 (Scientific and Educational Software). Sequences were masked for repetitive elements using RepeatMasker [142] prior to primer design. Primer sequences are provided in Table S2 (See supplementary file: Markey_thesis_supplement.xlsx).

Candidate gene analysis

Candidate genes were selected based on known function in hair and location relative to the homozygous region found in the affected calves on BTA5. Exons were determined using mRNA sequences from NCBI and the software SPIDEY [144]. Primers were designed to amplify exons in 1kb fragments. Exons were amplified in 20µl reactions using ~ 20-50 ng genomic DNA with final concentrations of 1x PCR buffer (containing 1.5 mM $MgCl_2$; Qiagen, Valencia, CA), 200 µM of each dNTP (Fermentas, Glen Burnie, MD), 0.5 µM of each primer and 0.5 units of HotStarTaq *Plus* DNA Polymerase (Qiagen, Valencia, CA). Cycling parameters varied for each amplicon, but typically involved an initial activation of 96°C for 5 minutes followed by 35 cycles of 95°C for 45 s, 56-62°C for 45 s and 72°C for 60 s followed by a final extension at 72°C for 10 minutes. Amplified PCR products were subjected to electrophoresis on a 1.2% 0.5X Tris-borate EDTA agarose gel to verify amplicon quality and primer specificity. Ten µl of amplified PCR product was then treated with 4 µl of a 1:12 dilution of ExoSAP-IT® (Affymetrix, Inc., Santa Clara, CA) and incubated at 37°C for 30 minutes then heat inactivated for 15 minutes at 80°C.

Re-sequencing was performed in 8µl reactions using 2µl ExoSAP-IT® treated PCR product, 3.62µl buffer (0.16 M Tris (pH 9.0), 3 mM $MgCl_2$, 4.9% tetramethylene sulfone,

0.0001% Tween-20® surfactant), 0.25µl BigDye v3.1, 0.08 µl BigDye dGTP v3.0, and 0.656 µM of each primer. Sequencing was done with an initial denaturation of 96°C for 90 s, followed by 55 cycles of 96°C for 15 s, 53°C for 15 s and 60°C for 3 minutes with a final extension of 60°C for 10 minutes. Sequence products were purified using Sephadex™ G-50 Fine (GE Healthcare) and ran on an ABI 3730XL (Applied Biosystems, Foster City, CA) capillary sequencer. Bases were called using Phred, sequences assembled using Phrap and sequence assemblies viewed and analyzed for polymorphisms using Codon Code Aligner (Codon Code Corporation).

Mutation assay

Primers were designed flanking the deletion (Figure 2.5) using Primer Designer v 2.0 (Scientific and Educational Software). The primer located 5′ to the deletion was fluorescently tagged via an M13 protocol [143]. PCR was performed in 10 µl reactions using 20-50 ng genomic DNA with final concentrations of 1x buffer buffer (containing 1.5 mM $MgCl_2$; Qiagen, Valencia, CA), 200 µM of each dNTP (Fermentas), 0.0125 µM M13 tailed primer, 0.25 µM standard primer, 0.15 µM M13 (Applied Biosystems, Foster City, CA) and 0.25 units HotStarTaq *Plus* DNA Polymerase (Qiagen, Valencia, CA). Cycling parameters included initial activation of 5 min at 96°C followed by 32 cycles of 30s at 95°C, 30s at 60°C and 45 sec at 70°C followed by a final extension of 1 hour. Two µl of each product was diluted in 100 µl Optima water and 3 µl of diluted product was submitted to fragment analysis using an ABI Prism 3730xl Analyzer (Applied Biosystems, Foster City, CA).

Abbreviations

AA: amino acid, BLAD: bovine leukocyte adhesion deficiency, CVM: complex vertebral malformation, EBS: epidemolysis bullosa simplex, HIP: helix initiation peptides, HTP: helix termination peptides, HY: hypotrichosis, IRS: inner root sheath, kb: kilobase, KIF: keratin

intermediate filaments, KRT71: Keratin 71 protein, *KRT71*: Keratin 71 gene, m: minute, Mb: megabase, s: second, TMAH: tetramethylammonium hydroxide

Additional materials

Supplementary file: Markey_thesis_supplement.xlsx

Table S1. Genotypes of reported Hypotrichosis affected calves on BTA5

The table contains SNP and microsatellite genotypes of the purported Hypotrichosis affected calves used in the analysis. The Mb position is based on Btau4.0. Tan shading indicates a run of homozygosity and yellow shading indicates a heterozygous genotype bordering a run of homozygosity. A red shaded calf indicates the calf was not heterozygous or homozygous for the HY deletion mutation. Group A represents calves resulting from the initial PLINK homozygosity analysis while group B represents calves that resulted from relaxed homozygosity commands. Group C represents calves that did not appear in either of the homozygosity analysis from PLINK.

Table S2. Primer sequences used for Hypotrichosis region on BTA5

Primer sequences and annealing temperatures used to amplify exon, microsatellite and HY mutation assay amplicons.

Figures

Figure 2.1. Manhattan plots of results from whole genome case/control association analysis. Each chromosome is color coded along the X axis. The Y axis represents the –Log10 of the corrected empirical P-value (max(T)) after 10,000 permutations from PLINK **A)** All purported HY affected calves **B)** KRT71 deletion calves p = 0.0001

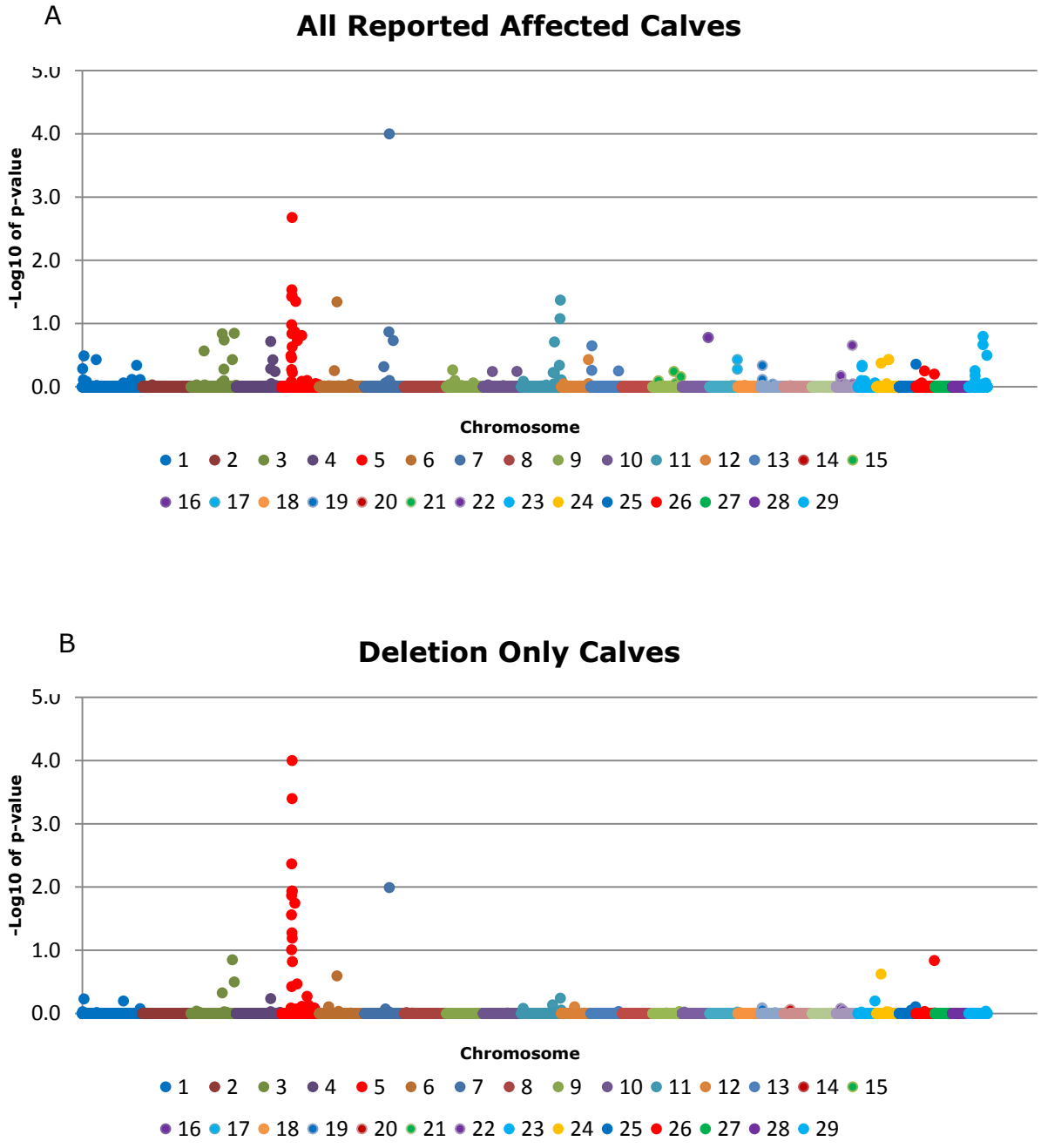


Figure 2.2. Schematic representing the genotypes of affected calves on BTA5. Each horizontal lane represents one calf with blue shading indicating a run of homozygosity. Grey shading indicates a heterozygous genotype and white indicates a homozygous genotype that is not within a run of homozygosity. The approximate Mb position based on Btau4.0 is indicated below the figure. The red box encompasses the consensus homozygous region which spans approximately 0.8 Mb. The genes that are encompassed within the consensus homozygous region are listed at the top of the figure along with corresponding approximate Mb positions. The asterisk (*) indicates the position of KRT71.

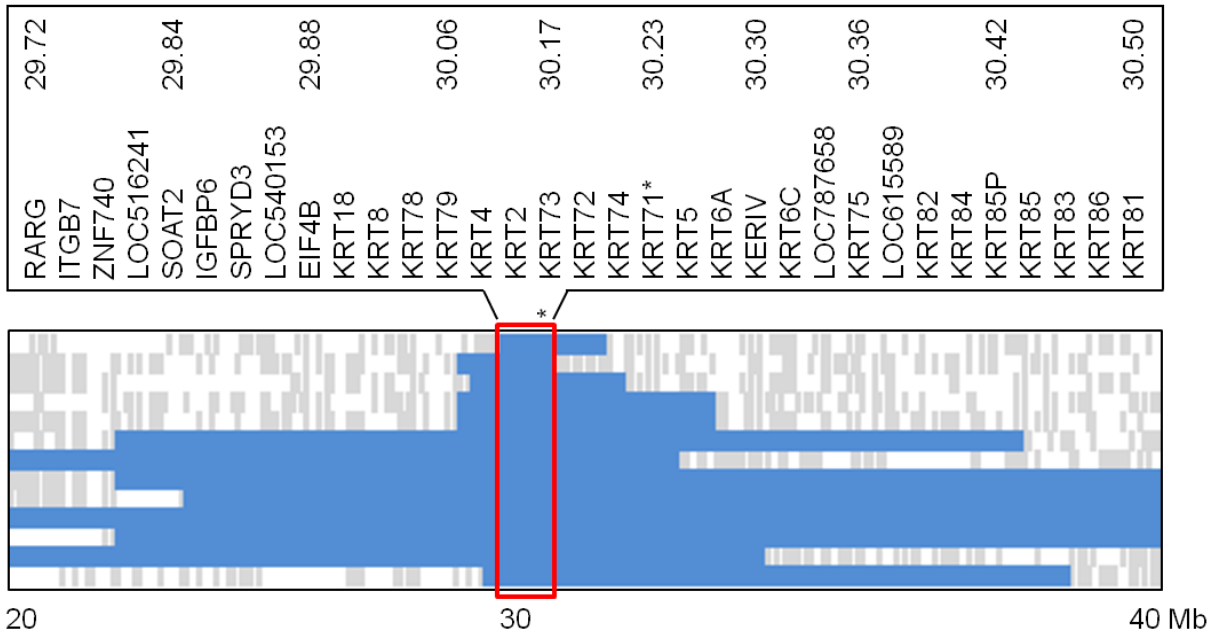


Figure 2.3. Alignment of genomic DNA sequences encompassing exon 1 of the bovine KRT71 gene. Arrows indicate the exon 1 boundaries. The expected translation of the normal and HY allelic sequences is provided. The deleted bases of the HY allele are represented by a dash (-). A star (*) indicates a stop codon. Note that the HY allele results in a stop codon in exon 1 of the KRT71 gene. Underlined bases indicate the primer sequences used for the DNA based diagnostic test.

```
                                                                 exon 1
                                                                    ↓
Normal        CATATAAAGGAGGGCACCTGCCAGTCCTCACTACAACCTGCAGAACCGTGTGGGAATTTG
HY            CATATAAAGGAGGGCACCTGCCAGTCCTCACTACAACCTGCAGAACCGTGTGGGAATTTG

                                                          M  S  R  Q  F  T
Normal        CCTTCGTTCCTCCAGCATCCAAGCTCCATCTCCACCAGAAGCATGAGCCGCCAATTCACC
HY            CCTTCGTTCCTCCAGCATCCAAGCTCCATCTCCACCAGAAGCATGAGCCGCCAATTCACC
                                                          M  S  R  Q  F  T

              C  K  S  G  A  A  A  K  G  G  F  S  G  C  S  A  V  L  S  G
Normal        TGCAAGTCGGGAGCTGCTGCCAAGGGAGGCTTCAGCGGCTGCTCAGCTGTGCTCTCCGGA
HY            TGCAAGTCGGGAGCTGCTGCCAAGGGAGGCTTCAGCGGCTGCTCAGCTGTGCTCTCCGGA
              C  K  S  G  A  A  A  K  G  G  F  S  G  C  S  A  V  L  S  G

              G  S  T  S  S  Y  R  A  G  G  K  G  L  S  G  G  F  G  S  R
Normal        GGCAGCACATCCTCCTACCGGGCAGGAGGCAAAGGGCTCAGCGGGGGCTTCGGAAGTCGG
HY            GGCAGCACATCCTCCTACCGGGCAGGAGGCAAAGGGCTCAGCGGGGGCTTCGGAAGTCGG
              G  S  T  S  S  Y  R  A  G  G  K  G  L  S  G  G  F  G  S  R

              S  L  Y  N  L  G  G  V  R  S  I  S  F  N  V  A  S  G  S  G
Normal        AGCCTTTACAACCTGGGCGGCGTCCGGAGCATCTCCTTCAATGTGGCCAGCGGCAGTGGG
HY            AGCCTTTACAACCTGGGCGGCGTCCGGAGCATCTCCTTCAATGTGGCCAGCGGCAGTGGG
              S  L  Y  N  L  G  G  V  R  S  I  S  F  N  V  A  S  G  S  G

              K  S  G  G  Y  G  F  G  R  G  R  A  S  G  F  A  G  S  M  F
Normal        AAGAGTGGAGGTTATGGATTTGGCCGGGGCCGGGCCAGTGGTTTCGCCGGCAGCATGTTT
HY            AAGAGTGGAGGTTATGGATTTGGCCGGGGCCGGGCCAGTGGTTTCGCCGGCAGCATGTTT
              K  S  G  G  Y  G  F  G  R  G  R  A  S  G  F  A  G  S  M  F

              G  S  V  A  L  G  P  M  C  P  T  V  C  P  P  G  G  I  H  Q
Normal        GGCAGCGTGGCCCTGGGGCCCATGTGCCCAACTGTGTGCCCACCTGGAGGCATCCACCAG
HY            GGCAGCGTGGCCCTGGGGCCCA--------ACTGTGTGCCCACCTGGAGGCATCCACCAG
              G  S  V  A  L  G  P  N   -   -  C  V  P  T  W  R  H  P  P  G

              V  T  V  N  E  S  L  L  A  P  L  N  V  E  L  D  P  E  I  Q
Normal        GTCACTGTCAATGAGAGCCTCCTGGCCCCCCTCAACGTGGAGCTGGACCCCGAGATCCAG
HY            GTCACTGTCAATGAGAGCCTCCTGGCCCCCCTCAACGTGGAGCTGGACCCCGAGATCCAG
                H  C  Q  *

              K  V  R  A  Q  E  R  E  Q  I  K  A  L  N  N  K  F  A  S  F
Normal        AAAGTGCGTGCCCAGGAGCGGGAGCAGATCAAGGCTCTGAACAACAAGTTCGCCTCCTTC
HY            AAAGTGCGTGCCCAGGAGCGGGAGCAGATCAAGGCTCTGAACAACAAGTTCGCCTCCTTC

              I  D  K
Normal        ATCGACAAGGTGGGTCTGCTAGGACCTGGGGACCTGCACACAGCCCTTGTATTTTCTTCT
HY            ATCGACAAGGTGGGTCTGCTAGGACCTGGGGACCTGCACACAGCCCTTGTATTTTCTTCT
                      ↑
```

Figure 2.4. Predicted protein sequences of a normal and HY K71 protein. A star (*) represents a stop codon. The normal protein is 525 amino acid residues in length. The HY protein is expected to be 108 amino acid residues in length due to the premature induction of a stop codon.

>Normal K71 protein sequence

MSRQFTCKSGAAAKGGFSGCSAVLSGGSTSSYRAGGKGLSGGFGSRSLYNLGGVRSISFN
VASGSGKSGGYGFGRGRASGFAGSMFGSVALGPMCPTVCPPGGIHQVTVNESLLAPLNVE
LDPEIQKVRAQEREQIKALNNKFASFIDKVRFLEQQNQVLETKWELLQQLDLNNCKNNLE
PILEGYISNLRKQLETLSGDRVRLDSELRSVRDVVEDYKKRYEEEINRRTAAENEFVLLK
KDVDAAYANKVELQAKVDSMDQEIKFFKCLYEAEIAQIQSHISDMSVILSMDNNRDLNLD
SIIDEVRAQYEDIALKSKAEAEALYQTKFQELQLAAGRHGDDLKNTKNEISELTRLIQRI
RSEIENVKKQASNLETAIADAEQRGDNALKDARAKLDELEAALHQSKEELARMMREYQEL
MSLKLALDMEIATYRKLLESEECRMSGEFPSPVSISIISSTSGSGGYGFRPSSVSGGYVA
NSGSCISGVCSVRGGESRSRSSTTDYKDALGKGSSLSAPSKKASR

>Congenital hypotrichosis (HY) K71 protein sequence

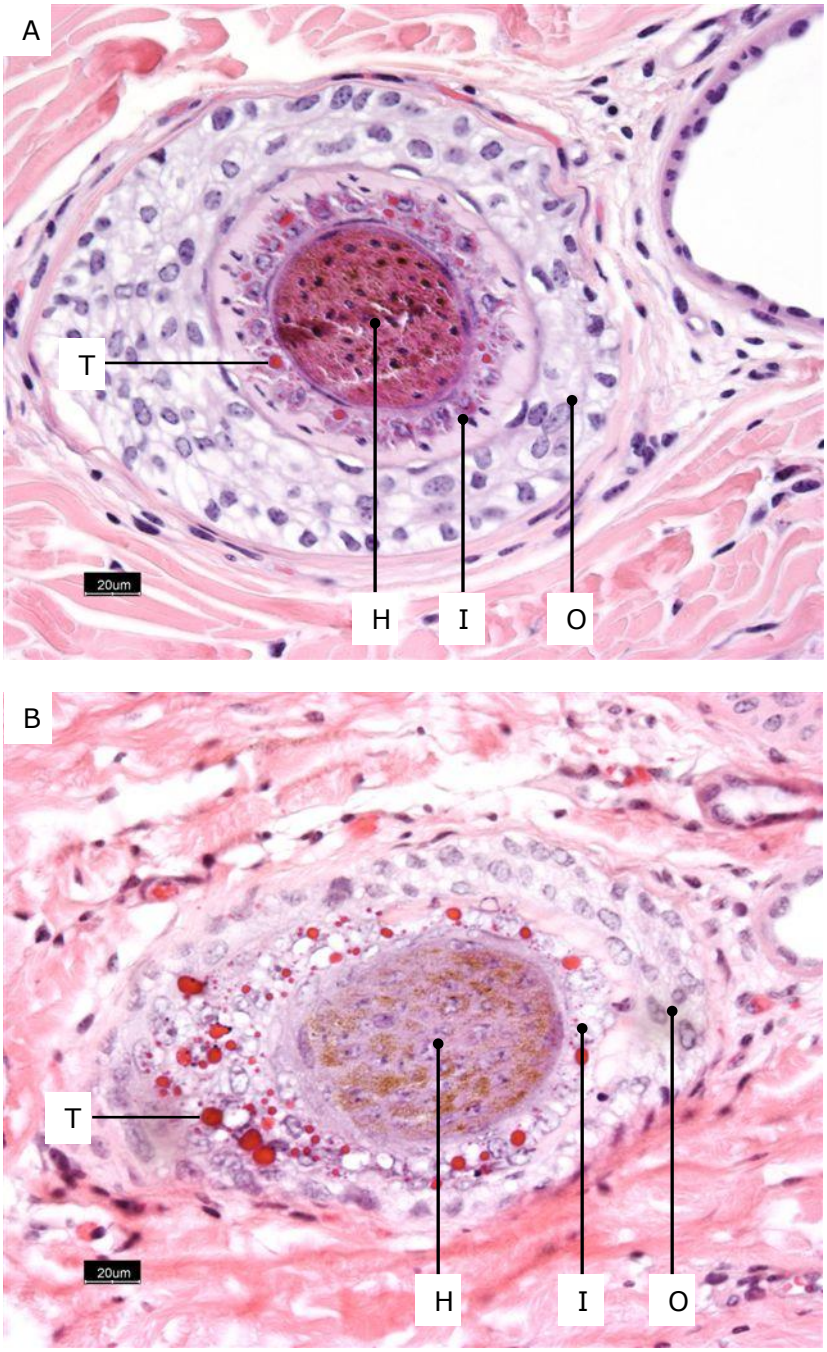
MSRQFTCKSGAAAKGGFSGCSAVLSGGSTSSYRAGGKGLSGGFGSRSLYNLGGVRSISFN
VASGSGKSGGYGFGRGRASGFAGSMFGSVALGPNCVPTWRHPPGHCQ*

Figure 2.5. The DNA sequences of a normal and HY allele as amplified by the DNA based diagnostic. A dashed line (-) represents a deleted base. Underlined bases indicate the primer sequences used for the DNA based diagnostic test.

>Normal DNA sequence
GCAGCATGTTTGGCAGCGTGGCCCTGGGGCCCATGTGCCCAACTGTGTGCCCACCTGGAG
GCATCCACCAGGTCACTGTCAATGAGAGCCTCCTGGCCCCCCTCAACGTGGAGCTGGACC
CCGAGATCCAGAAAGTGCGTGCCCAGGAGCGGGAGCAGATCAAGGCTCTGAACAACAAGT
TCGCCTCCTTCATCGACAAGGTGGGTCTGCTAGGACCTGGGGACCTGCACACAGCCCTTG
T

>Congenital hypotrichosis (HY) DNA sequence
GCAGCATGTTTGGCAGCGTGGCCCTGGGGCCCA--------ACTGTGTGCCCACCTGGAG
GCATCCACCAGGTCACTGTCAATGAGAGCCTCCTGGCCCCCCTCAACGTGGAGCTGGACC
CCGAGATCCAGAAAGTGCGTGCCCAGGAGCGGGAGCAGATCAAGGCTCTGAACAACAAGT
TCGCCTCCTTCATCGACAAGGTGGGTCTGCTAGGACCTGGGGACCTGCACACAGCCCTTG
T

Figure 2.6. Tissue section of normal (A) and Hypotrichosis (B) affected animals. Hair (H), inner root sheath (I) and outer root sheath (O) are indicated. HE stain with XX magnification. **A)** Normal animal. Note the trichohyaline granules (T) in the inner root sheath are smaller and more angular and cells surrounding the granules are intact. **B)** Affected animal. Note the trichohyaline granules (T) are larger and more globular with rounded margins and cells are vacuolated and degenerating.



Chapter 3. A genome wide association analysis for growth, carcass and meat quality traits in cattle

Abstract

Background

A whole genome scan was completed to discover quantitative trait loci that may be influencing growth, carcass and meat quality traits in a US Simmental-Angus population. Determining regions influencing quantitative traits can provide novel markers that can be integrated into marker assisted selection breeding schemes and also provide regions that can be explored with targeted re-sequencing to elucidate the genetic variation contributing to the complex quantitative phenotype.

Results

Association was detected for birth weight, back fat, yield grade, ribeye area and marbling. There were 81 SNP associations with birth weight that were distributed among 10 genomic regions. The carcass trait, back fat was significantly associated with 209 SNPs in 46 genomic regions and yield grade displayed associations with 172 SNPs in 69 genomic areas. Four SNPs within a single genomic region were associated with ribeye area. Intramuscular fat deposition or marbling was associated with 32 genomic regions containing 127 SNP markers.

Conclusion

A whole genome scan for growth, carcass and meat quality traits revealed several regions of the genome that are associated with these traits in beef cattle. These regions are ideal targets for development of marker assisted selection tools and are excellent regions to for further fine mapping studies and potentially the discovery of causal variants underlying the quantitative trait loci. The utilization of genotype imputation from the BovineSNP50 to

the high density 770K platform allowed localization of QTL to refined regions of the genome, often eliminating the need for further fine mapping efforts prior to targeted re-sequencing.

Background

Growth, carcass and quality traits that are of interest to animal producers are usually quantitative in nature and thus, influenced by many genes. Uncovering the numerous regions of the genome that influence these traits can provide an opportunity to gather new information about gene function and also lead to a better understanding of how genes interact to impact physiology. Additionally, the discovery of novel loci contributing to a trait may provide new markers that can be integrated into marker assisted selection programs with the potential to have considerable impact on genetic gain and improvement for many traits of interest to beef producers.

The ability to complete a genome wide association study (GWAS) to discover regions of the genome influencing quantitative traits has been improved by developments in bovine genomic technologies such as the BovineSNP50 and BovineHD Beadchip [5]. These platforms allow quantitative traits to be mapped efficiently and the high resolution of these platforms allows the quantitative trait loci to be localized to a finite region of the genome.

The advent of the high density single nucleotide polymorphism (SNP) platforms have provided an avenue for linkage disequilibrium mapping methods to locate genomic regions contributing to phenotypes of interest to beef producers such as calving ease [145-147], growth [56], meat quality [148] and carcass weight [149]. Several studies have investigated calving ease traits and association was detected with two chromosomes in Fleckvieh cattle [147] and in Danish and Swedish Holsteins association was detected on six chromosomes [146] and in a follow-up study with increases animal numbers association was detected on 22 chromosomes [145]. Meat quality traits have also been examined and association with Warner-Bratzler shear force was found in 79 genomic regions with eight

consensus regions across five beef breeds [148]. The association of carcass traits with regions of the genome was explored by Nishimura *et al.* [149] who utilized Japanese Black cattle to identify three genomic regions highly associated with carcass weight, confirming mapping data from previous linkage and fine mapping studies [150-152].

