

MAPPING NATURAL AND ARTIFICIAL SELECTION  
EVENTS IN ANIMAL GENOMES

---

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri – Columbia

---

In Partial Fulfilment

Of the Requirements for the Degree

Doctor of Philosophy

---

By

HOLLY RENÉ RAMEY

Dr. Jeremy Taylor, Dissertation Supervisor

MAY 2015

The undersigned, appointed by the Dean of the Graduate School, have examined the

Dissertation entitled

MAPPING NATURAL AND ARTIFICIAL SELECTION EVENTS  
IN ANIMAL GENOMES

Presented by Holly René Ramey,

A candidate for the degree of Doctor of Philosophy,

And hereby certify that, in their opinion, it is worthy of acceptance.

---

Dr. Jeremy F. Taylor

---

Dr. Robert D. Schnabel

---

Dr. William R. Lamberson

---

Dr. Gavin C. Conant

---

Dr. Gary S. Johnson

## *Dedication*

A long time ago I said that “It’s all about passion.... If we are passionate enough about something then we think about it all the time and that is where the break point between sanity and insanity lies.” I thought at the time, about seven years ago, that I understood what I was saying. But how wrong was I. It wasn’t until graduate school that I finally realized that the moment I wrote that down was a moment of wisdom that would ring no truer than when I was deep into my graduate studies. These past four and a half years have pushed me to the edge of the breaking point many times. People outside of the academic world ask me all the time whether I enjoy what I am studying in school. My response takes them by surprise occasionally when I say that some days I do, some days I don’t, but that it drives me crazy every day. Those moments and especially the moments where I am standing on the breaking point are the moments that I know I love what I do and most importantly they tell me I landed in the right place here at Mizzou.

Nothing about this process is easy. It requires a lot of prayer, late nights, mistakes, failures, and drinking. I know I am accomplished as a scientist, a genomicist, a soon-to-be Ph.D. but I have doubted myself the entire way. I have been my own worst enemy. Maybe it comes from the negative people in my life, the wonderful yet intimidating scientists I work alongside, or just my innate fear of failure. Whatever it is, wherever it comes from, I am slowly pushing it away. The past year and a half have really been the biggest steps towards this and being confident in my studies. Tackling Comprehensives really showed me what I am capable of. They’re scary as hell but I learned so much about the science and myself during that process that I am thankful for it. Now that the end is near I think I can finally look back at the last few years and say that I conquered it –

trials, tribulations, and all. I have made it through the tunnel and found the light on the other end.

The last question people ask me is “Do you think looking back on it now you would do it again if you knew then what you know now?” That answer I cannot provide yet, but I am hopeful in six months I can confidently say “Absolutely I would – without a shadow of a doubt...”

The people that have helped me along the way really kept me from falling off the edge at the breaking points. I have had many friends in the process and support from family and friends back home. Without these people I could not have completed this process sanely. Whether these friends were fleeting or lifelong friendships, I can be thankful for them all since each one has taught me something, both good and bad. Graduate school friends certainly come and go while you are doing a Ph.D. because you are here for a long time. That was a personally painful process a few times. Saying goodbye is not something I do easily. The friends who have been paramount to my sanity are Angie Schenewerk Rost, Veronica Negron Perez, Megan Rolf, Sarah Gregg, the entire women’s Columbia Rugby Football Club team – the Black Sheep (namely Carrie Marston and Christina Holzhauser), Andrew Alba, Maria Haag, Alex Stuckel, and Christine Hurley. Two names I have not included in this list are the ones that have truly been the best two people I have had the pleasure to call my friends. These special people are Dane Burmeister and Kira Marshall. I could not replace the effects these two have had on my life. Dane brought me back to life when I hit my first slump after the first year of grad school. He was the best to watch sports with, taught me to embrace alternative music and campy movies, unlocked my surprising love of hockey, and above all was the

absolute best drinking buddy! And Kira... I cannot even begin to put into words how thankful and blessed I have been to have her in my life. My soul feels more complete now that I know her. We have held each other up in the low times, shared a love of music and dance, and had some of the most fun and random adventures together. Simply put, she is the sister I never thought I would have.