A genome wide association study was completed to uncover quantitative trait loci (QTL) contributing to growth, carcass and meat quality traits in a Simmental-Angus population. A total of 1672 steers were genotyped with the Illumina BovineSNP50 Beadchip (50K) genotypes were imputed in the steers using 111 sires genotyped on the Illumina BovineHD Beadchip (HD) as a reference panel. The 50K and imputed HD genotypes were utilized in the whole genome scan to map associations with growth, carcass and meat quality traits in the population.

Results and Discussion

A total of 631,665 SNPs generated from imputation were used in the association analysis. The number of animals with phenotype data varied for each trait. After quality control measures, the number of animals with both genotypes and phenotypes ranged between 1423 and 1643 for growth traits and between 1639 to 1645 animals for carcass traits (Table 3.1). There were 630 animals with both genotypes and phenotypes for Warner-Bratzler shear force and 1645 for marbling.

Association reaching the significance threshold of $p \leq 1.0 \times 10^{-3}$ was detected for five of the traits analyzed on most of the autosomes with the exception of chromosomes 2, 16, 18, 19, and 23. There were 81 associations with birth weight (BW), 209 SNP associations with back fat, four associations with ribeye area (REA), 127 associations with marbling (MS) and 172 with yield grade (Table S3, supplementary file: Markey_thesis_supplement.xlsx). The SNP markers associated with these traits were positioned into ten regions for BW, 46 regions for BF, one region for REA, 32 regions for MS and 69 with yield grade (Table 3.2).

There were no associations exceeding the significance threshold for weaning weight, yearling weight, Warner-Bratzler shear force, hot carcass weight or kidney, pelvic and heart fat. Regions of the genome with SNP associations were investigated for plausible candidate genes that may be contributing the associated phenotype.

Birth weight

Association was detected with birth weight on BTA8 with six SNPs exceeding the significance threshold from 77.86 to 78.08 Mb. Two viable candidate genes, interleukin 11 receptor, alpha (IL11RA) and AT rich interactive domain 3C (BRIGHT-like) (ARID3C), reside in this region. IL11RA is expressed in all adult tissues, during embryonic development and in differentiating embryonic stem cells in mice and the IL11RA ligand (interleukin 11, IL11) may regulate cartilage and bone function [153]. In humans, a mutation in IL11RA results in short stature, cranial malformation and delayed tooth eruption [154]. IL11RA is strongly expressed in osteogenic mesenchyme but no expression was noted in differentiated osteoblasts associated with the bone matrix in embryonic mice [154]. IL11 is also expressed in mesenchyme and is involved in hematopoiesis, blastocyst implantation, bone remodeling and differentiation of both osteoblasts and osteoclasts [154]. As both osteoblasts and osteoclasts are integral to the balance of bone resorption and formation, alterations in receptor function may influence embryonic growth. The second candidate gene, ARID3C, is a member of the ARID family of DNA binding proteins that participate in embryonic development, cell cycle control and transcriptional regulation that likely occurs through modification of chromatin structure [155]. ARID3C forms a complex with ARID3A and actions perceived to be mediated by ARID3A may in fact be mediated by the complex between ARID3C and ARID3A [156]. The overexpression of ARID3A promotes transition from senescence into the cell cycle in mouse embryonic fibroblasts and ARID3A functions in the differentiation of mature erythrocytes from hematopoietic stem cells in mouse embryonic liver during embryonic development [157]. As IL11RA has known function in

bone development and ARID3A is known to be involved in regulation of cell cycle during embryonic development both are valid candidate genes to explore for genetic variants underlying the association seen with birth weight on BTA8.

Ribeye area

Association was detected with ribeye area phenotype on BTA20 with four SNPs reaching the significance threshold spanning from 25.60 to 26.75 Mb, the most significant of which was ARS-BFGL-NGS-42157 at 25.60 Mb. There are eight annotated genes in this region and of these, two are functional candidates. The gene pelota homolog (PELO), located at 26.28 Mb, is reported to be involved in cell proliferation and has been shown to associate with actin microfilaments in mammalian cells [158]. Over expression of PELO in human Hep2g cells displayed a marked effect in cell growth and cytoskeleton organization as well as cell spreading [158]. According to UniGene data EST profiles PELO is moderately expressed in bovine muscle (Bt. 76149). Also located in this region is pre-heat shock 27kDa protein 3 (HSPB3), a gene that is known to be abundantly expressed in muscle [159]. Heat shock proteins (HSPs) are thought to be antiproteolytic proteins that have a protective effect in skeletal muscle and are induced in response to cellular stress [160]. The small HSPs have been shown to confer stress resistance in cultured mammalian cells and are believed to stabilize actin fibers and intermediate filaments [161-163], thus maintaining or protecting the cytoskeletal structures [164]. The expression of HSPB3 is specific to heart and skeletal muscle and expression is induced in the late stages of myogenic differentiation by the myogenic regulatory factor MyoD [164]. As HSPB3 likely plays a role in cell survival and muscle cell differentiation and PELO is involved in cell proliferation and associates with actin microfilaments both are viable candidates that may contribute to the association seen with REA on BTA20.

Back Fat

Association with back fat was seen on BTA13 with 11 SNPs from 69.59 to 72.10 Mb and the most significant of these was BTA-33683-no-rs 72.10 Mb. Within this region there are seven annotated genes on the bovine assemble and of these, two are plausible candidate genes, phospholipase C, gamma 1 (PLCG1) and lipin 3 (LPIN3). PLCG1 is activated in response to leptin signaling [165] through phosphorylation by protein tyrosine kinases [166] and activity of PLCG1 is increased by the presence of lipids. PLCG1 has been implicated in metabolic syndrome displaying dyslipidemia, elevated plasma glucose and obesity in humans [167] and PLCG1 is also a member of the vasoactive intestinal peptide pathway that has been associated with fat mass and body mass index in humans [168]. LPIN3 belongs to the group of phosphatidate phosphatases that are integral in the biosynthesis of triacylglycerol that can then be converted into very-low-density lipoproteins that are available for uptake by peripheral tissues [169]. Lipin 3 interacts with lipin 1 to form a heteroligomer that acts as an integrated unit [168]. Lipin 1 is expressed in adipose tissue and is necessary for development of normal adipose tissue and mice that lack functional Lipin 1 exhibit adipose tissue deficiency, glucose intolerance, insulin resistance and reduced expression of lipoprotein lipase in white adipose tissue [170].

The region associated with BF on BTA20 consisted of 5 SNPs reaching the significance threshold, spanning from 32.07 to 32.36 Mb with the most significant association being with ARS-BFGL-NGS-97963 at 32.07 Mb. There are seven annotated genes in this region on the bovine assembly. Growth hormone receptor (GHR) spans this interval and is known to play a substantial role in modulating the biological actions of growth hormone (GH). Polymorphisms within GHR gene have been associated with milk yield and composition in dairy cattle [171] and lean muscle growth and fat deposition in beef cattle [172]. Classically, the role of GH in growth is described as GH released from the anterior pituitary stimulates the production and release of insulin-like growth factor 1 (IGF-1) from the liver and other tissues and IGF-1 then stimulates growth and differentiation

through tissue specific receptors. While GHR expression is most prevalent in the liver it is also expressed in numerous other tissues such as adipose, skeletal muscle, heart and testis. Tissue specific roles for GHR have been suggested with GHR in liver being involved in growth while GHR in adipose is involved in metabolism [173]. Increases in circulating GH result in decreased lipogenesis and increased lipolyisis thus, making free fatty acids available for energy utilization. The regulation of growth and metabolism through GHR is complex as many hormones and transcription factors control the expression of GHR such as GH, insulin, IGF, somatostatin, cortisol and thyroid hormones [173]. As GHRs are thought to have tissue specific roles, changes in ligand binding efficiency or changes in expression of GHR may have a substantial impact on mobilization or conservation of lipids in adipose tissue.

Association with back fat was detected on BTA4 with 25 SNPs spanning from 56.41 to 66.16 Mb. Three overlapping regions were also associated with yield grade: 12 SNPs from 60.22 to 60.25 Mb, one SNP at 63.80 Mb and one SNP at 66.16 Mb. There were 42 annotated genes and 22 uncharacterized loci in the region and of these adenylate cyclase activating polypeptide 1 receptor 1 (ADCYAP1R1), Kelch repeat and BTB (POZ) domain containing 2 (KBTDB2) and growth hormone releasing hormone receptor (GHRHR) were plausible candidates that may be contributing to the association seen in the region. ADCYAP1R1 is thought to play a role in energy and lipid metabolism and regulation of body weight [174]. The ADCYAP1R1 ligand, adenylate cyclase activating polypeptide 1 (ADCYAP1), has been associated with lipid and carbohydrate metabolism [175] and has been shown to stimulate insulin release from B-cells of pancreas of mice and rats [176]. Additionally, mice that are null for ADCYAP1 exhibit reduced body mass and adipose tissue [177]. ADCYAP1R1 null mice exhibit impaired insulin secretion [178, 179], reduced food intake, lower body weight and reduced white adipose tissue [180, 181]. The second candidate in the region, KBTDB2, is expressed in bovine adipose and pancreas (UniGene, Bt.28701) and has been implicated in the development of adipose tissue and fat

accumulation in cattle, pigs and mice [182]. Growth hormone releasing hormone receptor (GHRHR) is widely expressed in the anterior pituitary gland and interacts with growth hormone releasing hormone to provoke the release of growth hormone from the anterior pituitary gland. Growth hormone acts directly on adipose and muscle to regulate lipolysis and protein synthesis but also regulates growth indirectly through release of IGF-1 from the liver. Mutations within GHRHR can impair ligand binding and autosomal recessive forms of familial growth hormone deficiency resulting have been associated with mutations in GHRHR [183]. GHRHR mutations can also influence body composition and fat deposition in humans. Individuals heterozygous for a null GHRHR exhibited normal adult stature but displayed changes in body composition including reduced BMI and decreased lean tissue in addition to increased insulin sensitivity [184]. While ADCYAP1R1, KBTDB2 and GHRHR are viable candidate genes for the association seen on BTA4 due to their involvement with metabolism and adipose deposition the associated region is quite large, encompassing almost 10 Mb, thus further analysis will be necessary to effectively localize the QTL.

Marbling

Association was detected with marbling on BTA9 with 3 SNPs at 35.00 Mb the most significant of which was BovineHD0900009588 at 34.99 Mb. There were ten annotated genes in this region on the bovine genome assembly including two valid candidate genes fyn-related kinase (FRK) and 5'-nucleotidase domain containing 1 (NT5DC1). The gene FRK has been implicated in serum total cholesterol and low-density lipoprotein cholesterol in humans [185] through meta-analysis of GWAS. The lead SNP from the GWAS displayed cis acting association with transcript levels of FRK in both the liver and subcutaneous fat. Association was seen with transcripts of both FRK and NT5DC1 in human omental fat although little is known about the function of NT5DC1 [185], a knockout mouse model displayed a significant decrease in plasma T3 levels and subtle differences in expression of genes regulated by thyroid hormones [186, 187].

A single SNP, BovineHD1000005477, exceeding the threshold was associated with marbling on BTA10 at 16.42 Mb. There were six annotated genes in this region on the bovine assembly. The gene glucuronic acid epimerase (GLCE) is located in this region and has been associated with plasma triglyceride and high-density lipoprotein cholesterol levels in humans [188]. GLCE is a key enzyme in the biosynthesis of heparin sulfate proteogylcans (HSPG) that are expressed in endothelial cells of the liver and mediate the clearance of lipoprotein remnants by interacting with both lipases and lipoproteins. Triglyceride rich lipoproteins from the diet are acted upon by lipoprotein lipase to generate free fatty acids for energy in skeletal muscle or storage in adipose tissue. Once the triglyceride has been metabolized, the cholesterol rich remnant vesicle is sequestered by HSPG at the liver. The HSPGs also interact with apoB, apoE, LPL and hepatic lipase thus, allowing the interaction of these enzymes and the lipoprotein remnants in order to facilitate the clearance of the remnants from circulation. When HSPG and low density lipoprotein receptor pathway is impaired, circulating cholesterol and triglyceride increased more than 10 fold in mice [189] and high levels of circulating fatty acids may influence insulin signaling [190] and thus modulate lipogenesis or lipolysis in tissues.

The marbling phenotype was associated with two loci on BTA11, the first of which consisted of 11 SNPs spanning from 32.89 to 34.79 Mb and the most significant association occurred with BovineHD1100009915 at 32.89 Mb. There only annotated gene in the region was neurexin 1 (NRXN1). NRXN1 is highly expressed at the β-cell membrane of the pancreas in humans, rats and mice and interacts with components of the insulin secretory granule docking machinery at the β-cell membrane. Insulin release is bi-phasic with one pool of insulin sequestered in granules at the β-cell plasma membrane that can be rapidly released while the secondary pool held in reserve requiring transport to the membrane before secretion can occur [191]. In rats, increases in circulating glucose resulted in reduced expression of NRXN1 and increased insulin secretion, while targeted silencing of

NRXN1 also resulted in increased insulin secretion indicating the participation of NRXN1 in insulin secretion mechanisms [192].

The second locus associate with marbling on BTA11 was with 18 SNPs extending from 80.25 to 81.50 Mb the most significant of these was with BovineHD1100022978 at 80.25 Mb. There were nine annotated genes in this region of the bovine assembly including visinin-like 1 (VSNL1). VSNL1 is expressed in murine pancreatic islets and β-cells and is thought to mediate insulin secretion in the presence of glucose. In mice, over-expression of VSNL1 results in elevated cAMP levels, increased insulin secretion in the presence of glucose, and increased insulin gene expression in pancreatic β-cells [193]. As the cAMP pathway is known to enhance insulin gene transcription in the presence of glucose [194] it is likely that the VSNL1 stimulated cAMP accumulation resulted in elevated expression of insulin [193].

There were two regions were associated with marbling on BTA21. The first associated region contained 7 SNPs ranging from 21.14 to 21.22 Mb, the most significant of which were BovineHD2100006234 and BovineHD2100006237 located at 21.19 and 21.22 Mb respectively. There were 15 annotated genes within this region on the bovine genome assembly. Perilipin (PLIN1), a modulator of adipoctye metabolism, is localized on the surface of triacylglycerol containing lipid droplets in adipose tissue. Perilipin controls access of lipases to lipid droplets and upon phosphorylation by cAMP dependent protein kinase A, perilipin undergoes a conformational change, thus allowing lipases access to lipid droplets in order to liberate free fatty acids that will serve as fuel for tissues of the body [195, 196]. Mice that are null for PLIN1 exhibit increased lean body mass and alterations in glucose homeostasis, adipoctye lipolysis and leptin production and eventually develop insulin resistance [196, 197]. Mutations in PLIN1 in humans have been associated with altered lipolyisis and free fatty acid concentrations [198, 199] as well as risk for insulin resistance [200].

The second locus associated with marbling on BTA21 was comprised of 2 SNPs reaching the significance threshold at 59.67 Mb. The gene serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 12 (SERPINA12; also known as vaspin) resides in this area and elevated concentrations of this compound have been shown to be expressed in visceral and subcutaneous adipose tissue in humans and SERPINA12 has also been correlated with obesity and impaired insulin sensitivity [201]. SERPINA12 has been isolated from the visceral adipose tissue of Otsuka Long-Evans Tokushima Fatty (OLETF) rats, a model for type 2 diabetes, and it is suggested that the up-regulation of SERPINA12 may protect against insulin resistance.

Combined association with birth weight, back fat and marbling

Association was detected between a single locus on BTA6 and the phenotypes birth weight, marbling and back fat. There were associations with 64 SNPs for BW spanning from 36.99 to 42.06 Mb, 40 associations with BF reaching from 36.99 to 41.19 Mb and 8 associations with marbling spanning from 39.36 to 39.45 Mb. The largest association with birth weight was with BovineHD0600010810 located at 39.26 Mb and the most significant association with BF was with BovineHD0600010843 at 39.45 Mb. Three SNP associations with marbling were equally significant those being association with BovineHD0600010838 at 39.43 Mb, and BovineHD0600010841 and BovineHD0600010842 both of which were located at 39.44 Mb. This region has been associated with many QTL in beef cattle and there are several positional and functional candidate genes within this interval, including ATP-binding cassette, sub-family G (WHITE), member 2 (ABCG2), secreted phosphoprotein 1 (SPP1), non-SMC condensin I complex, subunit G (NCAPG) and ligand dependent nuclear receptor corepressor-like (LCORL) in addition to 16 other annotated genes in the bovine assembly. ABCG2 is located at 37.96 Mb and is expressed in bovine muscle, adipose, mammary and fetal tissues (UniGene, Bt.51973). ABCG2 has been implicated in dairy cattle as influencing milk yield and composition [202, 203] but the putative causative allele from dairy cattle was

not segregating in a beef breed diversity panel [204], thus this gene is unlikely to contribute to the association with phenotypes in this region. SPP1 (38.12 Mb) encodes the osteopontin protein that is a glycoprotein involved in cell signaling through binding with integrin, as well as bone growth and remodeling [205]. SPP1 is expressed in numerous bovine tissues including adipose, muscle, cartilage and fetal tissues (UniGene Bt.10536) and has been associated with growth and carcass traits in beef cattle [172]. Additionally, SPP1 has been implicated as containing the causative variant underlying a milk protein and milk fat QTL in Holsteins [205, 206]. NCAPG (38.77 Mb) is expressed in pancreas, liver, skin, intestine (UniGene, Bt.100379) as well as bovine fetal tissues such as skeletal muscle, kidney, liver, bone, fetal placenta and maternal placenta [207]. In horses, NCAPG has been implicated in height [208] and in cattle, as impacting prenatal growth [207], body frame size at puberty [209], carcass weight, ribeye area and subcutaneous fat deposition [152]. The fourth functional candidate gene in the region, LCORL, (38.84 Mb) has been associated with human height, skeletal size and height growth in infancy [210-212] as well as horse height [208] and horse body size [213]. LCORL is expressed in bovine fetal tissue, liver and intestine (UniGene, Bt.38533) but expression in other tissues has not been assessed in cattle. As NCAPG and LCORL are located in close proximity, they have been implicated jointly in horse height at withers [208] as well as fat thickness, hot carcass weight and ribeye area in cattle [214]. As this region on BTA6 is gene rich there are several positional and functional candidate genes that may contribute to the association with birth weight, marbling and back fat.

There was no association detected near some of the known growth influencing genes such as insulin-like growth factor 1 (IGF1), myostatin (MSTN) or pleiomorphic adenoma gene 1 (PLAG1). Association was also not observed near genes implicated as influencing meat quality such as thryoglobulin (TG), calpastatin (CAST), calpain 1, (mu/I) large subunit (CAPN1), melanocortin 4 receptor (MC4R) or diacylglycerol O-acyltransferase 1 (DGAT1).

Conclusion

A whole genome scan for growth, carcass and meat quality traits revealed several regions of the genome that are associated with birth weight, back fat, ribeye area, yield grade and marbling in beef cattle. These regions are ideal targets for development of markers to use in marker assisted selection in Angus-Simmental cattle as these markers are already segregating in these populations. The regions identified are also excellent targets to investigate further in an attempt to discover causal variants contributing to the association seen with the phenotypes.

The use of imputation from 50K to HD genotypes may prove useful in determining finite regions for targeted re-sequencing to discover causal variants underlying these phenotypes. In many instances the density of the imputed genotypes allowed localization of the QTL to a precise region containing just a handful of genes with only one or two plausible functional candidate genes. However, there were also examples where further refinement and analysis will be necessary to thoroughly interrogate the associated locus.

Many of the associations detected for the adiposity traits, marbling and back fat, were near genes that may have known or implicated function in metabolism and energy partitioning through modulation of insulin or growth hormone pathways. As insulin is known to increase lipase activity and fat storage it is not surprising that association with adipose related traits has been detected near genes that mediate insulin expression and secretion. Several of the associations were also detected near genes implicated in human obesity, diabetic and dyslipidemia conditions and elucidation of genetic variants underlying these phenotypes may provide further insight into the complex gene interactions that contribute to the development and progression of these human diseases.

Methods

Steers of a Simmental, Angus or Simmental x Angus crossbred background were obtained from five ranches in Montana over a five year period. Dams were randomly assigned to sire. Birth weight and weaning weight were recorded on each ranch. Animal management, dietary treatments and data collection methods are reviewed in Trejo 2010 (Thesis). Traits evaluated are categorized as growth, meat quality or carcass traits. Growth traits were birth weight (BW), weaning weight (WW), yearling weight (YW). Meat quality traits were marbling score (MS) and Warner-Bratzler shear force (SF). Carcass traits were hot carcass weight (HCW), ribey area (REA), back fat (BF), kidney, pelvic and heart fat (KPH) and yield grade (YG). Multiple blood samples were collected on each individual by jugular puncture. DNA was isolated using the simple salting-out method [140] and samples frozen at -20 C. Steers were parentally validated to their respective sires using a panel of 15 microsatellite markers consisting of 10 of the approved ISAG bovine markers and five additional markers. There were a total of ~1800 parentally verified steers from 133 sires over the five year period.

Animals were selected for genotyping based on divergent adjusted residual phenotypes which were attained using the PROC MIXED option of the SAS® system (Version 9.2, SAS Inst., Inc., Cary, NC). A stepwise reduction method with backward selection was used and variables significant at $p \leq 0.05$ were retained in the model. The full model for carcass and meat quality traits included year, ranch, days on feed (DOF), treatment, days of age (DOA), HCW, sire breed, dam breed and the interaction between sire and dam breeds. The full model for growth traits included year, ranch, birth month, birth weight, DOA, DOF, sire breed, dam breed and the interaction between sire and dam breeds. The upper and lower 20 animals were selected for each trait for the preliminary analysis. An additional 97 and 50 genotypes were available from years two and four, respectively, resulting in a total of 742 animals in the preliminary analysis. Additional animals were

selected for genotyping based on completeness of phenotypic data resulting in a total of 1672 genotyped offspring representing 127 sires. Those sires with six or more genotyped offspring were selected for genotyping on the BovineHD BeadChip resulting in a total of 111 sires.

DNA was sent to Geneseek (Lincoln, NE) for genotyping on the Illumina® BovSNP50 BeadChip and the BovineHD BeadChip platforms. Quality control measures were completed using GoldenHelix SNP & Variation Suite v 7.6.11 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). Individuals or markers with more than 10% missing genotypes and markers with minor allele frequency less than 1% were excluded from the analysis. After quality control measures there were 46,334 and 646,510 markers from the BovSNP50 and the BovineHD, respectively, and 1,644 steers and 110 sires. There were 41,728 markers from the BovSNP50 that were common between the SNP50 and BovineHD that were used during imputation. A hidden markov model procedure as implemented in Beagle (version 3.3.2) was used for imputation. The default parameters of BEAGLE v 3.3.2 [77] were used with 10 phasing iterations and four haplotype pairs were sampled for each individual during each iteration of phasing. The sire HD genotypes were phased and then used as a reference panel [75] to impute genotypes in the steers. The r2 score generated by BEAGLE was used as a threshold for imputation quality and only markers exceeding an r2 of 0.5 were used for the whole genome scan resulting in 631,665 markers. In order to assess the accuracy of imputation a total of 52 steers were genotyped on the BovineHD assay. Concordance was assessed using Plink v 1.06 [110] and at a Beagle quality control r2 of 0.5 concordance was found to be 93.2%.

After quality control measures, the number of animals with both genotypes and phenotypes ranged between 1412 and 1417 for growth traits and from 1632 to 1640 for carcass traits (Table 3.1). A total of 672 animals had both genotypes and phenotypes for Warner-Bratzler shear force and 1638 for marbling.

A genome wide association scan using the imputed genotypes and the residual phenotypes for each trait was conducted using linear regression with an additive model in GoldenHelix SVS v 7.6.11. Results were reported for associations with SNPs with a Bonferroni corrected $p \leq 1.0 \times 10^{-3}$.

Cattle and humans have been found to be similar in regards to LD and haplotype structure for distances of 1 to 100 Kb [215]. Investigation of known disease risk variants in human populations has shown that an association of SNPs with the risk variant does not exceed 500 Kb under the largest distance scenario [216]. An evaluation of LD in this population showed that LD lessened considerably after 500 Kb. Thus, 500 Kb flanking either the most significant SNP of a region or of a single significant SNP will be scanned for probable candidate genes based on known function and tissue expression.

Abbreviations

BW: birth weight, WW: weaning weight, YW: yearling weight, MS: marbling score, SF: Warner-Bratzler shear force, HCW: hot carcass weight, LMA: longissimus muscle area, BF: back fat, KPH: kidney, pelvic and heart fat, DP: dressing percentage

Additional materials

Supplementary file: Markey_thesis_supplement.xlsx

Table S3. Single SNP associations with traits and genomic regions

Associations reported with Bonferroni corrected $p \leq 1.0 \times 10^{-3}$.

Figures

Figure 3.1. Manhattan plots of single nucleotide polymorphism Bonferonni corrected associations. Red line corresponds to p = 1.0 x 10^{-3} that was used as a significance threshold. The X axis displays the bovine chromosomes and the Y axis corresponds to the negative log 10 of the Bonferroni corrected p-value.