My family has also been a huge support for me. I must first mention my lovely niece Mikayla "Kayla Bug" Ramey. I do not get to talk to or see her often but knowing how much she looks up to me drives me to be my best. In particular, there are four other people that have really been there for me. Tammy Brown, one of my cousins though I cannot say exactly which generation, has been there to hold me up when others in my family have let me fall. I am thankful for her almost weekly calls and cards that I get in the mail. She is definitely a great lady. My self-proclaimed little brother, Stephen Kleisner, has been there for me regardless of time, circumstance, or distance. He is one of the best people you will ever meet and I am beyond thankful for him. My wonderful, loving mother Cindy Ramey has been a rock for me. Things in my personal life have not changed as drastically as they have over the past four and a half years but she has been steady throughout. I am blessed beyond measure to call her my mother and I strive every day to be as wonderful a person as she is. And lastly, and almost as important as my Mom, is my boyfriend William Atkinson. The road we have taken through graduate school has been interesting. Being a thousand miles away has been the most difficult part. He has put up with a lot of venting, tears, questions, and general grumpiness. I still do not understand how or why he has stayed with me but I am appreciative that he has and that

we have grown closer rather than apart. Thank you for not giving up on me – even when I've wanted to give up on myself.

And more importantly than all of the things I have mentioned, I want to thank God for giving me the strength and will to be able to go through this entire process. You and prayer have been a friend and a source of hope for me – almost daily. Thank you for everything.

*My sincerest thanks,*

*Holly*

## **Acknowledgements**

When people and students approach me about the biggest decisions that they will encounter when deciding on a graduate education, I cannot preach enough that choosing the right advisor is the most important thing you will do. I was lucky enough to find mine during an internship in the summer of 2009 by happenstance. Dr. Jerry Taylor has been the absolute best mentor I could have ever asked for. He has not only provided me travel opportunities and scientific experiences beyond what a lot of professors can provide but he has allowed me the autonomy in my studies that have shaped me into a real scientist. The lab culture and his teaching doctrine have been wonderful. Dr. Taylor is a wonderful man and inspiring mentor. I credit him with all of my successes in graduate school. He is truly a world class mentor.

Not only have I had the best doctoral mentor but I've had the privilege of having a second mentor, Dr. Bob Schnabel. He is the heart and soul of the data in the Taylor laboratory. I do not think anyone could replace him and the high quality work that he does. I have learned just as many things from him as Dr. Taylor. He is quick to keep students grounded and cognizant of what they are doing and what the limitations of methods and the data are – even when you don't want to hear it. It has been a true pleasure to have Drs. Taylor and Schnabel behind me throughout my studies. I do not know how to thank them enough.

My Committee, including Drs. Bill Lamberson, Gavin Conant, and Gary Johnson, have been a great help to push me to achieve what they know is possible and help me find the right path if I get side tracked. They were thorough and thought provoking during my

Comprehensive examination process and I am thankful to them for that and all their continued help.

My lab mates, both past and present, have also impacted my graduate experience. These have been both positive and negative experiences at various times but valuable nonetheless. All of the following scientists are those involved with my time at the Taylor lab group: Dr. JaeWoo Kim, Dr. Stephanie McKay, Sarah Gregg, Helen Yampara-Iquise, Dr. Megan Rolf, Richard Chapple, Dr. Jared Decker, Jesse Hoff, Lynsey Whitacre, Maria Haag, Miranda Wilson, and Tamar Adcock. One of my fellow graduate students, Zhiyuan Chen, has also been of the utmost help for anything I needed to have explained or to just bounce ideas around with. A few of the visiting scientists also left lasting impressions on my work including Olivia Mapholi and Dr. Polyana Tizioto. These are some of the best people on a personal level and great scientists who I hope to continue working with in the future. Thank you to all of you.



## Table of Contents

Acknowledgements .....	ii
List of Tables .....	vi
List of Figures.....	viii
<b>1. Introduction.....</b>	<b>1</b>
<b>2. Detection of selective sweeps in cattle using genome-wide SNP data.....</b>	<b>4</b>
<b>Abstract .....</b>	<b>5</b>
<b>Background .....</b>	<b>5</b>
<b>Results.....</b>	<b>5</b>
<b>Conclusions.....</b>	<b>5</b>
<b>Keywords.....</b>	<b>6</b>
<b>Background.....</b>	<b>6</b>
<b>Materials and Methods .....</b>	<b>9</b>
<i>Samples, design, and genotyping.....</i>	<i>9</i>
<i>SNP filtering .....</i>	<i>10</i>
<i>Identification of putative selective sweep regions using BovineSNP50 data .....</i>	<i>11</i>
<i>Identification of putative selective sweep regions using AFFXB1P data.....</i>	<i>13</i>
<i>Annotation and functional analysis .....</i>	<i>14</i>
<b>Results .....</b>	<b>14</b>
<b>Regions identified as harbouring selective sweeps using BovineSNP50 data ....</b>	<b>14</b>
<b>Regions identified as harbouring selective sweeps using AFFXB1P data .....</b>	<b>16</b>
<b>Annotation and casual candidates underlying selective sweep regions .....</b>	<b>21</b>
<b>Discussion.....</b>	<b>21</b>
<b>List of Abbreviations.....</b>	<b>41</b>
<b>Competing interests.....</b>	<b>41</b>
<b>Authors' contributions.....</b>	<b>41</b>
<b>Acknowledgements.....</b>	<b>42</b>
<b>3. A survey of epistasis in growth and carcass traits of Angus steers .....</b>	<b>43</b>
<b>Abstract .....</b>	<b>44</b>
<b>Keywords.....</b>	<b>45</b>
<b>Background.....</b>	<b>45</b>
<b>Materials and Methods .....</b>	<b>47</b>

<i>Samples and phenotyping</i> .....	47
<i>Genotyping and SNP filtering</i> .....	47
<i>Identification of two-locus epistatic interactions</i> .....	48
<i>Annotation and pathway analysis</i> .....	49
<b>Results and Discussion</b> .....	50
<b>Software selection</b> .....	50
<b>Single locus effects</b> .....	51
<b>Interaction effects</b> .....	51
<b>Network formation</b> .....	52
<b>Weight trait networks</b> .....	53
<i>Overlap between traits</i> .....	53
<i>Correlated traits with non-overlapping networks</i> .....	58
<b>Fatness trait networks</b> .....	59
<b>Pathway analysis</b> .....	63
<b>Data comparisons and adjustments</b> .....	66
<b>Conclusions</b> .....	67
<b>List of Abbreviations</b> .....	68
<b>Competing interests</b> .....	68
<b>Authors' contributions</b> .....	68
<b>Acknowledgements</b> .....	68
<b>4. BMP3 does not cause the brachycephaly phenotype within Pug dogs</b> .....	<b>103</b>
<b>Abstract</b> .....	104
<b>Keywords</b> .....	105
<b>Background</b> .....	105
<b>Materials and Methods</b> .....	107
<i>Sample selection, sequencing, and variant calling</i> .....	107
<i>Dataset filtering</i> .....	109
<i>Runs of homozygosity analysis</i> .....	113
<b>Results and Discussion</b> .....	114
<b>Conclusion</b> .....	119
<b>References</b> .....	<b>121</b>
<b>VITA</b> .....	<b>135</b>

## List of Tables

Table 2.1 Summary for genotyped individuals.....	12
Table 2.2 Putative selective sweep regions identified by analysis of Bovine SNP50 genotypes .....	18
Table 2.3 Putative selective sweep regions detected in at least two breeds using AFFXB1P .....	19
Table 2.4 Genomic regions predicted to harbour selective sweeps using BovineSNP50 data and validated by AFFXB1P data.....	20
Table 2.5 Putative breed-specific selective sweeps identified using AFFXB1P data ...	27
Table 2.6 Potential causal genes underlying selective sweep regions and their associated function or phenotype .....	32
Table 3.1 Summary of analysed traits.....	48
Table 3.2 Significance values of pairwise interactions by trait .....	52
Table 3.3 Relevant enrichment within traits .....	64
Table 3.4 Top five networks identified by trait .....	65
Table 3.S1 Pairwise interaction results – BW .....	70
Table 3.S2 Pairwise interaction results – FAT .....	74
Table 3.S3 Pairwise interaction results – HCW.....	77
Table 3.S4 Pairwise interaction results – USFT .....	81
Table 3.S5 Pairwise interaction results – USIMF.....	84
Table 3.S6 Pairwise interaction results – USWT.....	88
Table 3.S7 Pairwise interaction results – WW .....	92

Table 4.1 Sample demography and classification.....	110
Table 4.2 Detected runs of homozygosity .....	118