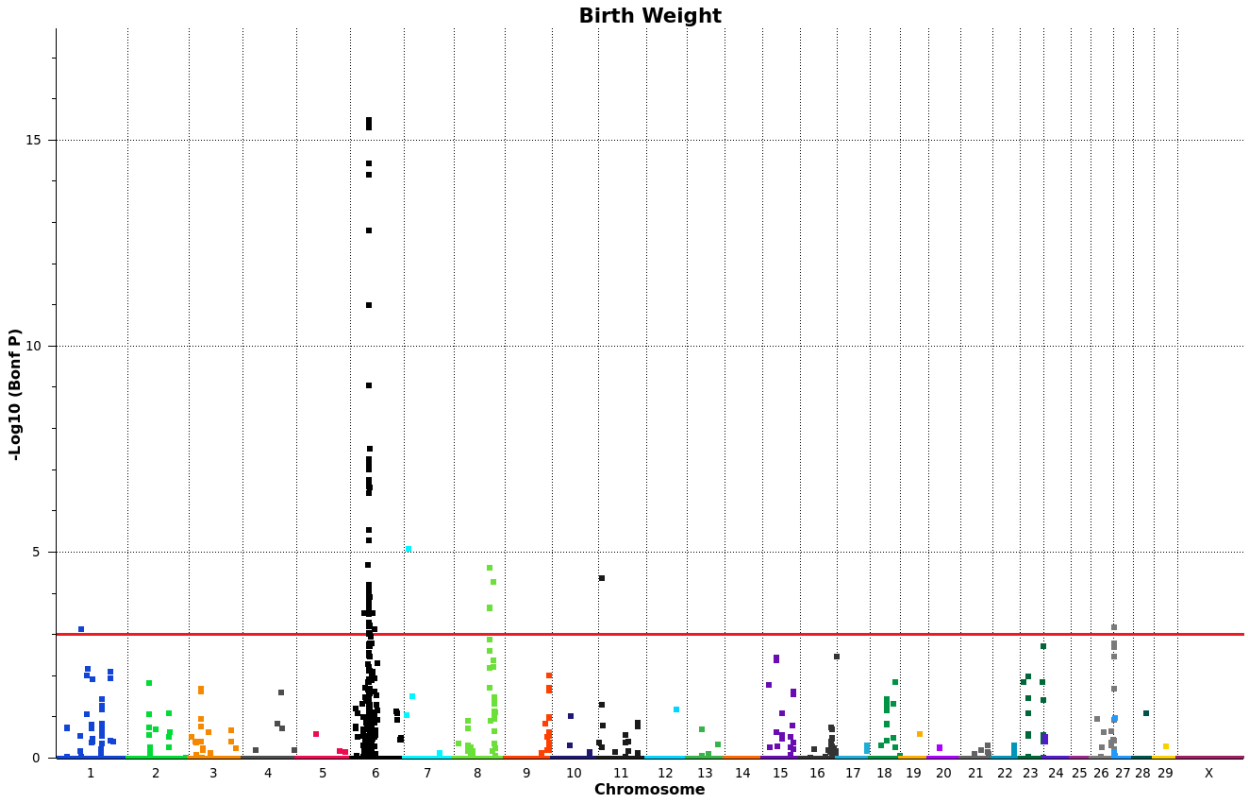
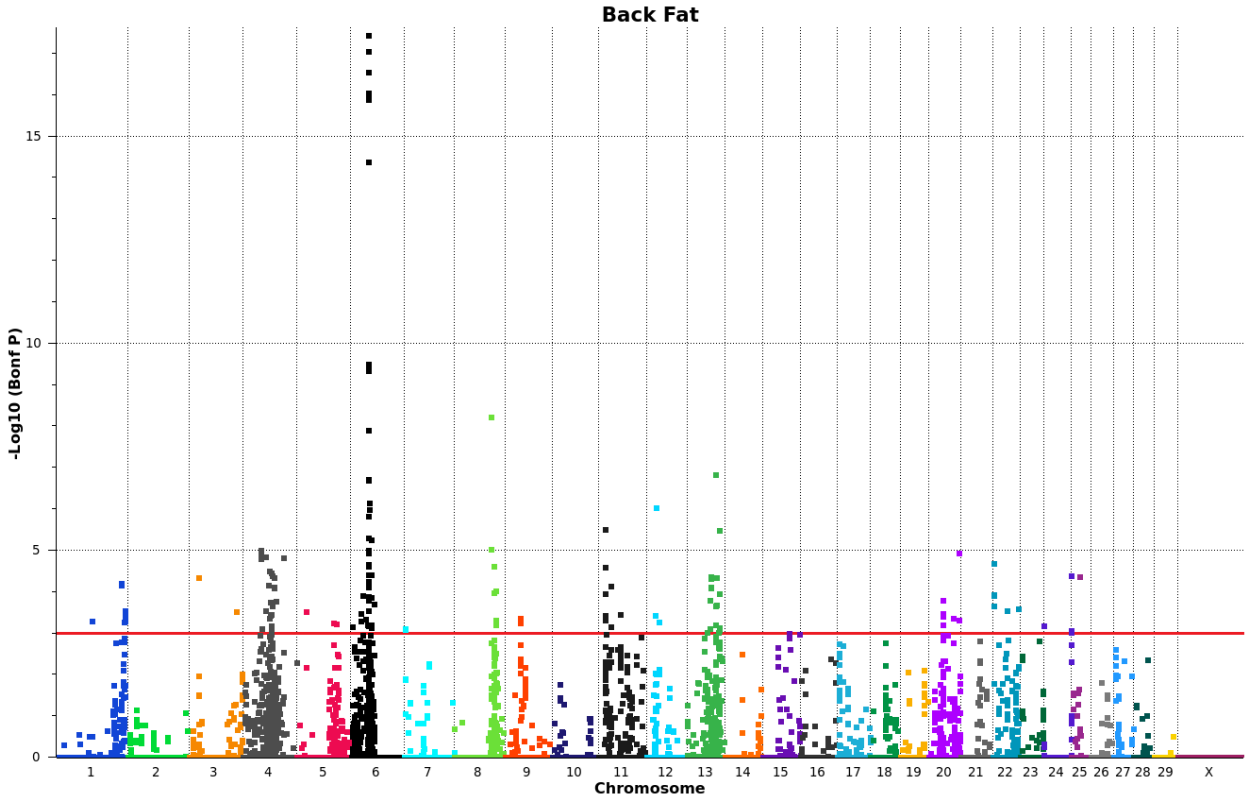


Figure 3.1. Continued



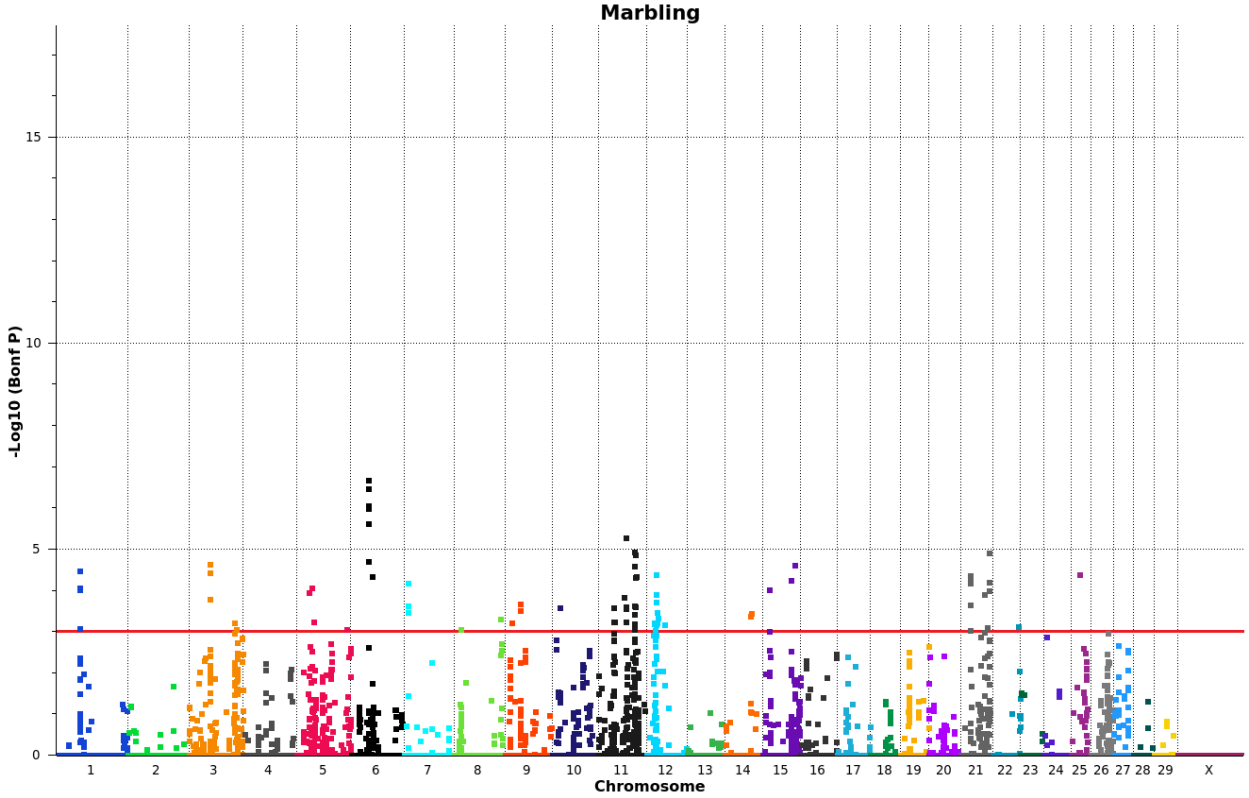
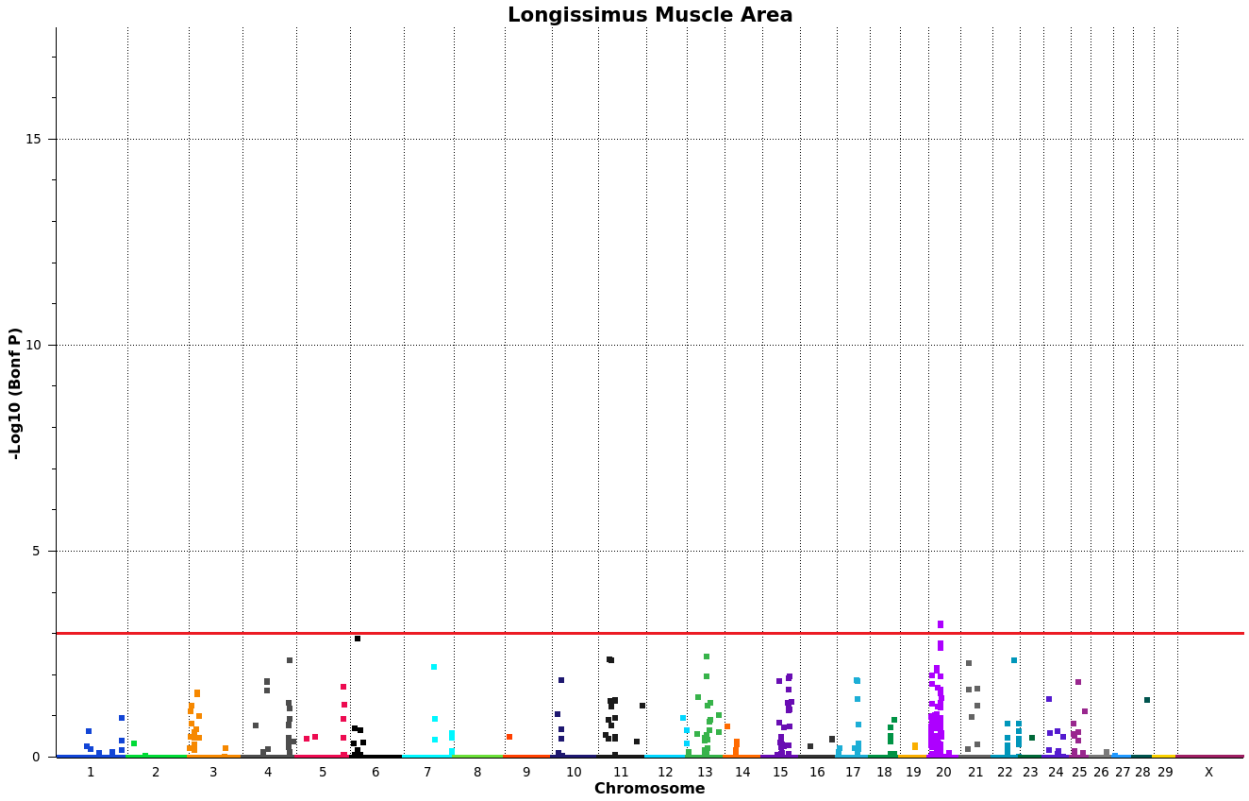
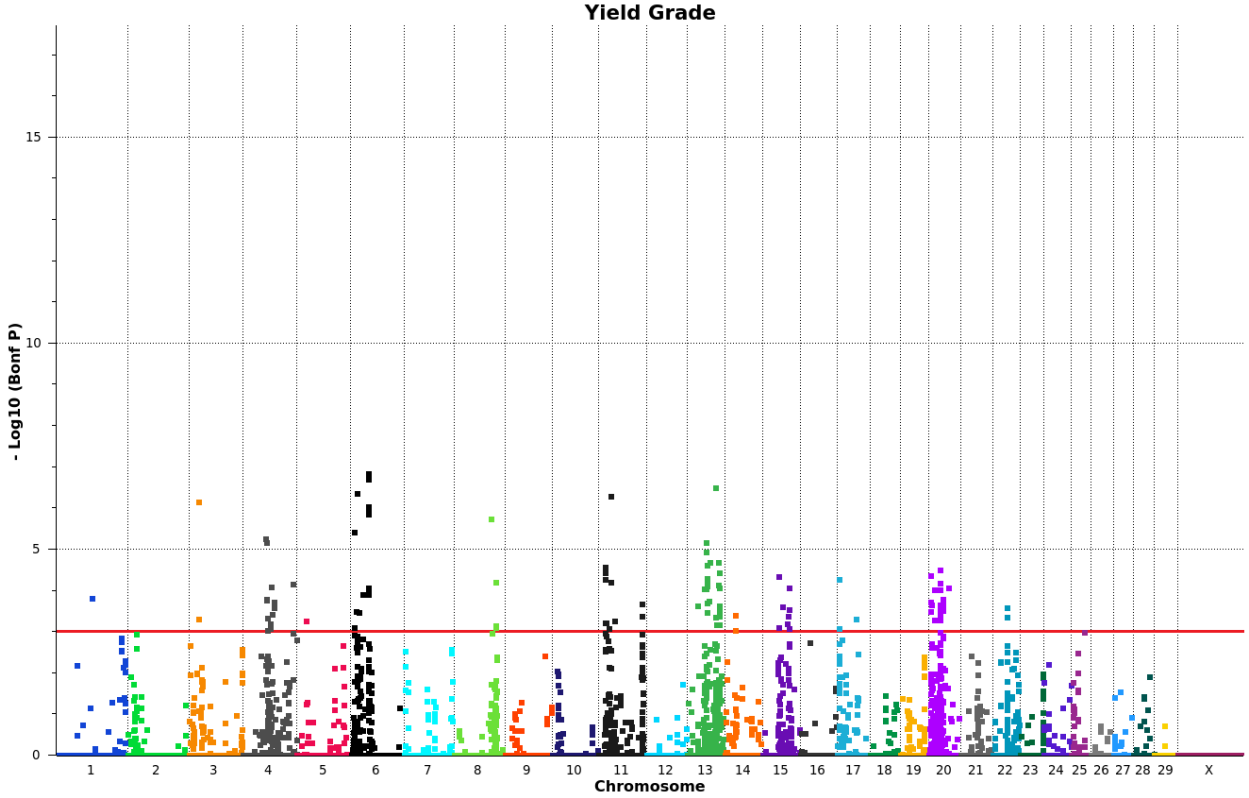


Figure 3.1. Continued





Tables

Table 3.1. Descriptive statistics of animals with both genotype and phenotype data for each trait.

Variable	N	Mean	StdDev	Min	Max
BW	1423	91.9677	12.1947	55	138
WW	1642	553.8624	84.9105	320	885
YW	1643	1216.1800	111.6795	849	1580
BF	1645	0.4795	0.1476	0.084	1.3
HCW	1645	825.4517	80.7096	492	1047
KPH	1645	2.3067	0.3805	1	3.5
REA	1639	13.7229	1.5537	8.72	20.44
YG	1645	2.9073	0.6782	0.56	6.36
MS	1645	532.6523	102.5113	240	890
SF	630	6.9957	2.1975	3.1548	14.5967

Table 3.2. Selected single SNP associations with traits and genomic regions. Associations reported with Bonferroni corrected $p \leq 1.0 \times 10^{-3}$.

1- Back fat (BF), birth weight (BW), ribeye area (REA), marbling score (MS)
2- Starting and ending base pair positions of regions or starting position of single SNPs that were associated with trait. Positions assigned based on UMD3.1 genomic assembly.
3- Number of SNPs in the region
4- Bonferroni adjusted p-value. Alpha of 0.05 adjusted for 631,665 tests.

Trait[1]	Chr	Start[2]	End[2]	No.SNP[3]	Bonf P[4]
BF	1	79241473	n/a	1	0.000511
BF	1	142804212	142904469	2	6.31E-05
BF	1	149345154	149352473	5	0.000544
BF	1	151299584	151345033	3	0.000296
BF	3	22643397	n/a	1	4.55E-05
BF	3	107397431	107405234	3	0.000311
BF	4	39748902	42939320	6	1.00E-05
BF	4	50732670	50798071	2	1.46E-05
BF	4	56409650	66161260	25	3.19E-05
BF	4	69569494	69569494	3	4.59E-05
BF	4	74536065	74536993	2	0.000171
BF	4	90647249	n/a	1	1.56E-05
BF	5	21665418	n/a	1	0.000312
BF	5	82482567	87890685	2	0.000575
BF	6	3469170	n/a	1	0.000720
BF	6	21402101	21467193	3	0.000346
BF	6	26205206	n/a	1	0.000126
BF	6	32331126	n/a	1	0.000477
BF	6	36994858	41190295	40	3.61E-18
BF	6	45340181	45948116	6	5.67E-06
BF	6	51890653	51892795	3	0.000206
BF	7	1692558	1698531	5	0.000780
BF	8	82776953	82882610	2	6.23E-09
BF	8	88068371	93445702	6	2.46E-05
BF	9	35032307	35040539	3	0.000439
BF	11	13589385	n/a	1	2.65E-05
BF	11	15082625	15165236	8	3.11E-06
BF	11	27174743	27512855	3	7.25E-05
BF	11	48210157	48232545	4	0.000365
BF	12	18909491	20487657	6	9.26E-07
BF	12	27211274	n/a	1	0.000534
BF	13	44473013	44483786	3	0.000962
BF	13	49869007	50142942	2	0.000162
BF	13	53753506	53855588	5	4.44E-05
BF	13	63085508	65978290	10	1.48E-07
BF	13	69593010	72098324	11	3.33E-06
BF	20	32074342	32359247	5	0.000161
BF	20	56351713	n/a	1	0.000444
BF	20	68958561	68965151	4	1.16E-05
BF	22	3658334	3750886	5	2.10E-05
BF	22	32775364	n/a	1	0.000288

Table 3.2. Continued

Trait[1]	Chr	Start[2]	End[2]	No.SNP[3]	Bonf P[4]
BF	22	56853113	56853848	2	0.000268
BF	24	632212	632760	2	0.000657
BF	24	61885106	61908842	6	4.09E-05
BF	25	18570338	18583815	2	4.50E-05
BW	1	52238554	52252862	5	0.000708
BW	6	28592769	n/a	1	0.000290
BW	6	36994858	42057261	58	3.14E-16
BW	6	48608641	48609889	2	0.000292
BW	6	51890653	51892795	3	0.000727
BW	7	8616879	n/a	1	8.00E-06
BW	8	77864103	78082969	6	2.29E-05
BW	8	86360973	86426013	2	5.17E-05
BW	11	5439764	5449469	2	4.21E-05
BW	26	51509776	n/a	1	0.000643
REA	20	25597446	26751470	4	0.000551
MS	1	51021507	51042849	6	3.41E-05
MS	3	47652200	48009232	4	2.38E-05
MS	3	102691519	n/a	1	0.000624
MS	3	107265306	n/a	1	0.000887
MS	5	26533028	36964780	4	8.65E-05
MS	5	111828522	n/a	1	0.000902
MS	6	39356766	39445731	8	2.16E-07
MS	6	46837171	n/a	1	4.57E-05
MS	7	8016739	8035883	6	6.71E-05
MS	8	14652026	14663575	2	0.000910
MS	8	103664822	n/a	1	0.000489
MS	9	14958833	n/a	1	0.000617
MS	9	34992028	35002290	3	0.000217
MS	10	16417578	n/a	1	0.000266
MS	11	32888659	34787108	11	0.000261
MS	11	56735630	60711686	7	5.44E-06
MS	11	80251088	81498504	18	1.16E-05
MS	11	84984134	84989388	4	4.53E-05
MS	12	17917095	17937018	2	0.000608
MS	12	19621759	21215798	10	4.14E-05
MS	12	24487401	24822715	3	0.000353
MS	12	41257544	n/a	1	0.000702
MS	14	58602358	58672380	2	0.000357
MS	15	16139861	16194101	2	9.68E-05
MS	15	63756692	n/a	1	5.80E-05
MS	15	71314601	n/a	1	2.45E-05
MS	21	21140744	21216547	7	4.45E-05
MS	21	53170535	53172678	3	0.000125
MS	21	59668657	59669168	2	0.000796
MS	21	64551867	64631179	11	1.27E-05
MS	22	58439832	n/a	1	0.000773
MS	25	18493692	n/a	1	4.10E-05

Chapter 4. Evaluation of a quantitative trait locus on BTA6 influencing growth, carcass and meat quality traits in beef cattle.

Abstract

Background

Numerous whole genome scans have been conducted in the search for regions of the genome that are contributing to growth, carcass and meat quality traits in cattle. The bovine autosome 6 has been repeatedly associated with traits such as birth weight and calving ease, back fat, ribeye area and carcass weight. Previous research conducted with a population of Simmental-Angus cattle revealed a locus on BTA6 from 37 to 42 Mb that was associated with back fat, marbling and birth weight.

Results

The region on BTA6 was evaluated in an attempt to further refine the candidate locus interval and explore genetic variants that may be contributing to the association with back fat, marbling and birth weight. The haplotype analysis revealed a region near 38.83 Mb that was associated with both back fat and birth weight. A second region near 39.27 Mb was associated with back fat, marbling and birth weight. The first region, near 38.83 Mb, corresponds to the LCORL-NCAPG locus that has been implicated in body size in many species as well as contributing to growth and carcass phenotypes in cattle. Re-sequencing of the LCORL-NCAPG locus in animals of know haplotype status revealed several polymorphisms that were consistent with haplotype status.

Conclusion

While several polymorphisms were discovered that were congruent with known haplotype status none of these appear to be the causal variant underlying the association

detected in the region. This research contributes to the mounting evidence implicating this region as influencing growth, carcass and meat quality traits in cattle.

Background

Bovine autosome 6 (BTA6) has been associated with many quantitative phenotypes in cattle including milk production, growth and carcass traits. Quantitative trait loci (QTL) for milk yield, milk protein and fat yield have been discovered on BTA6 in dairy cattle [202, 203, 205, 206, 217-219] and QTL contributing to birth weight and calving ease have been mapped to BTA6 in both dairy and beef cattle [22, 49, 220-224]. QTL influencing carcass traits such as back fat, ribeye area, proportion of lean and carcass weight have also been mapped to BTA6 [49, 152, 222, 225]. There appear to be many loci on BTA6 that contribute to various quantitative phenotypes in cattle.

Recently, we discovered a quantitative trait locus on BTA6 contributing to meat quality and growth traits in a Simmental-Angus population. The association with back fat, marbling and birth weight localized to a locus spanning from approximately 37 to 42 Mb on BTA6. There are several positional and functional candidate genes within this interval on BTA6 including ATP-binding cassette, sub-family G (WHITE), member 2 (ABCG2), secreted phosphoprotein 1 (SPP1), non-SMC condensin I complex, subunit G (NCAPG) and ligand dependent nuclear receptor corepressor-like (LCORL) in addition to 16 other annotated genes in the bovine University of Maryland assembly [2]. ABCG2 is located at 37.96 Mb is expressed in bovine muscle, adipose, mammary and fetal tissues (UniGene, Bt.51973). ABCG2 has been implicated in dairy cattle as influencing milk yield and composition [202, 203] but the putative causative allele from dairy cattle was not segregating in a beef breed diversity panel [204], thus this gene is unlikely to contribute to the phenotypes associated with this region. SPP1 (38.12 Mb) encodes the osteopontin protein that is a glycoprotein involved in cell signaling through binding with integrin, as well as bone growth and

remodeling [205]. SPP1 is expressed in numerous bovine tissues including adipose, muscle, cartilage and fetal tissues (UniGene, Bt.10536) and has been associated with growth and carcass traits in beef cattle [172]. Additionally, SPP1 has been implicated as containing the causative variant underlying a milk protein and milk fat QTL in Holsteins [205, 206]. NCAPG, is a component of the condensin complex that is responsible for condensation and segregation of chromosomes during mitosis [226]. NCAPG is expressed in pancreas, liver, skin, intestine (UniGene, Bt.100379) as well as bovine fetal tissues such as skeletal muscle, kidney, liver, bone, fetal placenta and maternal placenta [207]. The ligand dependent nuclear receptor corepressor-like (LCORL) gene encodes a transcription factor that is known to interact with ubiquitin C [227], a polyubitquitin encoding gene that participates in ubiquitination events that play a role in various cellular processes such as cell cycle regulation, protein degradation and cell signaling [228]. LCORL also interacts with C-terminal binding protein 1 (CTBP1) [229], a transcriptional repressor that plays a role in cellular proliferation and gene expression during development [230]. LCORL is expressed in bovine fetal tissue, liver and intestine (UniGene, Bt.38533) but expression in other tissues has not been assessed in cattle.

While this region of BTA6 has been implicated in growth and carcass traits in cattle the genetic variant contributing to these phenotypes has not yet been elucidated. In an attempt to further refine the 5 Mb interval on BTA6 influencing back fat, marbling and birth weight we conducted haplotype analysis using BovineHD Beadchip genotypes. Based on results from the haplotype analysis we selected candidate genes for re-sequencing efforts in an attempt to uncover the genetic etiology underlying these complex traits.

Results and Discussion

A whole genome scan for QTL influencing growth, carcass and meat quality traits was conducted in a Simmental-Angus population (Chapter 3) and association with back fat,

marbling and birth weight was detected on BTA6 from approximately 37 to 42 Mb. There were associations (Bonferroni corrected $p \leq 1.0 \times 10^{-3}$) detected for 55 SNPs with back fat, nine SNPs with marbling and 64 SNPs with birth weight on BTA6. Linkage disequilibrium was found to be moderate to high at distances of 200 Kb over much of genic area of the region (Figure 4.5).

There were associations (Bonferroni corrected $p \leq 1.0 \times 10^{-3}$) with 40 single SNPS for back fat (Figure 4.1) covering from 36.99 to 41.19 Mb and the most significant association occurred with BovineHD0600010843 at 39.45 Mb (Bonf. p=3.61E-18). The haplotype analysis (Figure 4.2) exhibited two regions of association, the most significant of these was centered on BovineHD4100004608 (BF_B1) at 39.27 Mb (p= 3.25E-29). The haplotypes and their respective locations relative to one another are detailed in Figure 4.6. The BF_B1 haplotype effect decreased BF and accounted for 8.06% of the variance in BF (Table 4.2b). The second region associated with BF by the haplotype analysis contained two haplotypes of equal significance the first of these beginning with BovineHD4100004575 (BF1) at 38.83 Mb (p= 2.12E-26) and the second began with BovineHD4100004579 (BF2) at 38.84 Mb (p= 2.12E-26). Both BF1 and BF2 decreased BF and each accounted for 6.68% of the variance in BF (Table 4.2a). The BF1 and BF2 haplotypes overlapped by 17 SNPs and exhibited the same genotypes at those 17 SNPs. The combined region covered by the BF1 and BF2 haplotypes spanned from 38.83 to 38.97 Mb. The SNPs in the BF1/BF2 were not in high LD ($r^2 < 0.12$) with those SNPs in BF_B1.

There were eight single SNP associations (Bonferroni corrected $p \leq 1.0 \times 10^{-3}$) with marbling (Figure 4.1) spanning from 39.36 to 39.45 Mb and the most significant (p=2.16E-07) of these was with BovineHD0600010835. The haplotype analysis displayed one region of association and the most significant association (p=2.09E-13) was with the haplotype starting with BovineHD0600010825 (MS1) at 39.36 Mb. The MS1 haplotype effect decreased MS and accounted for 3.25% of the variance in MS (Table 4.2c).