## List of Figures

Figure 2.1 Selective sweep regions discovered in the analysis of the BovineSNP50 data that were predicted to be common to two or more breeds.....	17
Figure 2.2 Selective sweep surrounding the POLL locus in Angus cattle.....	29
Figure 3.1 Top epistatic network for WW .....	55
Figure 3.2 Top epistatic network for HCW .....	56
Figure 3.3 Top epistatic network for USWT .....	57
Figure 3.4 Top epistatic network for FAT .....	60
Figure 3.5 Top epistatic network for USFT.....	61
Figure 3.S1 Top epistatic network for WW based on top 600 pairwise interactions ....	96
Figure 3.S2 Top epistatic network for HCW based on top 600 pairwise interactions...	97
Figure 3.S3 Top epistatic network for BW.....	98
Figure 3.S4 Second-ranked epistatic network for FAT .....	99
Figure 3.S5 Second-ranked epistatic network for USFT.....	100
Figure 3.S6 Third-ranked epistatic network for USFT.....	101
Figure 3.S7 Fourth-ranked epistatic network for USFT .....	102
Figure 4.1 Selective sweep scan results on CFA 32 in Pugs .....	116
Figure 4.2 Selective sweep scan results on CFA 1 in Pugs .....	11

# 1. Introduction

Selection events, of both artificial and natural origin, have left their mark on animal genomes. These signatures are detectable through a variety of analyses including scans for quantitative trait loci (QTL) and selective sweep detection. Utilizing these types of studies in large datasets is advantageous for the discovery and identification of causal mutations of large phenotypic effect. Elucidating the mutations or variants responsible for large proportions of phenotypic variation is paramount to understanding the unique genetic architectures of individual breeds or biological types of domesticated animals. Which mutations create the change in nucleotide sequence that is responsible for black coat colour in Angus cattle, the brachycephaly phenotype in Pugs, or the carcass composition of an animal? These types of questions can be investigated by mapping QTL and selective sweeps because the performed analyses are based on the fundamental theory that advantageous alleles are either artificially or naturally selected due to their benefit to the industry or consumer which increases the fitness of animals carrying these alleles which are then driven to high frequencies or even to fixation in a population [1-3].

Over time, a vast number of QTL and sweep regions harbouring variants which underlie various carcass composition, reproductive, health, and aesthetic phenotypes have been discovered. These studies have also led to a further understanding of the polygenicity of quantitative traits. Selective sweeps are frequently associated with variants responsible for Mendelian traits where a single mutation of beneficial effect is responsible for a specific phenotype. Quantitative trait locus mapping generally results in the identification of multiple mutations or genomic regions having individually large effects on a phenotype. The polygenic nature of quantitative traits and the inability of

current methodologies to completely explain their heritabilities using molecular data [4] have opened up another field of study in genomics – epistasis. Epistasis, which involves the interaction among two or more genes, could begin to fill in the gaps in our knowledge about the regulation of quantitative traits. Understanding how the genotypes at multiple loci within an individual can have effects beyond simply what each individual locus contributes to phenotype is an interesting and emerging area of research. Epistatic interactions among loci within the genome is becoming an increasingly important frontier for study as no single locus explains a significant proportion of phenotypic variation and many QTL may be context dependent. Such studies require very large data sets and among the most important of phenotypes in the beef industry are carcass composition traits which drive US consumer preferences.

This dissertation pursues topics in selective sweep mapping, epistatic network identification, and causal mutation identification within the genomes of cattle and dogs. Each chapter represents a research endeavour complete with a review of the current literature and an addition to the understanding of a specific problem. The studies provide new findings that are relevant to not only the cattle and dog scientific communities, but to genomics as a whole. Methods developed and results described within the chapters are applicable to any population with a diploid genome. Simply put, the approaches are species agnostic. The results and discussion are presented according to the relevance to the respective species but are also placed in a broader picture of the field of genomics. The overarching goal of this dissertation was to provide insights into the catalogue of variation that makes individuals within a population unique – describing both the

similarities and differences of biological type, geographical origin, and uses of the animals while striving to define the mutations underlying these.



## **2. Detection of selective sweeps in cattle using genome-wide SNP data**

Holly R. Ramey<sup>1</sup>, Jared E. Decker<sup>1</sup>, Stephanie D. McKay<sup>1,2</sup>, Megan M. Rolf<sup>1,3</sup>,

Robert D. Schnabel<sup>1</sup>, Jeremy F. Taylor<sup>1§</sup>

<sup>1</sup>Division of Animal Sciences, University of Missouri, Columbia MO 65211 USA

<sup>2</sup>Department of Animal Science, University of Vermont, Burlington, VT 05405 USA

<sup>3</sup>Department of Animal Science, Oklahoma State University, Stillwater, OK 74074 USA

§Corresponding author

### **Corresponding Author:**

Dr. Jeremy F. Taylor

S135B Animal Sciences

920 East Campus Drive

University of Missouri

Columbia MO 65211-5300

E-mail: [taylorjerr@missouri.edu](mailto:taylorjerr@missouri.edu)