There were associations (Bonferroni corrected $p \leq 1.0 \times 10^{-3}$) with 64 single SNPs for birth weight (Figure 4.1) spanning from 36.99 to 42.06 Mb and the most significant single SNP association was with BovineHD0600010810 located at 39.26 Mb (p=3.14E-16). Further analysis utilizing 20 SNP haplotypes revealed two distinct regions that were associated with BW (Figure 4.2). The first region contained two equally significant associations (p=7.616E-26) with haplotypes beginning at BovineHD0600010750 (BW1) at 38.82 Mb and Hapmap27083-BTC-041166 (BW2) at 38.83 Mb. These two haplotypes overlap by 19 SNPs and display the same genotypes at those 19 SNPs. The BW1 and BW2 haplotypes increase BW and each account for 7.53% of the variance in BW (Table 4.2d). While BW1 and BW2 increase BW the BW23 and BW24 haplotypes located at the same locus decrease BW (p=6.60E-06) and account for 1.43% of the variance in BW. The second region associated with BW, centered on BovineHD4100004608 (BW_B5) at 39.27 Mb, increases BW (p= 2.04E-25) and accounts for 8.03% of the variance in BW (Table 4.2e). The SNPs in the BW1 and BW2 haplotypes were approximately 350 Kb from BW_B5 and the SNPs in BW1/BW2 were not in high LD ($r^2 < 0.13$) with those SNPs in BW_B5.

The BF_B1 and BW_B5 haplotypes overlap on all 20 SNPs and exhibit the same genotypes in haplotype (Table 4.2b & Table 4.2e). The BF_B1 haplotype decreases BF while the BW_B5 haplotype increases birth weight. The BF1 and BW2 haplotypes overlap by 17 SNPs but do not display the same genotypes in the two haplotypes. The BF1 haplotype decreases BF and the BW2 haplotype increases BW which agrees with the inverse correlation seen between the two traits (data not shown).

While both the BF_B1 and MS1 haplotypes are located in close proximity to one another the nearest annotated genes are approximately 300 Kb away. The conservation of this region between bovine, human and murine species was assessed using a percent identity plot [231] and no large highly conserved segment was detected, although this does not rule out the possibility of the presence of a transcriptional regulatory element or a regulatory region of chromatin [232, 233]. The BW1 haplotype reaches from 38.82 to

38.91 Mb and two genes, LCORL and NCAPG, are located near this region. LCORL spans from 38.84 to 38.99 Mb and NCAPG covers from and 38.77 to 38.81 Mb. This LCORL-NCAPG locus has been implicated as contributing to many traits in cattle including body frame size at puberty [209], carcass weight, ribeye area and subcutaneous fat deposition [152, 214]. In humans this locus has been implicated in height traits and has also been associated with height in horses [208, 210-212]. Additionally, LCORL-NCAPG locus has been linked to body size in both horses and dogs [213, 234] as well as body length in pigs [235].

As the BW1/BW2 and BF1/BF2 haplotypes were located in a genic region we chose to explore this region further and focus attention on the birth weight phenotype. The BW1 haplotype locus was selected to evaluate sires for their haplotype status at this locus. Eight sires with the BW1 haplotype and eight sires with the BW23 haplotype were chosen for further interrogation of the locus through direct sequencing. These two haplotypes were chosen as they are opposing in their effect with the BW1 haplotype displaying the largest positive effect on BW and the BW23 haplotype exhibiting the largest negative effect on BW at the locus. In theory, a mutation that is influencing the birth weight phenotype should be consistent with the opposing haplotypes. There were 25 sires homozygous for the BW1 haplotype and 8 sires homozygous for the BW23 haplotype with a frequency in the sires of 47% and 29.5%, respectively. Eight sires that were homozygous for each respective haplotype were selected for direct sequencing of LCORL and NCAPG.

Direct sequencing resulted in discovery of 16 polymorphisms (Table 4.3) in LCORL that were consistent with the BW1 and BW23 haplotypes. Fifteen of these polymorphisms occurred in non-coding regions of the gene and do not appear to disrupt intron/exon boundaries. One polymorphism occurred in the 3’ untranslated region of LCORL. While these mutations do not occur in coding regions of the gene they may still play some yet unknown role in regulation of expression and there is some indication that LCORL may play

a role in cellular process through interactions with ubiquiton and CTBP1 thus, further investigation of these polymorphisms in a larger subset of animals may be warranted.

There were 29 polymorphisms (Table 4.3) discovered through direct sequencing in NCAPG that were consistent with the BW1 and BW23 haplotype status. Twenty-three of the polymorphisms occurred in non-coding regions of the gene that did not disrupt the intron/exon boundary while six of the polymorphisms occurred in coding regions of NCAPG. Four of these coding polymorphisms in NCAPG are synonymous while two are non-synonymous. While four of these mutations are synonymous there is some evidence that even synonymous mutations may have functional consequences through aberrant mRNA splicing or changes in mRNA stability that may influence protein expression as well as possible changes in tertiary protein structure resulting from codon bias [236]. The polymorphism rs110251642 is predicted to result in a leucine to methionine substitution and ss539003533 is predicted to result in an isoleucine to methionine (Ile-442-Met) substitution.

The Ile-442-Met substitution has previously been implicated as contributing to carcass weight [152], body frame size [209] and subcutaneous fat [214, 237] in cattle. The 442Met (c.1326G) allele has been suggested to decrease NCAPG expression level in bovine fetal placenta [207] but this mutation was assayed in our sire population and was found to be incongruent with the BW1 haplotype status. Eberlein *et al.* [207] did suggest that the decreased NCAPG expression associated with the 442Met (c.1326G) allele resulted in increased fetal growth however; the result was not statistically significant.

The effect of the Ile-442-Met substitution was evaluated for association with metabolism and growth at puberty as well as carcass composition in cattle [237]. The 442Met allele was associated with increased average daily gain, increased protein accretion and decreased fat content in the carcass resulting from decreased subcutaneous fat deposition. The 442Met allele was associated with increased circulating arginine levels and as arginine is a conditionally indispensable amino acid that is thought to play a substantial

role in growth [238] the increase in circulating arginine seen may be impacting average daily gain, lean tissue and fat deposition in the cattle [237].

It was suggested by Eberlein *et al*. [207] that increased NCAPG expression in bovine fetal placentome resulted in decreased birth weight but this is contradictory to research in pigs [239] where decreased expression of fetal NCAPG resulted in decreased birth weights. Piglets from mothers fed low protein diets exhibited decreased NCAPG expression in utero, decreased birth weight, and while final body weights at harvest were not different from controls, the amount of lean tissue was decreased while subcutaneous fat was increased. While this research appears contradictory to data from Eberlein *et al.* [207] consideration has to be given to the involvement of the low protein diet that likely limited amino acids necessary for in utero growth. It is of interest that expression of NCAPG may be modulated by maternal dietary status and that this may impact growth and carcass characteristics later in life.

In addition to the SNP and insertion/deletion polymorphisms, a simple repeat insertion mutation (NCAPGins) was discovered in intron 5 of NCAPG that is located approximately 100 bp from the 3' end of exon 5 and approximately 977 bp from the 5' end of exon 6. The insertion mutation appears to be a simple "AT" repeat that is approximately 600-800 bp in length (Figure 4.3). While this mutation does not occur in coding sequence nor does it disrupt an exon/intron boundary, mutations in intronic regions may interfere with intronic splicing elements that regulate alternative splicing events [240]. The sire population was assayed and the NCAPGins status was consistent with BW1 haplotype in 59 of 61 sires. The NCAPGins polymorphism should be evaluated further in a larger population of animals as splicing events are known to play a role in disease modification and susceptibility [240], thus this mutation may indeed be contributing to the association with the birth weight phenotype at this locus.

Conclusion

The interrogation of a previously discovered locus on BTA6 contributing to back fat, birth weight and marbling phenotypes in a Simmental-Angus population was conducted in an attempt to refine the locus interval and potentially discover a genetic variant contributing to the association seen in the region. The use of haplotype analysis revealed two separate regions, one near 38.83 Mb and the second near 39.27 Mb, associated with both back fat and birth weight. The most significantly associated haplotypes with back fat and birth weight were overlapping at both of the regions. Haplotype analysis also revealed one locus contributing to the association with marbling near 39.26 Mb. The region at 38.83 Mb is in proximity of the LCORL- NCAPG locus that has been implicated in many species as controlling body size and contributing to growth and carcass characteristics in cattle. Re-sequencing efforts of this region revealed several polymorphisms that were consistent with haplotype status but none appear to be the causative genetic variant underlying the association detected with birth weight on BTA6. The evidence of the contribution of the LCORL-NCAPG locus to growth and carcass phenotypes in cattle is mounting and further investigation of the region is warranted.

Methods

Genotype imputation and genome wide association analysis

Steers of a Simmental, Angus or Simmental x Angus crossbred background were obtained from five ranches in Montana over a five year period. Dams were randomly assigned to sire. Birth weight and weaning weight were recorded on each ranch. Animal management, dietary treatments and data collection methods are reviewed in Trejo 2010 (Thesis). Traits evaluated are categorized as growth, meat quality or carcass traits. Growth traits were birth weight (BW), weaning weight (WW), yearling weight (YW). Meat

quality traits were marbling score (MS) and Warner-Bratzler shear force (SF). Carcass traits were hot carcass weight (HCW), ribeye area (REA), back fat (BF), kidney, pelvic and heart fat (KPH) and yield grade (YG). Multiple blood samples were collected on each individual by jugular puncture. DNA was isolated using the simple salting-out method [140] and samples frozen at -20 C. Steers were parentally validated to their respective sires using a panel of 15 microsatellite markers consisting of 10 of the approved ISAG bovine markers and five additional markers. There were a total of ~1800 parentally verified steers from 133 sires over the five year period.

Animals were selected for genotyping based on divergent adjusted residual phenotypes which were attained using the PROC MIXED option of the SAS® system (Version 9.2, SAS Inst., Inc., Cary, NC). A stepwise reduction method with backward selection was used and variables significant at $p \leq 0.05$ were retained in the model. The full model for carcass and meat quality traits included year, ranch, days on feed (DOF), treatment, days of age (DOA), HCW, sire breed, dam breed and the interaction between sire and dam breeds. The full model for growth traits included year, ranch, birth month, birth weight, DOA, DOF, sire breed, dam breed and the interaction between sire and dam breeds. The upper and lower 20 animals were selected for each trait for the preliminary analysis. An additional 97 and 50 genotypes were available from years two and four, respectively, resulting in a total of 742 animals in the preliminary analysis. Additional animals were selected for genotyping based on completeness of phenotypic data resulting in a total of 1672 genotyped offspring representing 127 sires. Those sires with six or more genotyped offspring were selected for genotyping on the BovineHD BeadChip resulting in a total of 111 sires.

DNA was sent to Geneseek (Lincoln, NE) for genotyping on the Illumina® BovSNP50 BeadChip and the BovineHD BeadChip platforms. Quality control measures were completed using GoldenHelix SNP & Variation Suite v 7.6.11 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). Individuals or markers with more than 10% missing genotypes and

markers with minor allele frequency less than 1% were excluded from the analysis. After quality control measures there were 46,334 and 646,510 markers from the BovSNP50 and the BovineHD, respectively, and 1,644 steers and 110 sires. There were 41,728 markers from the BovSNP50 that were common between the SNP50 and BovineHD that were used during imputation. A hidden markov model procedure as implemented in Beagle (version 3.3.2) was used for imputation. The default parameters of BEAGLE v 3.3.2 [77] were used with 10 phasing iterations and four haplotype pairs were sampled for each individual during each iteration of phasing. The sire HD genotypes were phased and then used as a reference panel [75] to impute genotypes in the steers. The r2 score generated by BEAGLE was used as a threshold for imputation quality and only markers exceeding an r2 of 0.5 were used for the whole genome scan resulting in 631,665 markers. In order to assess the accuracy of imputation a total of 52 steers were genotyped on the BovineHD assay. Concordance was assessed using Plink v 1.06 [110] and at a Beagle quality control r2 of 0.5 concordance was found to be 93.2%.

After quality control measures, the number of animals with both genotypes and phenotypes ranged between 1412 and 1417 for growth traits and from 1632 to 1640 for carcass traits. A total of 672 animals had both genotypes and phenotypes for Warner-Bratzler shear force and 1638 for marbling.

A genome wide association scan using the imputed genotypes and the residual phenotypes for each trait was conducted using linear regression with an additive model in GoldenHelix SVS v 7.6.11. Results were reported for associations with SNPs at $p \leq 1.0 \times 10^{-3}$.

Cattle and humans have been found to be similar in regards to LD and haplotype structure for distances of 1 to 100 Kb [215]. Investigation of known disease risk variants in human populations has shown that an association of SNPs with the risk variant does not exceed 500 Kb under the largest distance scenario [216]. An evaluation of LD in this population showed that LD lessened considerably after 500 Kb. Thus, 500 Kb flanking

either the most significant SNP of a region or of a single significant SNP will be scanned for probable candidate genes based on known function and tissue expression.

Haplotype analysis

Regions of the genome with associations that reached the significance threshold were investigated further with haplotype analysis. The E-M algorithm, available in PLINK v 1.06, was used to phase and develop 20 SNP haplotypes with a sliding window in the regions flanking significant associations. A quantitative haplotype association test was completed using PLINK v 1.06.

Primer design and PCR

All primers were designed using Primer Designer v 2.0 (Scientific and Educational Software). Sequences were masked for repetitive elements using RepeatMasker [142] prior to primer design. Primer sequences are provided in Table S7 (See supplementary file Markey_thesis_supplement.xlsx). Exons were determined using mRNA sequences from NCBI and the software SPIDEY [144]. Primers were designed to amplify exons in 1kb fragments. Exons were amplified in 20µl reactions using ~ 20-50 ng genomic DNA with final concentrations of 1x PCR buffer (containing 1.5 mM $MgCl_2$; Qiagen, Valencia, CA), 200 µM of each dNTP (Fermentas, Glen Burnie, MD), 0.5 µM of each primer and 0.5 units of HotStarTaq *Plus* DNA Polymerase (Qiagen, Valencia, CA). Cycling parameters varied for each amplicon, but typically involved an initial activation of 96°C for 5 minutes followed by 35 cycles of 95°C for 45 s, 56-62°C for 45 s and 72°C for 60 s followed by a final extension at 72°C for 10 minutes. Amplified PCR products were subjected to electrophoresis on a 1.2% 0.5X Tris-borate EDTA agarose gel to verify amplicon quality and primer specificity. Ten µl of amplified PCR product was then treated with 4 µl of a 1:12 dilution of ExoSAP-IT®

(Affymetrix, Inc., Santa Clara, CA) and incubated at 37°C for 30 minutes then heat inactivated for 15 minutes at 80°C.

Re-sequencing was performed in 8µl reactions using 2µl ExoSAP-IT® treated PCR product, 3.62µl buffer (0.16 M Tris (pH 9.0), 3 mM $MgCl_2$, 4.9% tetramethylene sulfone, 0.0001% Tween-20® surfactant), 0.25µl BigDye v3.1, 0.08 µl BigDye dGTP v3.0, and 0.656 µM of each primer. Sequencing was done with an initial denaturation of 96°C for 90 s, followed by 55 cycles of 96°C for 15 s, 53°C for 15 s and 60°C for 3 minutes with a final extension of 60°C for 10 minutes. Sequence products were purified using Sephadex™ G-50 Fine (GE Healthcare) and ran on an ABI 3730XL (Applied Biosystems, Foster City, CA) capillary sequencer. Bases were called using Phred, sequences assembled using Phrap and sequence assemblies viewed and analyzed for polymorphisms using Codon Code Aligner (Codon Code Corporation).

Figures

Figure 4.1. Manhattan plots of single nucleotide polymorphisms Bonferonni corrected associations. The red line corresponds to p = 1.0 x 10^{-3} that was used as a significance threshold in the initial whole genome association analysis. The X axis displays the bovine chromosomes and the Y axis corresponds to the negative log 10 of the Bonferroni corrected p-value.

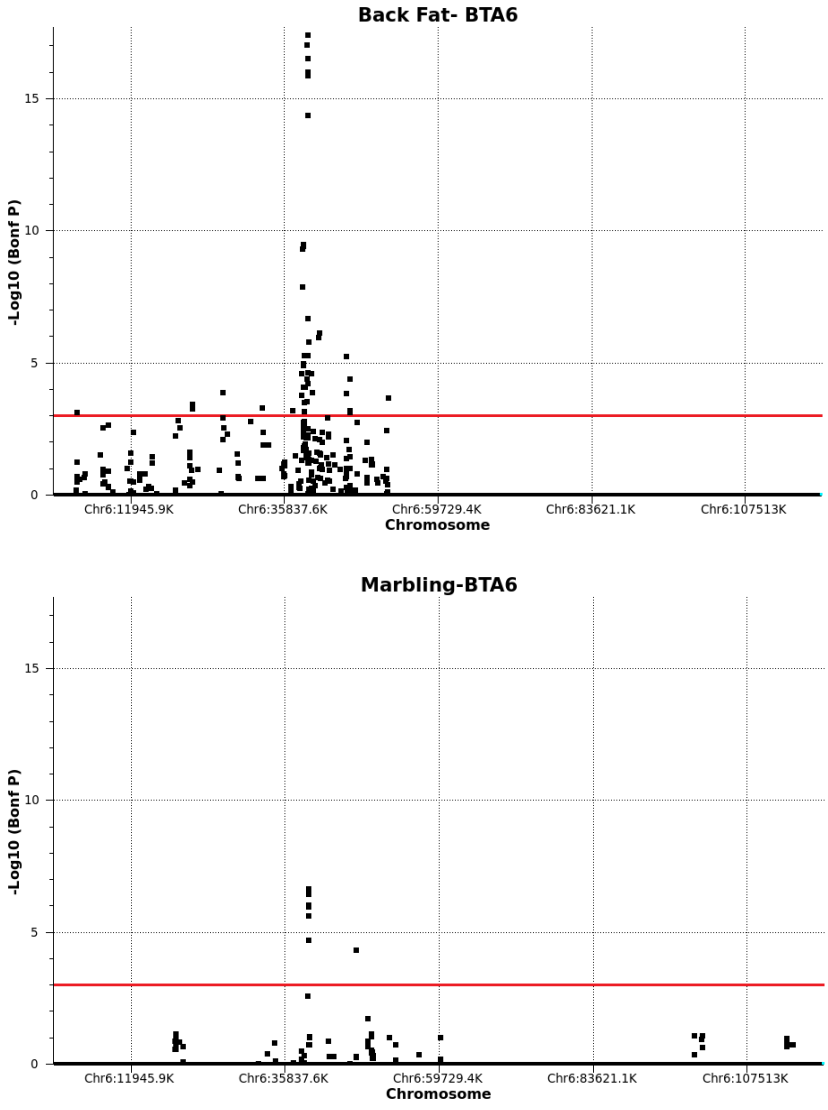


Figure 4.1. Continued

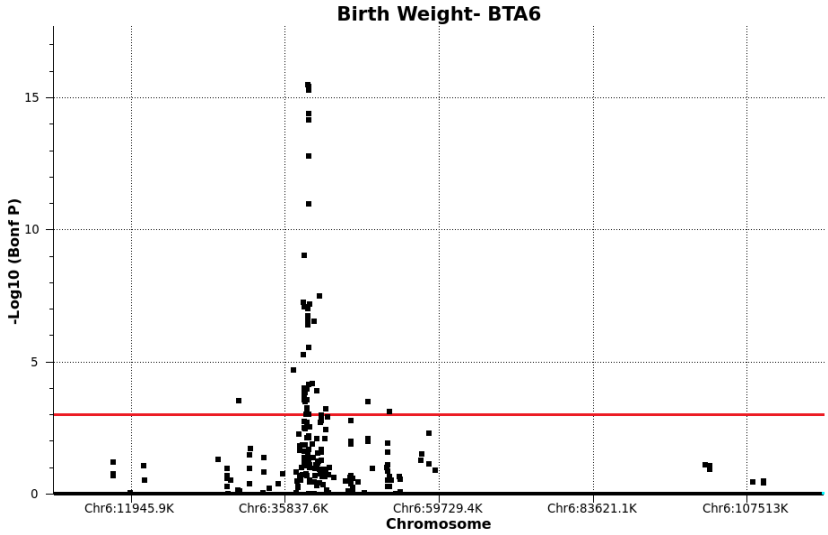


Figure 4.2. Manhattan plots of haplotype association analysis for back fat, marbling and birth weight on BTA6. A haplotype analysis using a twenty SNP sliding window was conducted for each trait using quantitative association analysis in PLINK to interrogate the region from 36 to 43 Mb. A schematic representing genes from University of Maryland genome assembly v 3.0 is beneath each Manhattan plot. The X axis represents the position along BTA6 and the Y axis represents the asymptotic p-value from the association analysis.

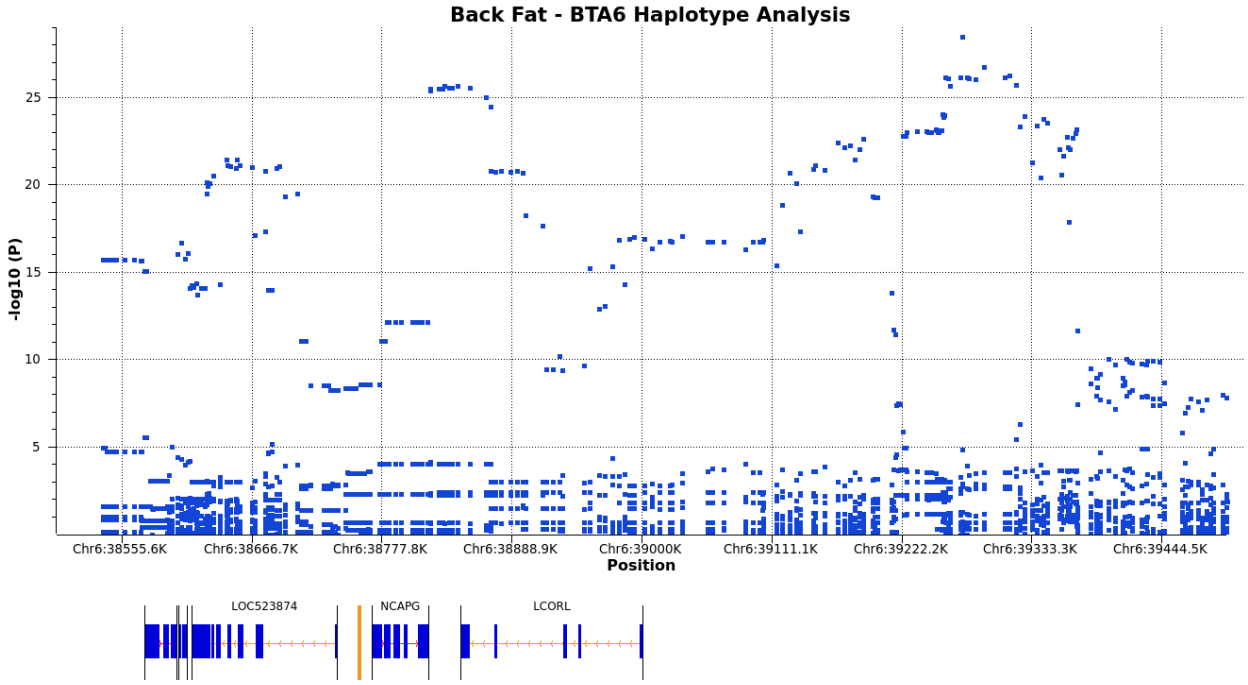
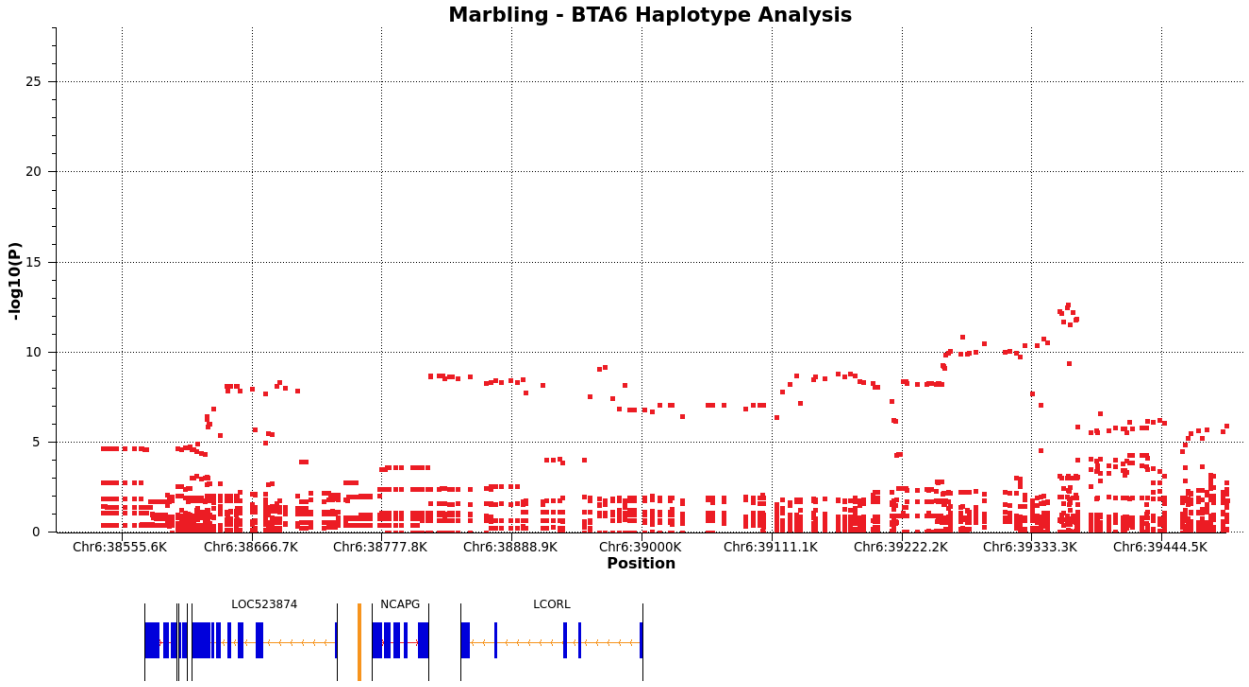


Figure 4.2. Continued



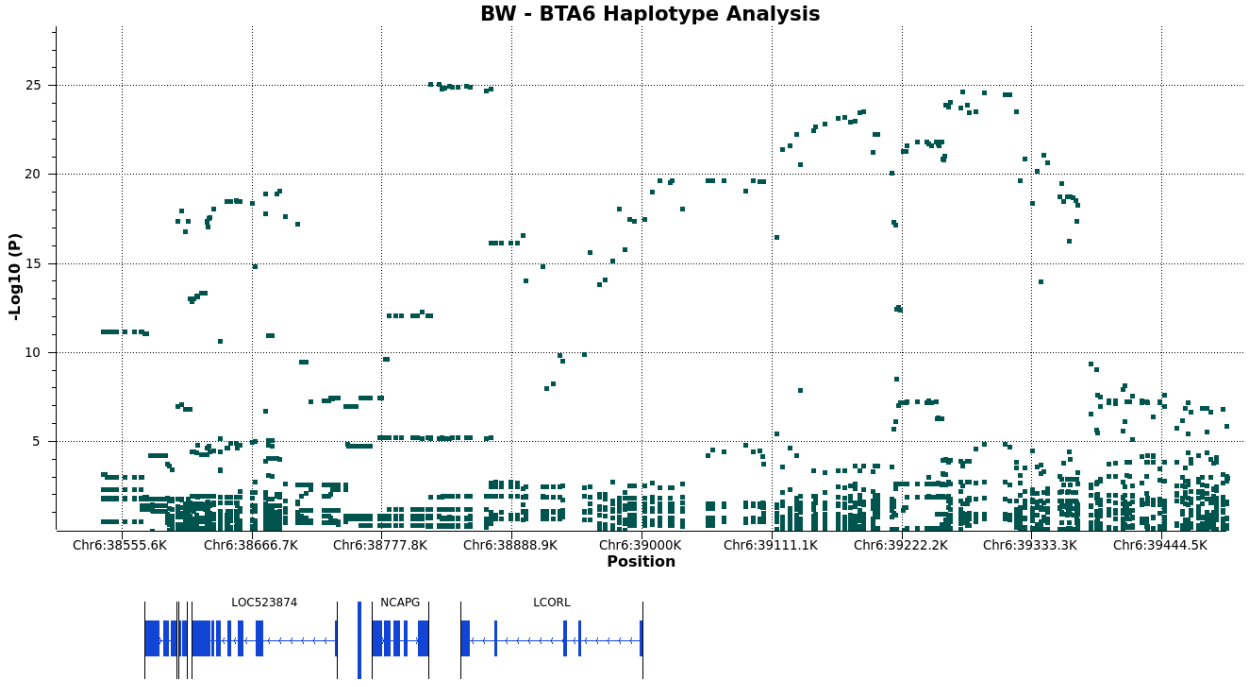


Figure 4.3. DNA sequence from normal and NCAPGins animals. Underlined sequence indicates primer sequences used as a diagnostic to amplify the normal and mutated alleles.

>NCAPG_InsertSeq
AAGCCATAATCTTAAAATACCTTTGAAAAATATTTG
GCATGTAACAATTTCTTCAATTGAGATTTTTATGGACAGAGTATAAATAA
TTTAGGTAGAATCCTATCGCATTATTGTGTGGCCAGGTATTAATTTTTTT
GGAGATTGTAGATAATTTTCTTAGGTGAAGGAGATATTACTAGTTATAAA
ATGCATAAATTATATTGAGTTTAATGGCCATTTTTTACTGTCTCTTAAAG
GTTTTAGCTGAAAAGGTTCACATGAGAGCTCTGTCCATTGCTCAGAGAGT
AATGCTCCTTCAACAAGGTCTCAATGACCGATCAGGTAATTTAAACATCT
TTATACATAAAACTTTAGTAGATTTTGAGGTCAAACTAAAAGGTTTAGGA
AAGAGTGTCCTATTAATAAATAATTATAATATATATAATAATATATATAT
ATATATTTATATAATATATATATAATATATATATAATATATATATATAAT
TATAATATATATATAATATATATATATATATATATAATATAATATATT
ATATATATAATAATATATAATATATAATAATATATATAATATATAATATA
TAATAATTAATTAATATTATATATATAATAATATATAATATATAATAATA
TATATAATATATAATATATAATAATTAATTAATATAATATATATATAATT
AAAAAAATAATTAATATAATATATAATAATTAATTATATATATATATATA
TATATATAATAAATAAATATATATATATATATATATTATATAAATAAATA
AATAAATAATAAAAAAAAATAGATCAGGTAAATCTATCATTTCCACAGCC
TGTCAAAGTTTATCTAGGTGTATTGTATAGAAGACATTTTTCAAGATCTT
GGTTCTTAGAAGTAGCTCAGTTTGTTTTTTTCTGTGTTTGGTGTATGGAG
AAACTAGAGTTTAAGGCTGTAAAAATAACCTACGTCAGCAGTAGGGCAGA
ATCTCTTTTTCAGCCCTCTCTCTTAAAAAAAAAAATGCTATCTTATCCCT
TTTGTGTATATCCTGCTATGAATGTGTTCTGTACATCTAAAACAAGTAGA
AATGGTCAGACCAATGTAATTTCCCTGTTTCCCACCTTTTCTCTTTTATA
ATTGGAGAATTGACTTACCGCAG

>NCAPG_normal
AAGCCATAATCTTAAAATACCTTTGAAAAATATTTGGCATGTAACAATTT
CTTCAATTGAGATTTTTATGGACAGAGTATAAATAATTTAGGTAGAATCC
TATCGCATTATTGTGTGGCTGGGTATTAATTTTTTTGGAGATTGTAGATA
ATTTTCTTAGGTGAAGGAGATATTACTAGTTATAAAATGCATAAATTATA
TTGAGTTTAATGGCCATTTTTTACTGTCTCTTAAAGGTTTTAGCTGAAAA
GGTTCACATGAGAGCTCTGTCCATTGCTCAGAGAGTAATGCTCCTTCAAC
AAGGTCTCAATGACCGATCAGGTAATTTAAACATCTATACATAAAACTTT
AGTAGATTTTGAGGTCAAACTAAAAGGTTTAGGAAAAGAGTGTCCTATTAA
TAAATTAGATCAGGTAAATCTATCATTTCCACAGCCTGTCAAAGTTTATC
TAGGTGTATTGTATAGAAGACATTTTTCAAGATCTTGGTTCTTAGAAGTA
GCTCAGTTTGTTTTTTTCTGTGTTTGGTGTATGGAGAAACTAGAGTTTAA
GGCTGTAAAAATAACCTACGTCAGCAGTAGGGCAGAATCTCTTTTTCAGC
CCTCTCTCTTAAAAAAAAAATGCTATCTTATCCCTTTTGTGTAT

Figure 4.4. Alignment of sequence generated from Sanger sequencing of normal and NCAPGins alleles.

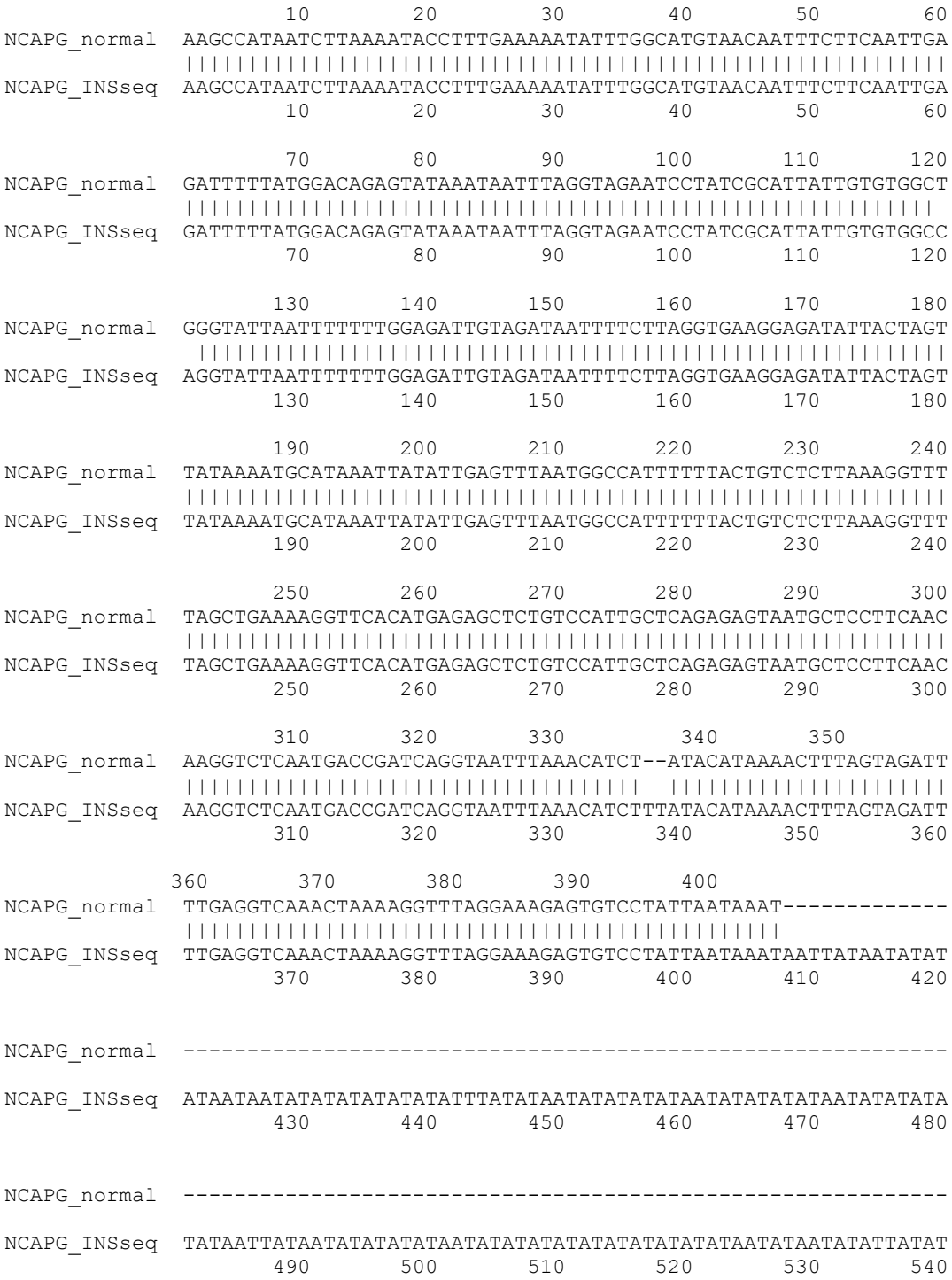


Figure 4.4. Continued

```
NCAPG_normal  ------------------------------------------------------------

NCAPG_INSseq  ATATAATAATATATAATATATAATAATATATATAATATATAATATATAATAATTAATTAA
                     550       560       570       580       590       600


NCAPG_normal  ------------------------------------------------------------

NCAPG_INSseq  TATTATATATATAATAATATATAATATATAATAATATATATAATATATAATATATAATAA
                     610       620       630       640       650       660


NCAPG_normal  ------------------------------------------------------------

NCAPG_INSseq  TTAATTAATATAATATATATATAATTAAAAAAATAATTAATATAATATATAATAATTAAT
                     670       680       690       700       710       720


NCAPG_normal  ------------------------------------------------------------

NCAPG_INSseq  TATATATATATATATATATATATAATAAATAAATATATATATATATATATATTATATAAA
                     730       740       750       760       770       780

                                       410       420       430       440
NCAPG_normal  -------------------------TAGATCAGGTAAATCTATCATTTCCACAGCCTGTC
                                       |||||||||||||||||||||||||||||||||||
NCAPG_INSseq  TAAATAAATAAATAATAAAAAAAAATAGATCAGGTAAATCTATCATTTCCACAGCCTGTC
                     790       800       810       820       830       840

                     450       460       470       480       490       500
NCAPG_normal  AAAGTTTATCTAGGTGTATTGTATAGAAGACATTTTTCAAGATCTTGGTTCTTAGAAGTA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NCAPG_INSseq  AAAGTTTATCTAGGTGTATTGTATAGAAGACATTTTTCAAGATCTTGGTTCTTAGAAGTA
                     850       860       870       880       890       900

                     510       520       530       540       550       560
NCAPG_normal  GCTCAGTTTGTTTTTTTCTGTGTTTGGTGTATGGAGAAACTAGAGTTTAAGGCTGTAAAA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NCAPG_INSseq  GCTCAGTTTGTTTTTTTCTGTGTTTGGTGTATGGAGAAACTAGAGTTTAAGGCTGTAAAA
                     910       920       930       940       950       960

                     570       580       590       600       610       620
NCAPG_normal  ATAACCTACGTCAGCAGTAGGGCAGAATCTCTTTTTCAGCCCTCTCTCTTAAAAAAAAAA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NCAPG_INSseq  ATAACCTACGTCAGCAGTAGGGCAGAATCTCTTTTTCAGCCCTCTCTCTTAAAAAAAAAA
                     970       980       990      1000      1010      1020

                      630       640
NCAPG_normal  -TGCTATCTTATCCCTTTTGTGTAT
               |||||||||||||||||||||||
NCAPG_INSseq  ATGCTATCTTATCCCTTTTGTGTAT
                    1030      1040
```

Figure 4.5. Pairwise linkage disequilibrium calculated as r^2 between all SNPs in BTA6 interval from 38.5-39.5 Mb.

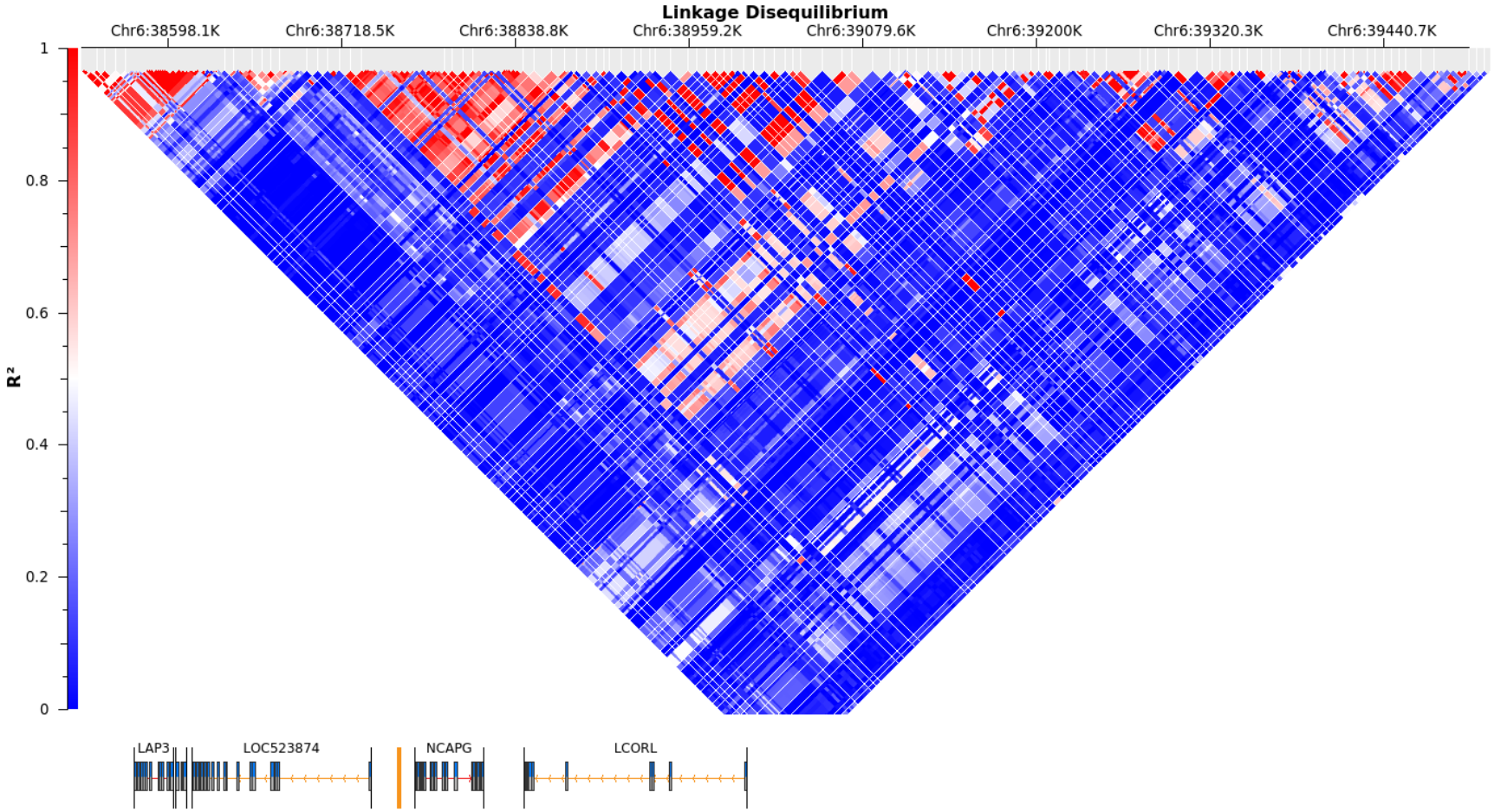
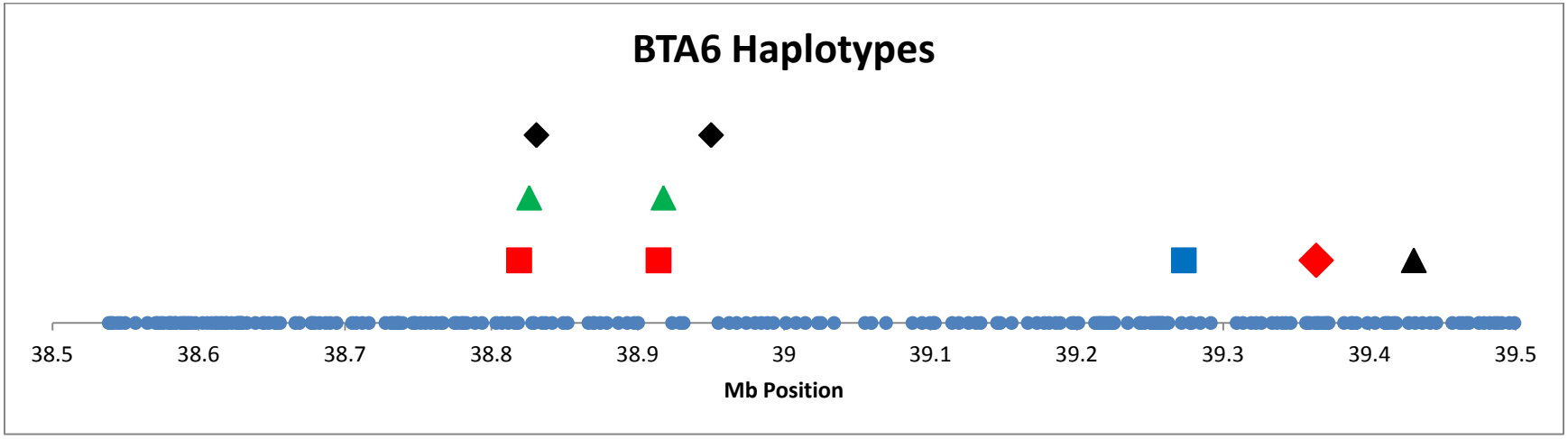


Figure 4.6. Graphical representation of the relative position of the 20 SNP haplotypes in the region from 38.5 to 39.5 Mb on BTA6. The X-axis represents the Mb position based on the University of Maryland bovine genome assembly 3.0. There were 239 SNPs in the 1 Mb window. Each blue circle along the X axis represents an individual SNP. The beginning and end of each 20 SNP haplotype near the 38.8 to 38.9 Mb region is represented by the respective matching shape: BW1 (red squares), BW2 (green triangles), BF1 (black diamonds). The beginning of the BF_B1/BW_B5 haplotypes is represented by the blue square and the end of BF_B1/BW_B5 haplotype is represented by the red diamond. The red diamond also represents the beginning of the MS1 haplotype and the black triangle represents the end of the MS1 haplotype.



Tables

Table 4.1. Summary of selected single SNP associations with back fat, birth weight and marbling on BTA6. SNPs represented exhibited Bonferroni corrected $p \leq 1.0 \times 10^{-3}$.

Trait	Marker	Bonf P	Chr	Position
BF	BovineHD0600000855	0.000720	6	3469170
BF	BovineHD0600005912	0.000506	6	21402101
BF	BovineHD0600005914	0.000506	6	21413804
BF	BovineHD0600005927	0.000346	6	21467193
BF	Hapmap27407-BTA-143867	0.000126	6	26205206
BF	BovineHD0600009070	0.000477	6	32331126
BF	BovineHD4100004461	0.000638	6	36994858
BF	BovineHD0600010627	0.000160	6	38387091
BF	BovineHD4100004559	2.49E-05	6	38390690
BF	BovineHD0600010631	2.49E-05	6	38410258
BF	BovineHD0600010665	1.26E-08	6	38571977
BF	Hapmap26308-BTC-057761	4.54E-10	6	38576012
BF	BovineHD0600010697	3.62E-10	6	38644886
BF	BovineHD0600010698	3.62E-10	6	38647897
BF	BovineHD0600010699	7.85E-05	6	38652252
BF	BovineHD0600010700	7.85E-05	6	38653201
BF	BovineHD0600010701	3.13E-10	6	38655605
BF	BovineHD0600010713	1.16E-05	6	38686919
BF	BovineHD0600010716	1.01E-05	6	38704872
BF	BovineHD4100004580	5.16E-06	6	38852093
BF	Hapmap31285-BTC-041097	0.000292	6	38869785
BF	BovineHD0600010759	0.000688	6	38897628
BF	BovineHD4100004586	7.76E-05	6	38955154
BF	BovineHD0600010810	8.87E-18	6	39262655
BF	BovineHD4100004607	0.000288	6	39271661
BF	BovineHD0600010812	3.89E-05	6	39278657
BF	BovineHD0600010824	4.23E-15	6	39356766
BF	BovineHD4100004616	1.98E-07	6	39358026
BF	BovineHD4100004617	5.59E-05	6	39359444
BF	BovineHD4100004620	2.18E-05	6	39367782
BF	BovineHD0600010827	5.21E-06	6	39372172
BF	BovineHD0600010835	1.05E-16	6	39413342
BF	BovineHD0600010836	2.84E-17	6	39416405
BF	BovineHD0600010838	1.33E-16	6	39426854
BF	BovineHD0600010839	2.05E-07	6	39430325
BF	BovineHD0600010840	9.08E-17	6	39431268
BF	BovineHD0600010841	1.33E-16	6	39436297
BF	BovineHD0600010842	1.33E-16	6	39441548
BF	BovineHD0600010843	3.61E-18	6	39445731
BF	BovineHD0600010865	1.49E-06	6	39579598
BF	BovineHD0600010866	1.49E-06	6	39580784
BF	BovineHD0600010938	2.39E-05	6	39935697
BF	BovineHD4100004688	0.000132	6	40072096
BF	BovineHD0600011159	1.05E-06	6	41079201
BF	BovineHD0600011178	7.24E-07	6	41186810

Table 4.1. Continued

Trait	Marker	Bonf P	Chr	Position
BF	BovineHD0600011179	7.24E-07	6	41190295
BF	BovineHD0600012352	5.67E-06	6	45340181
BF	BovineHD0600012356	0.000137	6	45353131
BF	BovineHD0600012477	4.03E-05	6	45920252
BF	Hapmap48206-BTA-119876	0.000622	6	45945033
BF	BovineHD0600012491	0.000732	6	45946691
BF	BovineHD0600012492	0.000732	6	45948116
BF	BovineHD0600014301	0.000206	6	51890653
BF	BovineHD0600014302	0.000206	6	51891293
BF	BovineHD0600014304	0.000206	6	51892795
BW	BovineHD0600007939	0.000290	6	28592769
BW	BovineHD4100004461	1.98E-05	6	36994858
BW	BovineHD0600010665	5.42E-08	6	38571977
BW	Hapmap26308-BTC-057761	5.05E-06	6	38576012
BW	BovineHD0600010697	8.90E-10	6	38644886
BW	BovineHD0600010698	8.90E-10	6	38647897
BW	BovineHD0600010699	0.000129	6	38652252
BW	BovineHD0600010700	0.000129	6	38653201
BW	BovineHD0600010701	7.62E-08	6	38655605
BW	BovineHD0600010704	0.000250	6	38666081
BW	BovineHD0600010705	0.000196	6	38668893
BW	BovineHD0600010713	0.000123	6	38686919
BW	BovineHD0600010716	9.63E-05	6	38704872
BW	BovineHD4100004580	0.000140	6	38852093
BW	Hapmap31285-BTC-041097	0.000306	6	38869785
BW	BovineHD4100004586	0.000922	6	38955154
BW	BovineHD0600010787	0.000104	6	39119600
BW	BovineHD0600010794	0.000631	6	39177795
BW	BovineHD0600010795	0.000256	6	39185743
BW	BovineHD0600010796	0.000505	6	39188643
BW	BovineHD4100004600	2.11E-07	6	39234851
BW	BovineHD4100004602	2.44E-07	6	39244034
BW	BovineHD0600010803	2.44E-07	6	39250536
BW	BovineHD4100004604	2.44E-07	6	39251631
BW	BovineHD0600010804	2.44E-07	6	39252552
BW	BovineHD0600010805	2.44E-07	6	39253604
BW	BovineHD0600010806	2.44E-07	6	39254959
BW	BovineHD0600010807	2.44E-07	6	39255646
BW	BovineHD4100004605	2.44E-07	6	39256262
BW	Hapmap27537-BTC-060891	2.44E-07	6	39257620
BW	BovineHD0600010808	2.44E-07	6	39258226
BW	BovineHD0600010809	2.44E-07	6	39259257
BW	BovineHD4100004606	3.70E-07	6	39261628
BW	BovineHD0600010810	3.14E-16	6	39262655
BW	BovineHD4100004607	9.51E-08	6	39271661
BW	BovineHD0600010812	1.76E-07	6	39278657
BW	BovineHD0600010824	9.92E-12	6	39356766
BW	BovineHD0600010827	0.000902	6	39372172
BW	BovineHD0600010830	6.95E-05	6	39398681
BW	BovineHD0600010831	6.95E-05	6	39403857

Table 4.1. Continued

Trait	Marker	Bonf P	Chr	Position
BW	BovineHD0600010835	3.57E-16	6	39413342
BW	BovineHD0600010836	3.69E-15	6	39416405
BW	BovineHD0600010838	4.80E-16	6	39426854
BW	BovineHD0600010839	2.76E-06	6	39430325
BW	BovineHD0600010840	1.53E-13	6	39431268
BW	BovineHD0600010841	4.80E-16	6	39436297
BW	BovineHD0600010842	4.80E-16	6	39441548
BW	BovineHD0600010843	6.75E-15	6	39445731
BW	BovineHD0600010865	6.24E-08	6	39579598
BW	BovineHD0600010866	6.24E-08	6	39580784
BW	BovineHD0600010938	6.06E-05	6	39935697
BW	BovineHD0600011009	2.65E-07	6	40275071
BW	BovineHD4100004732	0.000120	6	40700584
BW	BovineHD0600011088	0.000120	6	40705941
BW	BovineHD4100004733	0.000120	6	40707879
BW	BovineHD0600011159	3.01E-08	6	41079201
BW	BovineHD0600011206	0.000947	6	41313370
BW	BovineHD4100004749	0.000947	6	41320812
BW	BTB-00251468	0.000589	6	42057261
BW	BovineHD0600013364	0.000292	6	48608641
BW	BovineHD0600013365	0.000292	6	48609889
BW	BovineHD0600014301	0.000727	6	51890653
BW	BovineHD0600014302	0.000727	6	51891293
BW	BovineHD0600014304	0.000727	6	51892795
Marb	BovineHD0600010824	1.98E-05	6	39356766
Marb	BovineHD0600010835	2.16E-07	6	39413342
Marb	BovineHD0600010836	1.02E-06	6	39416405
Marb	BovineHD0600010838	3.44E-07	6	39426854
Marb	BovineHD0600010840	8.76E-07	6	39431268
Marb	BovineHD0600010841	3.44E-07	6	39436297
Marb	BovineHD0600010842	3.44E-07	6	39441548
Marb	BovineHD0600010843	2.38E-06	6	39445731
Marb	BovineHD0600012785	4.57E-05	6	46837171

Table 4.2a. Selected results from haplotype association analysis with back fat on BTA6.