Tel.: +1-573-884-4946

Fax: +1-573-882-6827

## **Abstract**

### **Background**

The domestication and subsequent selection by humans to create breeds and biological types of cattle undoubtedly altered the patterning of variation within their genomes. Strong selection to fix advantageous large-effect mutations underlying domesticability, breed characteristics or productivity created selective sweeps in which variation was lost in the chromosomal region flanking the selected allele. Selective sweeps have now been identified in the genomes of many animal species including humans, dogs, horses, and chickens. Here, we attempt to identify and characterise regions of the bovine genome that have been subjected to selective sweeps.

### **Results**

Two datasets were used for the discovery and validation of selective sweeps via the fixation of alleles at a series of contiguous SNP loci. BovineSNP50 data were used to identify 28 putative sweep regions among 14 diverse cattle breeds. Affymetrix BOS 1 prescreening assay data for five breeds were used to identify 85 regions and validate 5 regions identified using the BovineSNP50 data. Many genes are located within these regions and the lack of sequence data for the analysed breeds precludes the nomination of selected genes or variants and limits the prediction of the selected phenotypes. However, phenotypes that we predict to have historically been under strong selection include horned-polled, coat colour, stature, ear morphology, and behaviour.

### **Conclusions**

The bias towards common SNPs in the design of the BovineSNP50 assay led to the identification of recent selective sweeps associated with breed formation and common to

only a small number of breeds rather than ancient events associated with domestication which could potentially be common to all European taurines. The limited SNP density, or marker resolution, of the BovineSNP50 assay significantly impacted the rate of false discovery of selective sweeps, however, we found sweeps in common between breeds which were confirmed using an ultra-high-density assay scored in a small number of animals from a subset of the breeds. No sweep regions were shared between indicine and taurine breeds reflecting their divergent selection histories and the very different environmental habitats to which these sub-species have adapted.

### **Keywords**

selective sweep, reduced heterozygosity, signatures of selection, single nucleotide polymorphisms

### **Background**

The transition from hunter-gather lifestyles to permanent dwelling societies was facilitated by both plant and animal domestication [5]. The domestication of cattle occurred between 8,000 and 10,000 years ago and led to changes in the genome of the species due to the effects of demography and selection [6, 7]. Much of the variation within the genetically diverse ancestral population was either lost due to the limited sampling of animals within the sites of domestication or was partitioned into the subpopulations which went on to become recognised as distinct breeds. Selection for the phenotypes contributing to domesticability, biological type (draught, milk, meat) and the aesthetically appealing morphologies that have become breed hallmarks (polled, coat colour and patterning [8-12]) have also impacted the extent and distribution of variability within the genome.

Strong on-going selection for variants of large effect leads to a loss of variation within the chromosomal region flanking the selected variant and ultimately the complete fixation of the haplotype which harbours the variant. This phenomenon is known as the “hitch-hiking effect” [1] and a region of the genome in which artificial selection has driven a haplotype to complete fixation is defined as having been subjected to a “selective sweep.” Such regions may also occur within the genome due to random drift and these regions are not distinguishable from regions subjected to selective sweeps. Selective sweep studies differ from the classical “forward” genetics approach, which progresses from a phenotype to the identification of underlying causal genes and mutations. Rather, they follow a “reverse” genetics approach that begins with a signature of selection and attempts to infer the selected mutation and its associated phenotype [5]. An important reason for seeking selective sweeps is that these regions can elucidate the identities of genes and mutations with large phenotypic effect even if they are no longer segregating within any one population and thus cannot be detected by forward genetics without the formation of expensive crosses.

Several methods have been used to identify regions of the genome which have been subjected to selective sweeps, including those based on modelling allele frequency spectra, linkage disequilibrium and haplotype structure [13-15]. These approaches require the use of high-density single nucleotide polymorphism (SNP) data which have previously been shown to be useful for detecting selective sweeps in human populations [16, 17]. Studies aimed at localizing signatures of selection and selective sweeps have been performed in many animal species using SNP and microsatellite loci. These studies have pointed to interesting phenotypes which are important to understanding the nature of