[a]Beta statistic is an indicator of the direction of the effect

[b]Percent of variation explained by the haplotype

[c]Asymptotic p-value

[d]SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

[e]Base pair position from University of Maryland bovine genome assembly 3.0 for SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

HapID	Haplotype	NoAnm	Beta[a]	%Var[b]	P[c]	SNP_1[d]	SNP_20	Pos1[e]	Pos2
BF6	AAABABBBBBBABAAABBBAB	1638	-0.05786	6.65	2.75E-26	BovineHD0600010749	Hapmap33630-BTC-041044	38818283	38914175
BF14	BABABAAAAABABABBAAAB	1638	0.01838	0.94	8.32E-05	BovineHD0600010749	Hapmap33630-BTC-041044	38818283	38914175
BF10	AABABBBBBABAAABBBABB	1636	-0.05771	6.62	3.76E-26	BovineHD0600010750	BovineHD0600010761	38818907	38917456
BF13**	ABABAAAAABABABBAAABB	1636	0.01859	0.96	6.94E-05	BovineHD0600010750	BovineHD0600010761	38818907	38917456
BF7	ABABBBBBABAAABBBABBB	1638	-0.05786	6.65	2.75E-26	Hapmap27083-BTC-041166	BovineHD4100004583	38825835	38923533
BF15	BABAAAAABABABBAAABBA	1638	0.01838	0.94	8.32E-05	Hapmap27083-BTC-041166	BovineHD4100004583	38825835	38923533
BF8	BABBBBBABAAABBBABBBA	1638	-0.05786	6.65	2.75E-26	BovineHD0600010751	BovineHD4100004585	38827869	38928941
BF21	ABAAAAABABABBAAABBAA	1638	0.01825	0.93	9.24E-05	BovineHD0600010751	BovineHD4100004585	38827869	38928941
BF9	ABBBBBABAAABBBABBBAB	1638	-0.05786	6.65	2.75E-26	BovineHD0600010752	BovineHD0600034280	38829248	38931436
BF22	BAAAAABABABBAAABBAAB	1638	0.01825	0.93	9.24E-05	BovineHD0600010752	BovineHD0600034280	38829248	38931436
BF1*	BBBBBABAAABBBABBBABB	1638	-0.05797	6.68	2.12E-26	BovineHD4100004575	BovineHD0600010764	38830725	38950070
BF23	AAAAABABABBAAABBAABA	1638	0.01825	0.93	9.24E-05	BovineHD4100004575	BovineHD0600010764	38830725	38950070
BF3	BBBBABAAABBBABBBABBB	1635	-0.05799	6.68	2.42E-26	BovineHD4100004576	BovineHD4100004586	38834676	38955154
BF17	AAAABABABBAAABBAABAA	1635	0.01836	0.94	8.65E-05	BovineHD4100004576	BovineHD4100004586	38834676	38955154
BF4	BBBABAAABBBABBBABBBB	1635	-0.05799	6.68	2.42E-26	BovineHD4100004577	BovineHD0600010766	38837159	38962544
BF18	AAABABABBAAABBAABAAB	1635	0.01836	0.94	8.65E-05	BovineHD4100004577	BovineHD0600010766	38837159	38962544
BF2*	BBABAAABBBABBBABBBBA	1638	-0.05797	6.68	2.12E-26	BovineHD4100004579	BovineHD0600010767	38841588	38967708
BF16	AABABABBAAABBAABAABA	1638	0.01838	0.94	8.32E-05	BovineHD4100004579	BovineHD0600010767	38841588	38967708
BF5	BABAAABBBABBBABBBBAA	1635	-0.05799	6.68	2.42E-26	BovineHD4100004580	BovineHD0600010769	38852093	38974284
BF19	ABABABBAAABBAABAABAA	1635	0.01836	0.94	8.65E-05	BovineHD4100004580	BovineHD0600010769	38852093	38974284
BF11	ABAAABBBABBBABBBBAAA	1638	-0.05738	6.51	9.34E-26	BovineHD0600010755	BovineHD0600010770	38866381	38980227
BF24	BABABBAAABBAABAABAAA	1638	0.01825	0.93	9.24E-05	BovineHD0600010755	BovineHD0600010770	38866381	38980227
BF12	BAAABBBABBBABBBBAAAB	1635	-0.05703	6.39	3.05E-25	Hapmap31285-BTC-041097	BovineHD0600010771	38869785	38984478
BF20	ABABBAAABBAABAABAAAB	1635	0.01836	0.94	8.65E-05	Hapmap31285-BTC-041097	BovineHD0600010771	38869785	38984478

* most significant haplotype decreasing the phenotype

** most significant haplotype increasing the phenotype

Table 4.2b. Selected results from haplotype association analysis with back fat in region 2 on BTA6.

[a]Beta statistic is an indicator of the direction of the effect

[b]Percent of variation explained by the haplotype

[c]Asymptotic p-value

[d]SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

[e]Base pair position from University of Maryland bovine genome assembly 3.0 for SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

HapID	Haplotype	NoAnm	Beta[a]	%Var[b]	P[c]	SNP_1[d]	SNP_20	Pos1[e]	Pos2
BF_B4	BABAABAAABAAAABBAABA	1638	-0.05843	6.81	6.51E-27	BovineHD0600010809	BovineHD0600010824	39259257	39356766
BF_B26	ABABABBAABAAAABBBBAB	1638	0.01352	0.51	0.003769	BovineHD0600010809	BovineHD0600010824	39259257	39356766
BF_B9	ABAABAAABAAAABBAABAA	1637	-0.05839	6.80	7.51E-27	BovineHD4100004606	BovineHD4100004616	39261628	39358026
BF_B25	AAAAAAAABAAAABBAABBB	1637	0.02935	0.66	0.000971	BovineHD4100004606	BovineHD4100004616	39261628	39358026
BF_B12	BAABAAABAAAABBAABAAA	1631	-0.05798	6.71	1.98E-26	BovineHD0600010810	BovineHD4100004617	39262655	39359444
BF_B24	AAAAAAABAAAABBAABBBB	1631	0.02975	0.68	0.000889	BovineHD0600010810	BovineHD4100004617	39262655	39359444
BF_B5	AABAAABAAAABBAABAAAA	1638	-0.05838	6.81	6.59E-27	BovineHD4100004607	BovineHD4100004618	39271661	39362575
BF_B23	AAAAAABAAAABBAABBBBA	1638	0.03142	0.73	0.000521	BovineHD4100004607	BovineHD4100004618	39271661	39362575
BF_B1*	ABAAABAAAABBAABAAAAB	1501	-0.0625	8.06	3.25E-29	BovineHD4100004608	BovineHD0600010825	39273164	39363582
BF_B14	AAAAABAAAABBAABBBBAB	1501	0.03968	1.26	1.34E-05	BovineHD4100004608	BovineHD0600010825	39273164	39363582
BF_B6	BAAABAAAABBAABAAAABA	1638	-0.05838	6.81	6.59E-27	BTB-01709638	BovineHD4100004619	39277640	39364589
BF_B15	AAAABAAAABBAABBBBABA	1638	0.02918	0.91	0.000114	BTB-01709638	BovineHD4100004619	39277640	39364589
BF_B7	AAABAAAABBAABAAAABAB	1632	-0.05844	6.82	7.28E-27	BovineHD0600010812	BovineHD0600010826	39278657	39365496
BF_B21	AAABAAAABBAABBBBABAB	1632	0.02459	0.76	0.000408	BovineHD0600010812	BovineHD0600010826	39278657	39365496
BF_B10	AABAAAABBAABAAAABABB	1638	-0.05832	6.79	8.01E-27	BovineHD0600010813	BovineHD4100004620	39284507	39367782
BF_B20	AABAAAABBAABBBBABABA	1638	0.02306	0.80	0.000286	BovineHD0600010813	BovineHD4100004620	39284507	39367782
BF_B2	ABAAAABBAABAAAABABBB	1620	-0.0592	7.04	1.75E-27	BovineHD0600010815	BovineHD4100004621	39291640	39369803
BF_B17	ABAAAABBAABBBBABABAA	1620	0.02331	0.82	0.000263	BovineHD0600010815	BovineHD4100004621	39291640	39369803
BF_B8	BAAAABBAABAAAABABBBB	1638	-0.05847	6.81	6.76E-27	BovineHD0600010817	Hapmap33170-BTC-071249	39309298	39371150
BF_B18	BAAAABBAABBBBABABAAA	1638	0.02328	0.81	0.000275	BovineHD0600010817	Hapmap33170-BTC-071249	39309298	39371150
BF_B3	AAAABBAABAAAABABBBBB	1621	-0.0589	6.91	4.97E-27	ARS-BFGL-NGS-27686	BovineHD0600010827	39313672	39372172
BF_B19	AAAABBAABBBBABABAAAA	1621	0.02332	0.81	0.000284	ARS-BFGL-NGS-27686	BovineHD0600010827	39313672	39372172
BF_B11	AAABBAABAAAABABBBBBB	1638	-0.05792	6.70	1.81E-26	BovineHD0600010818	BovineHD4100004623	39319249	39382729
BF_B13**	BBBBBBBABBAABABBBBBB	1638	0.05071	1.31	3.53E-06	BovineHD0600010818	BovineHD4100004623	39319249	39382729

* most significant haplotype decreasing the phenotype

** most significant haplotype increasing the phenotype

Table 4.2c. Selected results from haplotype association analysis with marbling on BTA6.

[a]Beta statistic is an indicator of the direction of the effect

[b]Percent of variation explained by the haplotype

[c]Asymptotic p-value

[d]SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

[e]Base pair position from University of Maryland bovine genome assembly 3.0 for SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

HapID	Haplotype	NoAnm	Beta[a]	%Var[b]	P[c]	SNP_1[d]	SNP_20	Pos1[e]	Pos2
MS3	AAAABABBBBBBBABAAAAB	1637	-21.28	3.15	4.76E-13	BovineHD0600010824	BovineHD0600010835	39356766	39413342
MS11**	BBBABABAAAABBBABBBBA	1637	10.86	0.70	0.000681	BovineHD0600010824	BovineHD0600010835	39356766	39413342
MS5	AAABABBBBBBBABAAAABA	1638	-21.02	3.11	6.68E-13	BovineHD4100004616	BovineHD0600010836	39358026	39416405
MS14	BBABABAAAABBBABBBBAB	1638	10.64	0.67	0.000911	BovineHD4100004616	BovineHD0600010836	39358026	39416405
MS8	AABABBBBBBBABAAAABAB	1522	-21.31	3.21	1.87E-12	BovineHD4100004617	BovineHD0600010837	39359444	39418286
MS20	BABABAAAABBBABBBBABA	1522	9.644	0.57	0.003315	BovineHD4100004617	BovineHD0600010837	39359444	39418286
MS2	ABABBBBBBBABAAAABABA	1638	-21.6	3.20	2.98E-13	BovineHD4100004618	BovineHD0600010838	39362575	39426854
MS15	ABABAAAABBBABBBBABAB	1638	10.64	0.67	0.000911	BovineHD4100004618	BovineHD0600010838	39362575	39426854
MS1*	BABBBBBBBABAAAABABAA	1637	-21.6	3.25	2.09E-13	BovineHD0600010825	BovineHD0600010839	39363582	39430325
MS12	BABAAAABBBABBBBABABB	1637	10.83	0.70	0.000701	BovineHD0600010825	BovineHD0600010839	39363582	39430325
MS10	ABBBBBBBABAAAABABAAA	1638	-18.2	2.37	3.74E-10	BovineHD4100004619	BovineHD0600010840	39364589	39431268
MS16	ABAAAABBBABBBBABABBB	1638	10.64	0.67	0.000911	BovineHD4100004619	BovineHD0600010840	39364589	39431268
MS9	BBBBBBBABAAAABABAAAA	1521	-21.15	3.17	2.77E-12	BovineHD0600010826	BovineHD0600010841	39365496	39436297
MS19	BAAAABBBABBBBABABBBB	1521	9.71	0.57	0.003127	BovineHD0600010826	BovineHD0600010841	39365496	39436297
MS4	BBBBBBABAAAABABAAAAB	1637	-21.34	3.13	5.69E-13	BovineHD4100004620	BovineHD0600010842	39367782	39441548
MS18	AAAABBBABBBBABABBBBA	1637	10.61	0.67	0.000938	BovineHD4100004620	BovineHD0600010842	39367782	39441548
MS7	BBBBBABAAAABABAAAABA	1637	-20.9	3.01	1.52E-12	BovineHD4100004621	BovineHD0600010843	39369803	39445731
MS13	AAABBBABBBBABABBBBAB	1637	10.83	0.70	0.000701	BovineHD4100004621	BovineHD0600010843	39369803	39445731
MS6	BBBBABAAAABABAAAABAB	1638	-21.01	3.03	1.34E-12	Hapmap33170-BTC-071249	BTB-01326697	39371150	39461621
MS17	AABBBABBBBABABBBBABA	1638	10.64	0.67	0.000911	Hapmap33170-BTC-071249	BTB-01326697	39371150	39461621

* most significant haplotype decreasing the phenotype

** most significant haplotype increasing the phenotype

Table 4.2d. Selected results from haplotype association analysis with birth weight in region 1 on BTA6.

[a]Beta statistic is an indicator of the direction of the effect

[b]Percent of variation explained by the haplotype

[c]Asymptotic p-value

[d]SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

[e]Base pair position from University of Maryland bovine genome assembly 3.0 for SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

HapID	Haplotype	NoAnm	Beta[a]	%Var[b]	P[c]	SNP_1[d]	SNP_20	Pos1[e]	Pos2
BW1**	AABABBBBBBABAAABBBAB	1415	5.269	7.53	7.62E-26	BovineHD0600010750	Hapmap33630-BTC-041044	38818907	38914175
BW23	ABABAAAABABABABBAAAB	1415	-1.95	1.43	6.60E-06	BovineHD0600010750	Hapmap33630-BTC-041044	38818907	38914175
BW2**	ABABBBBBBABAAABBBABB	1415	5.269	7.53	7.62E-26	Hapmap27083-BTC-041166	BovineHD0600010761	38825835	38917456
BW24	BABAAAABABABABBAAABB	1415	-1.95	1.43	6.60E-06	Hapmap27083-BTC-041166	BovineHD0600010761	38825835	38917456
BW10	BABBBBBBABAAABBBABBB	1416	5.245	7.45	1.31E-25	BovineHD0600010751	BovineHD4100004583	38827869	38923533
BW13*	ABAAAABABABABBAAABBA	1416	-1.966	1.45	5.55E-06	BovineHD0600010751	BovineHD4100004583	38827869	38923533
BW8	ABBBBBBABAAABBBABBBA	1416	5.244	7.45	1.30E-25	BovineHD0600010752	BovineHD4100004585	38829248	38928941
BW19	BAAAABABABABBAAABBAA	1416	-1.953	1.43	6.34E-06	BovineHD0600010752	BovineHD4100004585	38829248	38928941
BW9	BBBBBBABAAABBBABBBAB	1416	5.244	7.45	1.30E-25	BovineHD4100004575	BovineHD0600034280	38830725	38931436
BW20	AAAABABABABBAAABBAAB	1416	-1.953	1.43	6.34E-06	BovineHD4100004575	BovineHD0600034280	38830725	38931436
BW3	BBBBBABAAABBBABBBABB	1416	5.255	7.49	9.78E-26	BovineHD4100004576	BovineHD0600010764	38834676	38950070
BW21	AAABABABABBAAABBAABA	1416	-1.953	1.43	6.34E-06	BovineHD4100004576	BovineHD0600010764	38834676	38950070
BW5	BBBBABAAABBBABBBABBB	1413	5.257	7.49	1.11E-25	BovineHD4100004577	BovineHD4100004586	38837159	38955154
BW15	AABABABABBAAABBAABAA	1413	-1.963	1.45	5.82E-06	BovineHD4100004577	BovineHD4100004586	38837159	38955154
BW6	BBBABAAABBBABBBABBBB	1413	5.257	7.49	1.11E-25	BovineHD4100004579	BovineHD0600010766	38841588	38962544
BW16	ABABABABBAAABBAABAAB	1413	-1.963	1.45	5.82E-06	BovineHD4100004579	BovineHD0600010766	38841588	38962544
BW4	BBABAAABBBABBBABBBBA	1416	5.255	7.49	9.78E-26	BovineHD0600010753	BovineHD0600010767	38849193	38967708
BW14*	BABABABBAAABBAABAABA	1416	-1.966	1.45	5.55E-06	BovineHD0600010753	BovineHD0600010767	38849193	38967708
BW7	BABAAABBBABBBABBBBAA	1413	5.257	7.49	1.11E-25	BovineHD4100004580	BovineHD0600010769	38852093	38974284
BW17	ABABABBAAABBAABAABAA	1413	-1.963	1.45	5.82E-06	BovineHD4100004580	BovineHD0600010769	38852093	38974284
BW12	ABAAABBBABBBABBBBAAA	1416	5.238	7.41	1.78E-25	BovineHD0600010755	BovineHD0600010770	38866381	38980227
BW22	BABABBAAABBAABAABAAA	1416	-1.953	1.43	6.34E-06	BovineHD0600010755	BovineHD0600010770	38866381	38980227
BW11	BAAABBBABBBABBBBAAAB	1413	5.267	7.45	1.42E-25	Hapmap31285-BTC-041097	BovineHD0600010771	38869785	38984478
BW18	ABABBAAABBAABAABAAAB	1413	-1.963	1.45	5.82E-06	Hapmap31285-BTC-041097	BovineHD0600010771	38869785	38984478

* most significant haplotype decreasing the phenotype

** most significant haplotype increasing the phenotype

Table 4.2e. Selected results from haplotype association analysis with birth weight in region 2 on BTA6.

[a]Beta statistic is an indicator of the direction of the effect

[b]Percent of variation explained by the haplotype

[c]Asymptotic p-value

[d]SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

[e]Base pair position from University of Maryland bovine genome assembly 3.0 for SNP in the first (SNP1) and last (SNP20) position of 20 SNP haplotype window

HapID	Haplotype	NoAnm	Beta[a]	%Var[b]	P[c]	SNP_1[d]	SNP_20	Pos1[e]	Pos2
BW_B1	BABAABAAABAAAABBAABA	1416	5.122	7.18	1.08E-24	BovineHD0600010809	BovineHD0600010824	39259257	39356766
BW_B13	ABABABBAABAAAABBAABB	1416	-4.715	1.08	9.25E-05	BovineHD0600010809	BovineHD0600010824	39259257	39356766
BW_B2	ABAABAAABAAAABBAABAA	1415	5.109	7.14	1.44E-24	BovineHD4100004606	BovineHD4100004616	39261628	39358026
BW_B14	BABABBAABAAAABBAABBB	1415	-4.547	1.02	0.000138	BovineHD4100004606	BovineHD4100004616	39261628	39358026
BW_B3	BAABAAABAAAABBAABAAA	1409	5.152	7.26	7.41E-25	BovineHD0600010810	BovineHD4100004617	39262655	39359444
BW_B15	ABABBAABAAAABBAABBBB	1409	-4.693	1.07	0.000100	BovineHD0600010810	BovineHD4100004617	39262655	39359444
BW_B4	AABAAABAAAABBAABAAAA	1416	5.097	7.12	1.68E-24	BovineHD4100004607	BovineHD4100004618	39271661	39362575
BW_B16	BABBAABAAAABBAABBBBA	1416	-4.735	1.05	0.000115	BovineHD4100004607	BovineHD4100004618	39271661	39362575
BW_B5**	ABAAABAAAABBAABAAAAB	1300	5.315	8.03	2.04E-25	BovineHD4100004608	BovineHD0600010825	39273164	39363582
BW_B17	ABBAABAAAABBAABBBBAB	1300	-4.661	1.11	0.000142	BovineHD4100004608	BovineHD0600010825	39273164	39363582
BW_B18	AAAAABAAAABBAABBBBAB	1300	-3.007	1.04	0.000230	BovineHD4100004608	BovineHD0600010825	39273164	39363582
BW_B6	BAAABAAAABBAABAAAABA	1416	5.116	7.17	1.13E-24	BTB-01709638	BovineHD4100004619	39277640	39364589
BW_B19	BBAABAAAABBAABBBBABA	1416	-4.735	1.05	0.000115	BTB-01709638	BovineHD4100004619	39277640	39364589
BW_B7	AAABAAAABBAABAAAABAB	1410	5.086	7.07	2.99E-24	BovineHD0600010812	BovineHD0600010826	39278657	39365496
BW_B20	BAABAAAABBAABBBBABAB	1410	-4.738	1.05	0.000116	BovineHD0600010812	BovineHD0600010826	39278657	39365496
BW_B8	AABAAAABBAABAAAABABB	1416	5.078	7.06	2.73E-24	BovineHD0600010813	BovineHD4100004620	39284507	39367782
BW_B21	AABAAAABBAABBBBABABA	1416	-2.446	1.26	2.30E-05	BovineHD0600010813	BovineHD4100004620	39284507	39367782
BW_B9	ABAAAABBAABAAAABABBB	1398	5.242	7.47	2.20E-25	BovineHD0600010815	BovineHD4100004621	39291640	39369803
BW_B22*	ABAAAABBAABBBBABABAA	1398	-2.548	1.35	1.29E-05	BovineHD0600010815	BovineHD4100004621	39291640	39369803
BW_B10	BAAAABBAABAAAABABBBB	1416	5.187	7.34	2.96E-25	BovineHD0600010817	Hapmap33170-BTC-071249	39309298	39371150
BW_B23	BAAAABBAABBBBABABAAA	1416	-2.53	1.33	1.36E-05	BovineHD0600010817	Hapmap33170-BTC-071249	39309298	39371150
BW_B11	AAAABBAABAAAABABBBBB	1400	5.21	7.42	3.02E-25	ARS-BFGL-NGS-27686	BovineHD0600010827	39313672	39372172
BW_B24	AAAABBAABBBBABABAAAA	1400	-2.492	1.30	1.96E-05	ARS-BFGL-NGS-27686	BovineHD0600010827	39313672	39372172
BW_B12	AAABBAABAAAABABBBBBB	1416	5.082	7.07	2.50E-24	BovineHD0600010818	BovineHD4100004623	39319249	39382729
BW_B25	BBBBBBBABBAABABBBBBB	1416	-3.793	1.03	0.000132	BovineHD0600010818	BovineHD4100004623	39319249	39382729

* most significant haplotype decreasing the phenotype

** most significant haplotype increasing the phenotype

Table 4.3. Summary of polymorphisms detected by direct sequencing that were consistent with haplotype status for BW1 haplotype on BTA6.

dbSNP_SS#	Gene	Polymorphism	Coding Status	Notes
ss539003528	NCAPG	T/C	coding	synonymous
ss539003529	NCAPG	G/T	not coding	
ss539003530	NCAPG	T/C	not coding	
ss539003531	NCAPG	T/G	not coding	
ss539003532	NCAPG	T/G	coding	synonymous
ss539003533	NCAPG	C/A	coding	non-synonymous (Ile-442-Met)
ss538968754	NCAPG	T/G	not coding	
ss538968724	NCAPG	T/C	not coding	
ss538968727	NCAPG	T/G	not coding	
ss538968730	NCAPG	A/G	not coding	
ss538968732	NCAPG	G/A	coding	synonymous
ss538968756	NCAPG	G/C	not coding	
ss538968758	NCAPG	delA	not coding	
rs133222819	NCAPG	T/C	coding	synonymous
rs110251642	NCAPG	C/A	coding	non-synonymous (Leu>Met)
ss538968736	NCAPG	G/C	not coding	
ss538968762	NCAPG	A/G	not coding	
ss538968738	NCAPG	G/A	not coding	
ss538968740	NCAPG	A/G	not coding	
ss538968764	NCAPG	C/A	not coding	
ss538968742	NCAPG	C/T	not coding	
ss538968744	NCAPG	C/A	not coding	
ss538968746	NCAPG	G/T	not coding	
ss538968766	NCAPG	C/T	not coding	
ss538968752	NCAPG	delAC	not coding	
ss538968748	NCAPG	C/T	not coding	
ss538968750	NCAPG	delAT	not coding	
ss538969084	NCAPG	T/C	not coding	
ss538969086	NCAPG	C/G	not coding	
ss539003508	LCORL	A/G	not coding	
ss539003509	LCORL	A/G	not coding	
ss539003510	LCORL	T/C	coding	exon 7 -3'UTR
ss539003511	LCORL	G/C	not coding	
ss539003512	LCORL	G/A	not coding	
ss539003514	LCORL	C/G	not coding	
ss539003513	LCORL	T/C	not coding	
ss539003515	LCORL	C/A	not coding	
ss539003516	LCORL	T/C	not coding	
ss539003517	LCORL	A/G	not coding	
ss539003518	LCORL	delAC	not coding	
ss539003519	LCORL	G/C	not coding	
ss539003520	LCORL	A/T	not coding	
ss539003521	LCORL	G/A	not coding	
ss539003522	LCORL	C/A	not coding	
ss539003523	LCORL	G/A	not coding	

References

1. Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y, Zhang L, Sodergren E, Havlak P, Worley KC, Weinstock GM, Gibbs RA: **Bos taurus genome assembly.** *BMC Genomics* 2009, **10:**180.
2. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marcais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL: **A whole-genome assembly of the domestic cow, Bos taurus.** *Genome Biol* 2009, **10:**R42.
3. Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA: **Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories.** *Genome Res* 2009, **19:**770-777.
4. Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA: **A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**18526-18531.
5. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP: **Development and characterization of a high density SNP genotyping assay for cattle.** *PLoS One* 2009, **4:**e5350.
6. Dekkers JC: **Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons.** *J Anim Sci* 2004, **82 E-Suppl:**E313-328.
7. Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES: **Association mapping: critical considerations shift from genotyping to experimental design.** *Plant Cell* 2009, **21:**2194-2202.
8. Han B, Kang HM, Eskin E: **Rapid and accurate multiple testing correction and power estimation for millions of correlated markers.** *PLoS Genet* 2009, **5:**e1000456.
9. Weller JI, Song JZ, Heyen DW, Lewin HA, Ron M: **A new approach to the problem of multiple comparisons in the genetic dissection of complex traits.** *Genetics* 1998, **150:**1699-1706.
10. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.** *J Roy Stat Soc B Met* 1995, **57:**289-300.
11. Spielman RS, McGinnis RE, Ewens WJ: **Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM).** *Am J Hum Genet* 1993, **52:**506-516.
12. Abecasis GR, Cardon LR, Cookson WO: **A general test of association for quantitative traits in nuclear families.** *Am J Hum Genet* 2000, **66:**279-292.
13. Abecasis GR, Cookson WO, Cardon LR: **Pedigree tests of transmission disequilibrium.** *Eur J Hum Genet* 2000, **8:**545-551.
14. Fulker DW, Cherny SS, Sham PC, Hewitt JK: **Combined linkage and association sib-pair analysis for quantitative traits.** *Am J Hum Genet* 1999, **64:**259-267.
15. Purcell S, Sham P, Daly MJ: **Parental phenotypes in family-based association analysis.** *Am J Hum Genet* 2005, **76:**249-259.
16. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155:**945-959.
17. Wu C, DeWan A, Hoh J, Wang Z: **A comparison of association methods correcting for population stratification in case-control studies.** *Ann Hum Genet* 2011, **75:**418-427.

18. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38:**904-909.
19. Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M: **An Arabidopsis example of association mapping in structured samples.** *PLoS Genet* 2007, **3:**e4.
20. Meuwissen TH, Karlsen A, Lien S, Olsaker I, Goddard ME: **Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping.** *Genetics* 2002, **161:**373-379.
21. Gautier M, Barcelona RR, Fritz S, Grohs C, Druet T, Boichard D, Eggen A, Meuwissen TH: **Fine mapping and physical characterization of two linked quantitative trait loci affecting milk fat yield in dairy cattle on BTA26.** *Genetics* 2006, **172:**425-436.
22. Olsen HG, Meuwissen TH, Nilsen H, Svendsen M, Lien S: **Fine mapping of quantitative trait Loci on bovine chromosome 6 affecting calving difficulty.** *J Dairy Sci* 2008, **91:**4312-4322.
23. Uleberg E, Wideroe IS, Grindflek E, Szyda J, Lien S, Meuwissen TH: **Fine mapping of a QTL for intramuscular fat on porcine chromosome 6 using combined linkage and linkage disequilibrium mapping.** *J Anim Breed Genet* 2005, **122:**1-6.
24. Sahana G, Guldbrandtsen B, Janss L, Lund MS: **Comparison of association mapping methods in a complex pedigreed population.** *Genet Epidemiol* 2010, **34:**455-462.
25. Madsen P: **A User's Guide to DMU.** In *Book A User's Guide to DMU* (Editor ed.^eds.), vol. Version 6, release 5.0. City: University of Aarhus Faculty Agricultural Sciences; 2010.
26. Schwenger B, Schober S, Simon D: **DUMPS cattle carry a point mutation in the uridine monophosphate synthase gene.** *Genomics* 1993, **16:**241-244.
27. Kunieda T, Nakagiri M, Takami M, Ide H, Ogawa H: **Cloning of bovine LYST gene and identification of a missense mutation associated with Chediak-Higashi syndrome of cattle.** *Mamm Genome* 1999, **10:**1146-1149.
28. Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick BW, Ryan AM, Clayton D, Li L, Neibergs HL, Zhang N, et al.: **A genetic linkage map of the bovine genome.** *Nat Genet* 1994, **6:**227-235.
29. Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SL, Hawkins GA, Toldo SS, Fries R, Grosz MD, Yoo J, et al.: **A genetic linkage map for cattle.** *Genetics* 1994, **136:**619-639.
30. Ma RZ, Beever JE, Da Y, Green CA, Russ I, Park C, Heyen DW, Everts RE, Fisher SR, Overton KM, Teale AJ, Kemp SJ, Hines HC, Guerin G, Lewin HA: **A male linkage map of the cattle (Bos taurus) genome.** *J Hered* 1996, **87:**261-271.
31. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, et al.: **Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing.** *Genetics* 1995, **139:**907-920.
32. Barendse W, Vaiman D, Kemp SJ, Sugimoto Y, Armitage SM, Williams JL, Sun HS, Eggen A, Agaba M, Aleyasin SA, Band M, Bishop MD, Buitkamp J, Byrne K, Collins F, Cooper L, Coppettiers W, Denys B, Drinkwater RD, Easterday K, Elduque C, Ennis S, Erhardt G, Li L, et al.: **A medium-density genetic linkage map of the bovine genome.** *Mamm Genome* 1997, **8:**21-28.
33. Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS, Smith TP, Lopez-Corrales NL, Beattie CW: **A second-generation linkage map of the bovine genome.** *Genome Res* 1997, **7:**235-249.

34. Yoneda K, Moritomo Y, Takami M, Hirata S, Kikukawa Y, Kunieda T: **Localization of a locus responsible for the bovine chondrodysplastic dwarfism (bcd) on chromosome 6.** *Mamm Genome* 1999, **10:**597-600.
35. Ohba Y, Kitagawa H, Kitoh K, Asahina S, Nishimori K, Yoneda K, Kunieda T, Sasaki Y: **Homozygosity mapping of the locus responsible for renal tubular dysplasia of cattle on bovine chromosome 1.** *Mamm Genome* 2000, **11:**316-319.
36. Kobayashi N, Hirano T, Maruyama S, Matsuno H, Mukoujima K, Morimoto H, Noike H, Tomimatsu H, Hara K, Itoh T, Imakawa K, Nakayama H, Nakamaru T, Sugimoto Y: **Genetic mapping of a locus associated with bovine chronic interstitial nephritis to chromosome 1.** *Anim Genet* 2000, **31:**91-95.
37. Ohba Y, Kitagawa H, Kitoh K, Sasaki Y, Takami M, Shinkai Y, Kunieda T: **A deletion of the paracellin-1 gene is responsible for renal tubular dysplasia in cattle.** *Genomics* 2000, **68:**229-236.
38. Takeda H, Takami M, Oguni T, Tsuji T, Yoneda K, Sato H, Ihara N, Itoh T, Kata SR, Mishina Y, Womack JE, Moritomo Y, Sugimoto Y, Kunieda T: **Positional cloning of the gene LIMBIN responsible for bovine chondrodysplastic dwarfism.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99:**10549-10554.
39. Hirano T, Kobayashi N, Itoh T, Takasuga A, Nakamaru T, Hirotsune S, Sugimoto Y: **Null mutation of PCLN-1/Claudin-16 results in bovine chronic interstitial nephritis.** *Genome Res* 2000, **10:**659-663.
40. Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet JC, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ, Georges M: **Highly effective SNP-based association mapping and management of recessive defects in livestock.** *Nat Genet* 2008, **40:**449-454.
41. Fasquelle C, Sartelet A, Li W, Dive M, Tamma N, Michaux C, Druet T, Huijbers IJ, Isacke CM, Coppieters W, Georges M, Charlier C: **Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle.** *PLoS Genet* 2009, **5:**e1000666.
42. Meyers SN, McDaneld TG, Swist SL, Marron BM, Steffen DJ, O'Toole D, O'Connell JR, Beever JE, Sonstegard TS, Smith TP: **A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle.** *BMC Genomics* 2010, **11:**337.
43. Drogemuller C, Reichart U, Seuberlich T, Oevermann A, Baumgartner M, Kuhni Boghenbor K, Stoffel MH, Syring C, Meylan M, Muller S, Muller M, Gredler B, Solkner J, Leeb T: **An Unusual Splice Defect in the Mitofusin 2 Gene (MFN2) Is Associated with Degenerative Axonopathy in Tyrolean Grey Cattle.** *PLos One* 2011, **6:**e18931.
44. Salmon Hillbertz NH, Isaksson M, Karlsson EK, Hellmen E, Pielberg GR, Savolainen P, Wade CM, von Euler H, Gustafson U, Hedhammar A, Nilsson M, Lindblad-Toh K, Andersson L, Andersson G: **Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs.** *Nat Genet* 2007, **39:**1318-1320.
45. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ, 3rd, Comstock KE, Keller ET, Mesirov JP, von Euler H, Kampe O, Hedhammar A, Lander ES, Andersson G, Andersson L, Lindblad-Toh K: **Efficient mapping of mendelian traits in dogs through genome-wide association.** *Nat Genet* 2007, **39:**1321-1328.
46. Georges M, Dietz AB, Mishra A, Nielsen D, Sargeant LS, Sorensen A, Steele MR, Zhao X, Leipold H, Womack JE, et al.: **Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative**

trait locus. *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90:**1058-1062.
47. Coppieters W, Riquet J, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M: **A QTL with major effect on milk yield and composition maps to bovine chromosome 14.** *Mamm Genome* 1998, **9:**540-544.
48. Stone RT, Keele JW, Shackelford SD, Kappes SM, Koohmaraie M: **A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits.** *J Anim Sci* 1999, **77:**1379-1384.
49. Casas E, Shackelford SD, Keele JW, Stone RT, Kappes SM, Koohmaraie M: **Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin.** *J Anim Sci* 2000, **78:**560-569.
50. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M, Snell R: **Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition.** *Genome Res* 2002, **12:**222-231.
51. Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JA: **A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map.** *J Dairy Sci* 2008, **91:**3225-3236.
52. Kim ES, Berger PJ, Kirkpatrick BW: **Genome-wide scan for bovine twinning rate QTL using linkage disequilibrium.** *Anim Genet* 2009, **40:**300-307.
53. Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q: **Genome wide association studies for milk production traits in Chinese Holstein population.** *PLoS One* 2010, **5:**e13661.
54. Mai MD, Sahana G, Christiansen FB, Guldbrandtsen B: **A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip.** *J Anim Sci* 2010, **88:**3522-3528.
55. Sahana G, Guldbrandtsen B, Bendixen C, Lund MS: **Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle.** *Anim Genet* 2010, **41:**579-588.
56. Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TP, Sonstegard TS, Thallman RM, Bennett GL: **Genome-wide association study of growth in crossbred beef cattle.** *J Anim Sci* 2010, **88:**837-848.
57. Hirschhorn JN, Daly MJ: **Genome-wide association studies for common diseases and complex traits.** *Nat Rev Genet* 2005, **6:**95-108.
58. Grisart B, Farnir F, Karim L, Cambisano N, Kim JJ, Kvasz A, Mni M, Simon P, Frere JM, Coppieters W, Georges M: **Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101:**2398-2403.
59. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L: **A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig.** *Nature* 2003, **425:**832-836.
60. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F, Larzul C, Laville E, Meish F, Milenkovic D, Tobin J, Charlier C, Georges M: **A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep.** *Nat Genet* 2006, **38:**813-818.

61. Takeda H, Charlier C, Farnir F, Georges M: **Demonstrating polymorphic miRNA-mediated gene regulation in vivo: application to the g+6223G->A mutation of Texel sheep.** *Rna* 2010, **16:**1854-1863.
62. Heyen DW, Weller JI, Ron M, Band M, Beever JE, Feldmesser E, Da Y, Wiggans GR, VanRaden PM, Lewin HA: **A genome scan for QTL influencing milk production and health traits in dairy cattle.** *Physiol Genomics* 1999, **1:**165-175.
63. Looft C, Reinsch N, Karall-Albrecht C, Paul S, Brink M, Thomsen H, Brockmann G, Kuhn C, Schwerin M, Kalm E: **A mammary gland EST showing linkage disequilibrium to a milk production QTL on bovine Chromosome 14.** *Mamm Genome* 2001, **12:**646-650.
64. Farnir F, Grisart B, Coppieters W, Riquet J, Berzi P, Cambisano N, Karim L, Mni M, Moisio S, Simon P, Wagenaar D, Vilkki J, Georges M: **Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14.** *Genetics* 2002, **161:**275-287.
65. Cases S, Smith SJ, Zheng YW, Myers HM, Lear SR, Sande E, Novak S, Collins C, Welch CB, Lusis AJ, Erickson SK, Farese RV, Jr.: **Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95:**13018-13023.
66. Smith SJ, Cases S, Jensen DR, Chen HC, Sande E, Tow B, Sanan DA, Raber J, Eckel RH, Farese RV, Jr.: **Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat.** *Nat Genet* 2000, **25:**87-90.
67. Andersson-Eklund L, Marklund L, Lundstrom K, Haley CS, Andersson K, Hansson I, Moller M, Andersson L: **Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross.** *J Anim Sci* 1998, **76:**694-700.
68. Nezer C, Moreau L, Brouwers B, Coppieters W, Detilleux J, Hanset R, Karim L, Kvasz A, Leroy P, Georges M: **An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs.** *Nat Genet* 1999, **21:**155-156.
69. Jeon JT, Carlborg O, Tornsten A, Giuffra E, Amarger V, Chardon P, Andersson-Eklund L, Andersson K, Hansson I, Lundstrom K, Andersson L: **A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus.** *Nat Genet* 1999, **21:**157-158.
70. Excoffier L, Laval G, Schneider S: **Arlequin (version 3.0): an integrated software package for population genetics data analysis.** *Evol Bioinform Online* 2005, **1:**47-50.
71. Tregouet DA, Escolano S, Tiret L, Mallet A, Golmard JL: **A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm.** *Ann Hum Genet* 2004, **68:**165-177.
72. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** *Am J Hum Genet* 2001, **68:**978-989.
73. Scheet P, Stephens M: **A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase.** *Am J Hum Genet* 2006, **78:**629-644.
74. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: **MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes.** *Genet Epidemiol* 2010, **34:**816-834.
75. Howie BN, Donnelly P, Marchini J: **A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.** *PLoS Genet* 2009, **5:**e1000529.

76. Browning SR, Browning BL: **Haplotype phasing: existing methods and new developments.** *Nat Rev Genet* 2011, **12:**703-714.
77. Browning SR, Browning BL: **Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering.** *Am J Hum Genet* 2007, **81:**1084-1097.
78. Marchini J, Howie B: **Genotype imputation for genome-wide association studies.** *Nat Rev Genet* 2010, **11:**499-511.
79. Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A: **A comprehensive evaluation of SNP genotype imputation.** *Hum Genet* 2009, **125:**163-171.
80. Jayasekara MU, Leipold HW, Cook JE: **Pathological changes in congenital hypotrichosis in Hereford cattle.** *Zentralblatt fur VeterinarmedizinReihe A* 1979, **26:**744-753.
81. Hutt FB: **A Note on Six Kinds of Genetic Hypotrichosis in Cattle.** *J Hered* 1963, **54:**186-187.
82. Wipprecht C, Horlacher WR: **A Lethal Gene in Jersey Cattle.** *J Hered* 1935, **26:**363-368.
83. HuttT FB, Saunders LZ: **Viable Genetic Hypotrichosis in Guernsey Cattle.** *J Hered* 1953, **44:**97-103.
84. Olson TA, Harrove DD, Leipold HW: **Occurrence of Hypotrichosis in Polled Hereford Cattle.** *The Bovine Practitioner* 1985, **20:**4-8.
85. Shand A, Young GB: **A Note on Congenital Alopeica in a Fresian Herd.** *The Veterinary Record* 1964, **76:**907-909.
86. Steffen DJ, Leipold HW, Gibb J, Smith JE: **Congenital anemia, dyskeratosis, and progressive alopecia in Polled Hereford calves.** *Veterinary pathology* 1991, **28:**234-240.
87. Craft WA, Blizzard WL: **The Inheritance of Semi-Hairlessness in Cattle.** *J Hered* 1934, **25:**385-390.
88. Kidwell JF, Guilbert HR: **A recurrence of the semi-hairless gene in cattle.** *J Hered* 1950, **41:**190-192.
89. Schleger AV, Thompson BJ, Hewetson RW: **Histopathology of hypotrichosis in calves.** *Aust J Biol Sci* 1967, **20:**661-668.
90. Becker RB, Simpson CF, Wilcox CJ: **Hairless Guernsey Cattle: Hypotrichosis--A Non-lethal Character.** *J Hered* 1963, **54:**3-7.
91. Pirchner F, Grunberg W: **Alopecia in Pinzgau Cattle.** *Ann Genet Sel Anim* 1970, **2:**129-133.
92. Schleger AV: **Relationship between cyclic changes in the hair follicle and sweat gland size in cattle.** *Aust J Biol Sci* 1966, **19:**607-617.
93. Young JG: **A Note on the Occurrence of an Epithelial Defect in a Grade Hereford Herd.** *Aust Vet J* 1953, **29:**298-299.
94. Mohr OL, Wriedt C: **Hairless, A New Recessive Lethal in Cattle.** *J Genet* 1928, **19:**315-336.
95. Holmes JR, Young GB: **Symmetrical alopecia in cattle.** *Vet Rec* 1954, **66:**704.
96. Regan WM, Mead SW, Gregory PW: **AN IHERITED SKIN-DEFECT IN CATTLE: The Occurrence of a Sub-Lethal Epithelial Defect in a Jersey Herd, and a plan for Eliminating Lethal Genes.** *J Hered* 1935, **26:**357-362.
97. Bracho GA, Beeman K, Johnson JL, Leipold HW: **Further studies of congenital hypotrichosis in Hereford cattle.** *Zentralblatt fur VeterinarmedizinReihe A* 1984, **31:**72-80.
98. Sorensen AC, Sorensen MK, Berg P: **Inbreeding in Danish dairy cattle breeds.** *J Dairy Sci* 2005, **88:**1865-1872.
99. Zenger KR, Khatkar MS, Cavanagh JA, Hawken RJ, Raadsma HW: **Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into**

Australian and global population variability, including impact of selection. *Anim Genet* 2007, **38:**7-14.
100. Shuster DE, Kehrli ME, Jr., Ackermann MR, Gilbert RO: **Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle.** *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89:**9225-9229.
101. Thomsen B, Horn P, Panitz F, Bendixen E, Petersen AH, Holm LE, Nielsen VH, Agerholm JS, Arnbjerg J, Bendixen C: **A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation.** *Genome Res* 2006, **16:**97-105.
102. Schutz E, Scharfenstein M, Brenig B: **Implication of complex vertebral malformation and bovine leukocyte adhesion deficiency DNA-based testing on disease frequency in the Holstein population.** *J Dairy Sci* 2008, **91:**4854-4859.
103. Nagahata H: **Bovine leukocyte adhesion deficiency (BLAD): a review.** *J Vet Med Sci* 2004, **66:**1475-1482.
104. Meyers S, McDaneld T, Swist S, Marron B, Steffen D, O'Toole D, O'Connell J, Beever J, Sonstegard T, Smith TP: **A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle.** *BMC Genomics* 2010, **11:**337.
105. Karlsson EK, Baranowska I, Wade C, Salmon Hillbertz N, Zody M, Anderson N, Biagi T, Patterson N, Pielberg G, Kulbokas E, Comstock K, Keller E, Mesirov J, von Euler H, Kampe O, Hedhammar A, Lander E, Andersson G, Andersson L, Lindblad-Toh K: **Efficient mapping of mendelian traits in dogs through genome-wide association.** *Nature genetics* 2007, **39:**1321-1328.
106. Andersson L: **Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci.**
107. Rose R, Smith JE, Leipold HW: **Role of arginine-converting-enzyme in hypotrichosis of Hereford cattle.** *Zentralblatt fur VeterinarmedizinReihe A* 1983, **30:**369-372.
108. Rose R, Smith JE, Leipold HW: **Increased solubility of hair from hypotrichotic Herefords.** *Zentralbl Veterinarmed A* 1983, **30:**363-368.
109. Ogawa H, Goldsmith LA: **Human epidermal transglutaminase. Preparation and properties.** *J Biol Chem* 1976, **251:**7281-7288.
110. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81:**559-575.
111. Antonarakis SE, Krawczak M, Cooper DN: **Disease-causing mutations in the human genome.** *Eur J Pediatr* 2000, **159 Suppl 3:**S173-178.
112. Arin MJ: **The molecular basis of human keratin disorders.** *Hum Genet* 2009, **125:**355-373.
113. Langbein L, Rogers MA, Praetzel S, Winter H, Schweizer J: **K6irs1, K6irs2, K6irs3, and K6irs4 represent the inner-root-sheath-specific type II epithelial keratins of the human hair follicle.** *J Invest Dermatol* 2003, **120:**512-522.
114. Aoki N, Sawada S, Rogers MA, Schweizer J, Shimomura Y, Tsujimoto T, Ito K, Ito M: **A novel type II cytokeratin, mK6irs, is expressed in the Huxley and Henle layers of the mouse inner root sheath.** *J Invest Dermatol* 2001, **116:**359-365.
115. Porter RM, Gandhi M, Wilson NJ, Wood P, McLean WH, Lane EB: **Functional analysis of keratin components in the mouse hair follicle inner root sheath.** *Br J Dermatol* 2004, **150:**195-204.
116. Lyne AG, Heideman MJ: **The Pre-Natal Development of Skin and Hair in Cattle.** *Aust J Biol Sci* 1959, **12:**72-95.

117. Langbein L, Rogers MA, Praetzel S, Aoki N, Winter H, Schweizer J: **A novel epithelial keratin, hK6irs1, is expressed differentially in all layers of the inner root sheath, including specialized huxley cells (Flugelzellen) of the human hair follicle.** *J Invest Dermatol* 2002, **118:**789-799.
118. Langbein L, Schweizer J: **Keratins of the human hair follicle.** *Int Rev Cytol* 2005, **243:**1-78.
119. Ito M: **Biologic roles of the innermost cell layer of the outer root sheath in human anagen hair follicle: further electron microscopic study.** *Arch Dermatol Res* 1989, **281:**254-259.
120. Winter H, Langbein L, Praetzel S, Jacobs M, Rogers MA, Leigh IM, Tidman N, Schweizer J: **A novel human type II cytokeratin, K6hf, specifically expressed in the companion layer of the hair follicle.** *J Invest Dermatol* 1998, **111:**955-962.
121. Paus R, Cotsarelis G: **The biology of hair follicles.** *N Engl J Med* 1999, **341:**491-497.
122. Schneider MR, Schmidt-Ullrich R, Paus R: **The hair follicle as a dynamic miniorgan.** *Curr Biol* 2009, **19:**R132-142.
123. Moll R, Divo M, Langbein L: **The human keratins: biology and pathology.** *Histochem Cell Biol* 2008, **129:**705-733.
124. Parry DA, Steinert PM: **Intermediate filaments: molecular architecture, assembly, dynamics and polymorphism.** *Quarterly reviews of biophysics* 1999, **32:**99-187.
125. Bragulla HH, Homberger DG: **Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia.** *J Anat* 2009, **214:**516-559.
126. Steinert PM: **Keratins: dynamic, flexible structural proteins of epithelial cells.** *Curr Probl Dermatol* 2001, **13:**193-198.
127. Steinert PM: **Structure, function, and dynamics of keratin intermediate filaments.** *J Invest Dermatol* 1993, **100:**729-734.
128. O'Guin WM, Sun TT, Manabe M: **Interaction of trichohyalin with intermediate filaments: three immunologically defined stages of trichohyalin maturation.** *J Invest Dermatol* 1992, **98:**24-32.
129. Rothnagel JA, Rogers GE: **Trichohyalin, an intermediate filament-associated protein of the hair follicle.** *J Cell Biol* 1986, **102:**1419-1429.
130. Bornslaeger EA, Corcoran CM, Stappenbeck TS, Green KJ: **Breaking the connection: displacement of the desmosomal plaque protein desmoplakin from cell-cell interfaces disrupts anchorage of intermediate filament bundles and alters intercellular junction assembly.** *J Cell Biol* 1996, **134:**985-1001.
131. Schweizer J, Bowden PE, Coulombe PA, Langbein L, Lane EB, Magin TM, Maltais L, Omary MB, Parry DA, Rogers MA, Wright MW: **New consensus nomenclature for mammalian keratins.** *J Cell Biol* 2006, **174:**169-174.
132. Cooper D, Sun TT: **Monoclonal antibody analysis of bovine epithelial keratins. Specific pairs as defined by coexpression.** *J Biol Chem* 1986, **261:**4646-4654.
133. Lu X, Lane EB: **Retrovirus-mediated transgenic keratin expression in cultured fibroblasts: specific domain functions in keratin stabilization and filament formation.** *Cell* 1990, **62:**681-696.
134. Peters T, Sedlmeier R, Bussow H, Runkel F, Luers GH, Korthaus D, Fuchs H, Hrabe de Angelis M, Stumm G, Russ AP, Porter RM, Augustin M, Franz T: **Alopecia in a novel mouse model RCO3 is caused by mK6irs1 deficiency.** *J Invest Dermatol* 2003, **121:**674-680.
135. Yiasemides E, Trisnowati N, Su J, Dang N, Klingberg S, Marr P, Melbourne W, Tran K, Chow CW, Orchard D, Varigos G, Murrell DF: **Clinical heterogeneity in recessive**

epidermolysis bullosa due to mutations in the keratin 14 gene, KRT14. *Clin Exp Dermatol* 2008, **33:**689-697.
136. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, Vonholdt BM, Rhue A, Boyko A, Byers A, Wong A, Mosher DS, Elkahloun AG, Spady TC, Andre C, Lark KG, Cargill M, Bustamante CD, Wayne RK, Ostrander EA: **Coat variation in the domestic dog is governed by variants in three genes.** *Science* 2009, **326:**150-153.
137. Kuramoto T, Hirano R, Kuwamura M, Serikawa T: **Identification of the rat Rex mutation as a 7-bp deletion at splicing acceptor site of the Krt71 gene.** *The Journal of veterinary medical science / the Japanese Society of Veterinary Science* 2010, **72:**909-912.
138. Gandolfi B, Outerbridge CA, Beresford LG, Myers JA, Pimentel M, Alhaddad H, Grahn JC, Grahn RA, Lyons LA: **The naked truth: Sphynx and Devon Rex cat breed mutations in KRT71.** *Mamm Genome* 2010, **21:**509-515.
139. Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT: **The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics.** *Nucleic acids research* 2011, **39:**D842-848.
140. Miller SA, Dykes DD, Polesky HF: **A simple salting out procedure for extracting DNA from human nucleated cells.** *Nucleic Acids Res* 1988, **16:**1215.
141. **SSRIT - Simple Sequence Repeat Identification Tool.** In *Book SSRIT - Simple Sequence Repeat Identification Tool* (Editor ed.^eds.). City.
142. Smit AFA, Hubley R, Green P: **RepeatMasker.** In *Book RepeatMasker* (Editor ed.^eds.). City.
143. Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF: **M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods.** *BioTechniques* 2001, **31:**24-26, 28.
144. Wheelan SJ, Church DM, Ostell JM: **Spidey: a tool for mRNA-to-genomic alignments.** *Genome Res* 2001, **11:**1952-1957.
145. Hoglund JK, Guldbrandtsen B, Lund MS, Sahana G: **Analyzes of genome-wide association follow-up study for calving traits in dairy cattle.** *BMC Genet* 2012, **13:**71.
146. Sahana G, Guldbrandtsen B, Lund MS: **Genome-wide association study for calving traits in Danish and Swedish Holstein cattle.** *J Dairy Sci* 2011, **94:**479-486.
147. Pausch H, Flisikowski K, Jung S, Emmerling R, Edel C, Gotz KU, Fries R: **Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle.** *Genetics* 2011, **187:**289-297.
148. McClure MC, Ramey HR, Rolf MM, McKay SD, Decker JE, Chapple RH, Kim JW, Taxis TM, Weaber RL, Schnabel RD, Taylor JF: **Genome-wide association analysis for quantitative trait loci influencing Warner-Bratzler shear force in five taurine cattle breeds.** *Anim Genet* 2012, **43:**662-673.
149. Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, Sugimoto Y, Takasuga A: **Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle.** *BMC Genet* 2012, **13:**40.
150. Takasuga A, Watanabe T, Mizoguchi Y, Hirano T, Ihara N, Takano A, Yokouchi K, Fujikawa A, Chiba K, Kobayashi N, Tatsuda K, Oe T, Furukawa-Kuroiwa M, Nishimura-Abe A, Fujita T, Inoue K, Mizoshita K, Ogino A, Sugimoto Y: **Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping.** *Mamm Genome* 2007, **18:**125-136.

151. Mizoshita K, Takano A, Watanabe T, Takasuga A, Sugimoto Y: **Identification of a 1.1-Mb region for a carcass weight QTL onbovine Chromosome 14.** *Mamm Genome* 2005, **16:**532-537.
152. Setoguchi K, Furuta M, Hirano T, Nagao T, Watanabe T, Sugimoto Y, Takasuga A: **Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate.** *BMC Genet* 2009, **10:**43.
153. Davidson AJ, Freeman SA, Crosier KE, Wood CR, Crosier PS: **Expression of murine interleukin 11 and its receptor alpha-chain in adult and embryonic tissues.** *Stem Cells* 1997, **15:**119-124.
154. Nieminen P, Morgan NV, Fenwick AL, Parmanen S, Veistinen L, Mikkola ML, van der Spek PJ, Giraud A, Judd L, Arte S, Brueton LA, Wall SA, Mathijssen IM, Maher ER, Wilkie AO, Kreiborg S, Thesleff I: **Inactivation of IL11 signaling causes craniosynostosis, delayed tooth eruption, and supernumerary teeth.** *Am J Hum Genet* 2011, **89:**67-81.
155. Wilsker D, Patsialou A, Dallas PB, Moran E: **ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development.** *Cell Growth Differ* 2002, **13:**95-106.
156. Tidwell JA, Schmidt C, Heaton P, Wilson V, Tucker PW: **Characterization of a new ARID family transcription factor (Brightlike/ARID3C) that co-activates Bright/ARID3A-mediated immunoglobulin gene transcription.** *Mol Immunol* 2011, **49:**260-272.
157. Webb CF, Bryant J, Popowski M, Allred L, Kim D, Harriss J, Schmidt C, Miner CA, Rose K, Cheng HL, Griffin C, Tucker PW: **The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development.** *Mol Cell Biol* 2011, **31:**1041-1053.
158. Burnicka-Turek O, Kata A, Buyandelger B, Ebermann L, Kramann N, Burfeind P, Hoyer-Fender S, Engel W, Adham IM: **Pelota interacts with HAX1, EIF3G and SRPX and the resulting protein complexes are associated with the actin cytoskeleton.** *BMC Cell Biol* 2010, **11:**28.
159. Fontaine JM, Sun X, Benndorf R, Welsh MJ: **Interactions of HSP22 (HSPB8) with HSP20, alphaB-crystallin, and HSPB3.** *Biochem Biophys Res Commun* 2005, **337:**1006-1011.
160. Sainz N, Rodriguez A, Catalan V, Becerril S, Ramirez B, Gomez-Ambrosi J, Fruhbeck G: **Leptin administration downregulates the increased expression levels of genes related to oxidative stress and inflammation in the skeletal muscle of ob/ob mice.** *Mediators Inflamm* 2010, **2010:**784343.
161. Lavoie JN, Hickey E, Weber LA, Landry J: **Modulation of actin microfilament dynamics and fluid phase pinocytosis by phosphorylation of heat shock protein 27.** *J Biol Chem* 1993, **268:**24210-24214.
162. Iwaki T, Iwaki A, Tateishi J, Goldman JE: **Sense and antisense modification of glial alpha B-crystallin production results in alterations of stress fiber formation and thermoresistance.** *J Cell Biol* 1994, **125:**1385-1393.
163. Mehlen P, Preville X, Chareyron P, Briolay J, Klemenz R, Arrigo AP: **Constitutive expression of human hsp27, Drosophila hsp27, or human alpha B-crystallin confers resistance to TNF- and oxidative stress-induced cytotoxicity in stably transfected murine L929 fibroblasts.** *J Immunol* 1995, **154:**363-374.
164. Sugiyama Y, Suzuki A, Kishikawa M, Akutsu R, Hirose T, Waye MM, Tsui SK, Yoshida S, Ohno S: **Muscle develops a specific form of small heat shock protein complex composed of MKBP/HSPB2 and HSPB3 during myogenic differentiation.** *J Biol Chem* 2000, **275:**1095-1104.
165. Raju R, Balakrishnan L, Nanjappa V, Bhattacharjee M, Getnet D, Muthusamy B, Kurian Thomas J, Sharma J, Rahiman BA, Harsha HC, Shankar S, Prasad TS, Mohan

SS, Bader GD, Wani MR, Pandey A: **A comprehensive manually curated reaction map of RANKL/RANK-signaling pathway.** *Database (Oxford)* 2011, **2011:**bar021.
166. Sekiya F, Poulin B, Kim YJ, Rhee SG: **Mechanism of tyrosine phosphorylation and activation of phospholipase C-gamma 1. Tyrosine 783 phosphorylation is not sufficient for lipase activation.** *J Biol Chem* 2004, **279:**32181-32190.
167. Avery CL, He Q, North KE, Ambite JL, Boerwinkle E, Fornage M, Hindorff LA, Kooperberg C, Meigs JB, Pankow JS, Pendergrass SA, Psaty BM, Ritchie MD, Rotter JI, Taylor KD, Wilkens LR, Heiss G, Lin DY: **A phenomics-based strategy identifies loci on APOC1, BRAP, and PLCG1 associated with metabolic syndrome phenotype domains.** *PLoS Genet* 2011, **7:**e1002322.
168. Liu YJ, Guo YF, Zhang LS, Pei YF, Yu N, Yu P, Papasian CJ, Deng HW: **Biological pathway-based genome-wide association analysis identified the vasoactive intestinal peptide (VIP) pathway important for obesity.** *Obesity (Silver Spring)* 2010, **18:**2339-2346.
169. Manmontri B, Sariahmetoglu M, Donkor J, Bou Khalil M, Sundaram M, Yao Z, Reue K, Lehner R, Brindley DN: **Glucocorticoids and cyclic AMP selectively increase hepatic lipin-1 expression, and insulin acts antagonistically.** *J Lipid Res* 2008, **49:**1056-1067.
170. Reue K, Xu P, Wang XP, Slavin BG: **Adipose tissue deficiency, glucose intolerance, and increased atherosclerosis result from mutation in the mouse fatty liver dystrophy (fld) gene.** *J Lipid Res* 2000, **41:**1067-1076.
171. Blott S, Kim JJ, Moisio S, Schmidt-Kuntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D, Karim L, Simon P, Snell R, Spelman R, Wong J, Vilkki J, Georges M, Farnir F, Coppieters W: **Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition.** *Genetics* 2003, **163:**253-266.
172. White SN, Casas E, Allan MF, Keele JW, Snelling WM, Wheeler TL, Shackelford SD, Koohmaraie M, Smith TP: **Evaluation in beef cattle of six deoxyribonucleic acid markers developed for dairy traits reveals an osteopontin polymorphism associated with postweaning growth.** *J Anim Sci* 2007, **85:**1-10.
173. Reindl KM, Sheridan MA: **Peripheral regulation of the growth hormone-insulin-like growth factor system in fish and other vertebrates.** *Comp Biochem Physiol A Mol Integr Physiol* 2012, **163:**231-245.
174. Moody TW, Ito T, Osefo N, Jensen RT: **VIP and PACAP: recent insights into their functions/roles in physiology and disease from molecular and genetic studies.** *Curr Opin Endocrinol Diabetes Obes* 2011, **18:**61-67.
175. Gray SL, Cummings KJ, Jirik FR, Sherwood NM: **Targeted disruption of the pituitary adenylate cyclase-activating polypeptide gene results in early postnatal death associated with dysfunction of lipid and carbohydrate metabolism.** *Mol Endocrinol* 2001, **15:**1739-1747.
176. Jamen F, Persson K, Bertrand G, Rodriguez-Henche N, Puech R, Bockaert J, Ahren B, Brabet P: **PAC1 receptor-deficient mice display impaired insulinotropic response to glucose and reduced glucose tolerance.** *J Clin Invest* 2000, **105:**1307-1315.
177. Adams BA, Gray SL, Isaac ER, Bianco AC, Vidal-Puig AJ, Sherwood NM: **Feeding and metabolism in mice lacking pituitary adenylate cyclase-activating polypeptide.** *Endocrinol* 2008, **149:**1571-1580.
178. Winzell MS, Ahren B: **Role of VIP and PACAP in islet function.** *Peptides* 2007, **28:**1805-1813.
179. Asnicar MA, Koster A, Heiman ML, Tinsley F, Smith DP, Galbreath E, Fox N, Ma YL, Blum WF, Hsiung HM: **Vasoactive intestinal polypeptide/pituitary adenylate**

cyclase-activating peptide receptor 2 deficiency in mice results in growth retardation and increased basal metabolic rate. *Endocrinol* 2002, **143:**3994-4006.

180. Tomimoto S, Ojika T, Shintani N, Hashimoto H, Hamagami K, Ikeda K, Nakata M, Yada T, Sakurai Y, Shimada T, Morita Y, Ishida C, Baba A: **Markedly reduced white adipose tissue and increased insulin sensitivity in adcyap1-deficient mice.** *J Pharmacol Sci* 2008, **107:**41-48.
181. Nakata M, Kohno D, Shintani N, Nemoto Y, Hashimoto H, Baba A, Yada T: **PACAP deficient mice display reduced carbohydrate intake and PACAP activates NPY-containing neurons in the rat hypothalamic arcuate nucleus.** *Neurosci Lett* 2004, **370:**252-256.
182. Hishikawa D, Hong YH, Roh SG, Miyahara H, Nishimura Y, Tomimatsu A, Tsuzuki H, Gotoh C, Kuno M, Choi KC, Lee HG, Cho KK, Hidari H, Sasaki S: **Identification of genes expressed differentially in subcutaneous and visceral fat of cattle, pig, and mouse.** *Physiol Genomics* 2005, **21:**343-350.
183. Martari M, Salvatori R: **Diseases associated with growth hormone-releasing hormone receptor (GHRHR) mutations.** *Prog Mol Biol Transl Sci* 2009, **88:**57-84.
184. Pereira RM, Aguiar-Oliveira MH, Sagazio A, Oliveira CR, Oliveira FT, Campos VC, Farias CT, Vicente TA, Gois MB, Jr., Oliveira JL, Marques-Santos C, Rocha IE, Barreto-Filho JA, Salvatori R: **Heterozygosity for a mutation in the growth hormone-releasing hormone receptor gene does not influence adult stature, but affects body composition.** *J Clin Endocrinol Metab* 2007, **92:**2353-2357.
185. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, Konig IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Doring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA, Jr., Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR,

Boehnke M, Kathiresan S: **Biological, clinical and population relevance of 95 loci for blood lipids.** *Nature* 2010, **466:**707-713.
186. Chandrasekharan S, Qiu TH, Alkharouf N, Brantley K, Mitchell JB, Liu ET: **Characterization of mice deficient in the Src family nonreceptor tyrosine kinase Frk/rak.** *Mol Cell Biol* 2002, **22:**5235-5247.
187. **Online Mendelian Inheritance in Man, OMIM® McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) November 2012. World Wide Web. URL: http://omim.org/**
188. Hodoglugil U, Williamson DW, Yu Y, Farrer LA, Mahley RW: **Glucuronic acid epimerase is associated with plasma triglyceride and high-density lipoprotein cholesterol levels in Turks.** *Ann Hum Genet* 2011, **75:**398-417.
189. Mahley RW, Ji ZS: **Remnant lipoprotein metabolism: key pathways involving cell-surface heparan sulfate proteoglycans and apolipoprotein E.** *J Lipid Res* 1999, **40:**1-16.
190. Guilherme A, Virbasius JV, Puri V, Czech MP: **Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes.** *Nat Rev Mol Cell Biol* 2008, **9:**367-377.
191. Hou JC, Min L, Pessin JE: **Insulin granule biogenesis, trafficking and exocytosis.** *Vitam Horm* 2009, **80:**473-506.
192. Mosedale M, Egodage S, Calma RC, Chi NW, Chessler SD: **Neurexin-1alpha contributes to insulin-containing secretory granule docking.** *J Biol Chem* 2012, **287:**6350-6361.
193. Dai FF, Zhang Y, Kang Y, Wang Q, Gaisano HY, Braunewell KH, Chan CB, Wheeler MB: **The neuronal Ca2+ sensor protein visinin-like protein-1 is expressed in pancreatic islets and regulates insulin secretion.** *J Biol Chem* 2006, **281:**21942-21953.
194. Philippe J, Missotten M: **Functional characterization of a cAMP-responsive element of the rat insulin I gene.** *J Biol Chem* 1990, **265:**1465-1469.
195. Pidoux G, Witczak O, Jarnaess E, Myrvold L, Urlaub H, Stokka AJ, Kuntziger T, Tasken K: **Optic atrophy 1 is an A-kinase anchoring protein on lipid droplets that mediates adrenergic control of lipolysis.** *Embo J* 2011, **30:**4371-4386.
196. Miyoshi H, Souza SC, Endo M, Sawada T, Perfield JW, 2nd, Shimizu C, Stancheva Z, Nagai S, Strissel KJ, Yoshioka N, Obin MS, Koike T, Greenberg AS: **Perilipin overexpression in mice protects against diet-induced obesity.** *J Lipid Res* 2010, **51:**975-982.
197. Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA, Mouse Genome Database G: **The Mouse Genome Database (MGD): mouse biology and model systems.** *Nucleic Acids Res* 2008, **36:**D724-728.
198. Qi L, Tai ES, Tan CE, Shen H, Chew SK, Greenberg AS, Corella D, Ordovas JM: **Intragenic linkage disequilibrium structure of the human perilipin gene (PLIN) and haplotype association with increased obesity risk in a multiethnic Asian population.** *J Mol Med (Berl)* 2005, **83:**448-456.
199. Qi L, Corella D, Sorli JV, Portoles O, Shen H, Coltell O, Godoy D, Greenberg AS, Ordovas JM: **Genetic variation at the perilipin (PLIN) locus is associated with obesity-related phenotypes in White women.** *Clin Genet* 2004, **66:**299-310.
200. Corella D, Qi L, Tai ES, Deurenberg-Yap M, Tan CE, Chew SK, Ordovas JM: **Perilipin gene variation determines higher susceptibility to insulin resistance in Asian women when consuming a high-saturated fat, low-carbohydrate diet.** *Diabetes Care* 2006, **29:**1313-1319.
201. Rabe K, Lehrke M, Parhofer KG, Broedl UC: **Adipokines and insulin resistance.** *Mol Med* 2008, **14:**741-751.
202. Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A, Lee JH, Drackley JK, Band MR, Hernandez AG, Shani M, Lewin HA, Weller JI, Ron M:

Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res* 2005, **15:**936-944.
203. Ron M, Kliger D, Feldmesser E, Seroussi E, Ezra E, Weller JI: **Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by a daughter design.** *Genetics* 2001, **159:**727-735.
204. Heaton MP, Chitko-McKnown CG, Grosse WM, Keele JW, Keen JE, Laegreid WW: **Interleukin-8 haplotype structure from nucleotide sequence variation in commercial populations of U.S. beef cattle.** *Mamm Genome* 2001, **12:**219-226.
205. Sheehy PA, Riley LG, Raadsma HW, Williamson P, Wynn PC: **A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6.** *Anim Genet* 2009, **40:**492-498.
206. Schnabel RD, Kim JJ, Ashwell MS, Sonstegard TS, Van Tassell CP, Connor EE, Taylor JF: **Fine-mapping milk production quantitative trait loci on BTA6: analysis of the bovine osteopontin gene.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**6896-6901.
207. Eberlein A, Takasuga A, Setoguchi K, Pfuhl R, Flisikowski K, Fries R, Klopp N, Furbass R, Weikard R, Kuhn C: **Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (NCAPG) gene.** *Genetics* 2009, **183:**951-964.
208. Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, Rieder S: **A genome-wide association study reveals loci influencing height and other conformation traits in horses.** *PLoS One* 2012, **7:**e37282.
209. Setoguchi K, Watanabe T, Weikard R, Albrecht E, Kuhn C, Kinoshita A, Sugimoto Y, Takasuga A: **The SNP c.1326T>G in the non-SMC condensin I complex, subunit G (NCAPG) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle.** *Anim Genet* 2011, **42:**650-655.
210. Carty CL, Johnson NA, Hutter CM, Reiner AP, Peters U, Tang H, Kooperberg C: **Genome-wide association study of body height in African Americans: the Women's Health Initiative SNP Health Association Resource (SHARe).** *Hum Mol Genet* 2012, **21:**711-720.
211. Soranzo N, Rivadeneira F, Chinappen-Horsley U, Malkina I, Richards JB, Hammond N, Stolk L, Nica A, Inouye M, Hofman A, Stephens J, Wheeler E, Arp P, Gwilliam R, Jhamai PM, Potter S, Chaney A, Ghori MJ, Ravindrarajah R, Ermakov S, Estrada K, Pols HA, Williams FM, McArdle WL, van Meurs JB, Loos RJ, Dermitzakis ET, Ahmadi KR, Hart DJ, Ouwehand WH, Wareham NJ, Barroso I, Sandhu MS, Strachan DP, Livshits G, Spector TD, Uitterlinden AG, Deloukas P: **Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size.** *PLoS Genet* 2009, **5:**e1000445.
212. Sovio U, Bennett AJ, Millwood IY, Molitor J, O'Reilly PF, Timpson NJ, Kaakinen M, Laitinen J, Haukka J, Pillas D, Tzoulaki I, Hoggart C, Coin LJ, Whittaker J, Pouta A, Hartikainen AL, Freimer NB, Widen E, Peltonen L, Elliott P, McCarthy MI, Jarvelin MR: **Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966.** *PLoS Genet* 2009, **5:**e1000409.
213. Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, Loredo AI, Bellone RR, Mezey JG, Brooks SA, Sutter NB: **Four loci explain 83% of size variation in the horse.** *PLoS One* 2012, **7:**e39929.
214. Lindholm-Perry AK, Sexten AK, Kuehn LA, Smith TP, King DA, Shackelford SD, Wheeler TL, Ferrell CL, Jenkins TG, Snelling WM, Freetly HC: **Association, effects and validation of polymorphisms within the NCAPG - LCORL locus located on**

BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *BMC Genet* 2011, **12:**103.
215. Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ: **High-resolution haplotype block structure in the cattle genome.** *BMC Genet* 2009, **10:**19.
216. Chang D, Keinan A: **Predicting signatures of "synthetic associations" and "natural associations" from empirical patterns of human genetic variation.** *PLoS Comput Biol* 2012, **8:**e1002600.
217. Olsen HG, Gomez-Raya L, Vage DI, Olsaker I, Klungland H, Svendsen M, Adnoy T, Sabry A, Klemetsdal G, Schulman N, Kramer W, Thaller G, Ronningen K, Lien S: **A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle.** *J Dairy Sci* 2002, **85:**3124-3130.
218. Olsen HG, Lien S, Gautier M, Nilsen H, Roseth A, Berg PR, Sundsaasen KK, Svendsen M, Meuwissen TH: **Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6.** *Genetics* 2005, **169:**275-283.
219. Weikard R, Widmann P, Buitkamp J, Emmerling R, Kuehn C: **Revisiting the quantitative trait loci for milk production traits on BTA6.** *Anim Genet* 2012, **43:**318-323.
220. Davis GP, Moore SS, Drinkwater RD, Shorthose WR, Loxton ID, Barendse W, Hetzel DJ: **QTL for meat tenderness in the M. longissimus lumborum of cattle.** *Anim Genet* 2008, **39:**40-45.
221. Kolbehdari D, Wang Z, Grant JR, Murdoch B, Prasad A, Xiu Z, Marques E, Stothard P, Moore SS: **A whole-genome scan to map quantitative trait loci for conformation and functional traits in Canadian Holstein bulls.** *J Dairy Sci* 2008, **91:**2844-2856.
222. Gutierrez-Gil B, Williams JL, Homer D, Burton D, Haley CS, Wiener P: **Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle.** *J Anim Sci* 2009, **87:**24-36.
223. Kneeland J, Li C, Basarab J, Snelling WM, Benkel B, Murdoch B, Hansen C, Moore SS: **Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of Bos taurus.** *J Anim Sci* 2004, **82:**3405-3414.
224. Maltecca C, Weigel KA, Khatib H, Cowan M, Bagnato A: **Whole-genome scan for quantitative trait loci associated with birth weight, gestation length and passive immune transfer in a Holstein x Jersey crossbred population.** *Anim Genet* 2009, **40:**27-34.
225. Mizoguchi Y, Watanabe T, Fujinaka K, Iwamoto E, Sugimoto Y: **Mapping of quantitative trait loci for carcass traits in a Japanese Black (Wagyu) cattle population.** *Anim Genet* 2006, **37:**51-54.
226. Murphy LA, Sarge KD: **Phosphorylation of CAP-G is required for its chromosomal DNA localization during mitosis.** *Biochem Biophys Res Commun* 2008, **377:**1007-1011.
227. Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ, Harper JW, Gygi SP: **Systematic and quantitative assessment of the ubiquitin-modified proteome.** *Mol Cell* 2011, **44:**325-340.
228. Kimura Y, Tanaka K: **Regulatory mechanisms involved in the control of ubiquitin homeostasis.** *J Biochem* 2010, **147:**793-798.
229. Vinayagam A, Stelzl U, Foulle R, Plassmann S, Zenkner M, Timm J, Assmus HE, Andrade-Navarro MA, Wanker EE: **A directed protein interaction network for investigating intracellular signal transduction.** *Sci Signal* 2011, **4:**rs8.
230. Dorman K, Shen Z, Yang C, Ezzat S, Asa SL: **CtBP1 interacts with Ikaros and modulates pituitary tumor cell survival and response to hypoxia.** *Mol Endocrinol* 2012, **26:**447-457.

231. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W: **PipMaker--a web server for aligning two genomic DNA sequences.** *Genome Res* 2000, **10:**577-586.
232. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A, Demarche B, Gut I, Heath S, Foglio M, Liang L, Laukens D, Mni M, Zelenika D, Van Gossum A, Rutgeerts P, Belaiche J, Lathrop M, Georges M: **Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4.** *PLoS Genet* 2007, **3:**e58.
233. Maston GA, Evans SK, Green MR: **Transcriptional regulatory elements in the human genome.** *Annu Rev Genomics Hum Genet* 2006, **7:**29-59.
234. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, Fall T, Seppala EH, Hansen MS, Lawley CT, Karlsson EK, Bannasch D, Vila C, Lohi H, Galibert F, Fredholm M, Haggstrom J, Hedhammar A, Andre C, Lindblad-Toh K, Hitte C, Webster MT: **Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping.** *PLoS Genet* 2011, **7:**e1002316.
235. Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB, Archibald AL, Fredholm M, Groenen MA, Andersson L: **Strong signatures of selection in the domestic pig genome.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109:**19529-19536.
236. Sauna ZE, Kimchi-Sarfaty C: **Understanding the contribution of synonymous mutations to human disease.** *Nat Rev Genet* 2011, **12:**683-691.
237. Weikard R, Altmaier E, Suhre K, Weinberger KM, Hammon HM, Albrecht E, Setoguchi K, Takasuga A, Kuhn C: **Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition.** *Physiol Genomics* 2010, **42A:**79-88.
238. Wu G, Bazer FW, Davis TA, Kim SW, Li P, Marc Rhoads J, Carey Satterfield M, Smith SB, Spencer TE, Yin Y: **Arginine metabolism and nutrition in growth, health and disease.** *Amino Acids* 2009, **37:**153-168.
239. Altmann S, Murani E, Schwerin M, Metges CC, Wimmers K, Ponsuksili S: **Maternal dietary protein restriction and excess affects offspring gene expression and methylation of non-SMC subunits of condensin I in liver and skeletal muscle.** *Epigenetics* 2012, **7:**239-252.
240. Wang GS, Cooper TA: **Splicing in disease: disruption of the splicing code and the decoding machinery.** *Nat Rev Genet* 2007, **8:**749-761.

Appendix

Description of supplementary file: Markey_thesis_supplement.xlsx

Table S1. Genotypes of reported Hypotrichosis affected calves on BTA5

The table contains SNP and microsatellite genotypes of the purported Hypotrichosis affected calves used in the analysis. The Mb position is based on Btau4.0. Tan shading indicates a run of homozygosity and yellow shading indicates a heterozygous genotype bordering a run of homozygosity. A red shaded calf indicates the calf was not heterozygous or homozygous for the HY deletion mutation. Group A represents calves resulting from the initial PLINK homozygosity analysis while group B represents calves that resulted from relaxed homozygosity commands. Group C represents calves that did not appear in either of the homozygosity analysis from PLINK.

Table S2. Primer sequences used for Hypotrichosis region on BTA5

Primer sequences and annealing temperatures used to amplify exon, microsatellite and HY mutation assay amplicons.

Table S3. Single SNP associations with traits and genomic regions

Associations reported with Bonferroni corrected $p \leq 1.0 \times 10^{-3}$.

Table S4. Selected results from haplotype association analysis with back fat on BTA6

A haplotype analysis using a 20 SNP sliding window was conducted on BTA6 from 36 to 43 Mb using a quantitative association analysis in PLINK. The 12 most significant haplotypes from the two regions on BTA6 were selected and all haplotypes from the significant haplotype windows are represented in the table.

Table S5. Selected results from haplotype association analysis with marbling on BTA6

A haplotype analysis using a 20 SNP sliding window was conducted on BTA6 from 36 to 43 Mb using a quantitative association analysis in PLINK. The five most significant haplotypes from the region on BTA6 were selected and all haplotypes from the significant haplotype windows are represented in the table.

Table S6. Selected results from haplotype association analysis with birth weight on BTA6

A haplotype analysis using a 20 SNP sliding window was conducted on BTA6 from 36 to 43 Mb using a quantitative association analysis in PLINK. The 12 most significant haplotypes from the two regions on BTA6 were selected and all haplotypes from the significant haplotype windows are represented in the table.

Table S7. Primer sequences used to amplify the LCORL-NCAPG region on BTA6

Primer sequences, amplicon sizes and annealing temperatures used to amplify segments of the LCORL-NCAPG region on BTA6.