historic natural and artificial selection applied to these species. In chicken, selective sweeps have been found to involve loci believed to be inherent to domestication and include *BCDO2* which controls yellow and white skin colours, *SEMA3A* which plays a role in axonal path-finding important in brain development, and *THSR* which is postulated to derestrict the regulation of seasonal reproduction [18]. Selective sweeps have been found in the dog genome at *TRYPI* which controls black coat colour in Large Munsterlanders and at *FGFR3* in Dachshunds [19]. *FGFR3* mutations cause achondroplasia in humans and cattle. Other studies in dogs have identified a sweep surrounding *IGF1* which is responsible for size variation [20] and in a genomic region for which the selected phenotype is unknown in Boxers [21]. These sweeps range in size from 28 kb to 40 Mb suggesting considerable variation in the intensity of selection and also in the population census sizes of these breeds. A ~75 kb selective sweep at a locus influencing stature in the horse is upstream of a transcription factor (*LCORL*) that is associated with variation in human height [22]. A 28 kb selective sweep in a region of the swine genome harbouring *IGF2* has been implicated with selection for increased muscle mass and decreased fat deposition [23]. More recently, whole genome resequencing has been utilized in swine to identify selective sweeps in the *NR6A1*, *PLAG1* and *LCORL* genes which are associated with an increased number of vertebrae and an elongation of the animal's back [24].

Domestic animals have been demonstrated to be excellent models for genetic studies due to the availability of extensive pedigrees and because, as species, they are frequently more genetically diverse than human [25]. However, relatively few large-scale selective sweep studies have been conducted in cattle to elucidate the genes which have

historically been selected by humans to create the existing diversity of breeds and specialised biological types for draught, milk and meat production [26]. Rather, studies to date have tended to focus on specific breeds, individual biological type, or chromosomal regions [27-31]. In this study, we sought to identify signatures of completed selective sweeps genome-wide using 6,373 animals from 14 breeds genotyped with the high-density BovineSNP50 assay and 58 individuals from five breeds genotyped with the ultra-high-density Affymetrix BOS 1 prescreening assay (AFFXB1P). The sampled animals represent 14 meat and milk producing breeds as well as the taurine and indicine sub-species (Table 2.1). The detected sweeps were considered to be validated if they were found in more than one breed or if they were found in analyses of both the BovineSNP50 and AFFXB1P data. Our goals are to ultimately identify the selected mutations and phenotypes subjected to selection by our ancestral herdsmen in the processes of domestication and formation of breeds and biological types. We present here the most comprehensive genome-wide analysis of selective sweeps in cattle.

## **Materials and Methods**

### *Samples, design, and genotyping*

We utilized two data sets comprising SNPs scored in animals that were registered by their respective breed associations and we sampled across male lineages to ensure that the animals were not closely related and that they represented the diversity within each breed. The first data set comprised 6,373 full-blood animals from 13 taurine breeds (*Bos taurus taurus*) including Angus, Braunvieh, Charolais, Hanwoo, Hereford, Limousin, Salers, Shorthorn, Simmental, Brown Swiss, Finnish Ayrshire, Holstein, and Jersey, and the indicine (*Bos taurus indicus*) Brahman breed (Table 2.1). Genotypes scored in these

animals were generated using the Illumina (San Diego, CA) BovineSNP50 BeadChip which assayed 54,001 loci with a median intermarker interval of 37 kb [32, 33]. The second data set comprised 58 animals from the Angus, Hanwoo, Simmental, Wagyu and Brahman breeds (Table 2.1) which were genotyped with a prescreening assay comprising 2,787,037 SNPs with a median intermarker interval of 975 bp that was used by Affymetrix (Santa Clara, CA) in the design of the Axiom Genome-Wide BOS 1 assay [34]. The Angus, Hanwoo, Simmental and Brahman animals genotyped with the AFFXB1P assay were not selected to be full-bloods and not all animals were included in the BovineSNP50 data set. Thus the animal samples are partially independent and decreased from 4.4 to 126.8 fold in size between the assays, but the assay resolution, measured as SNP density, increased ~50 fold between the assays.

The sampled breeds were chosen based on their geographical origins, historical uses by human, diverse phylogenetic relationships and because of the availability of at least 40 BovineSNP50 genotyped full-blood individuals. Each genotyped individual was registered with its respective breed association and was proven by pedigree-analysis to be full-blood, since some associations (e.g., Simmental, Limousin) allow the registration of crossbred cattle. This sampling strategy was employed to ensure that there would be minimal effects of recent introgression between the breeds following breed formation.

#### *SNP filtering*

All X-linked loci were removed from the analysis due to the greater number of assembly issues that are associated with this chromosome and also because the studied animals were male resulting in a halving of the number of chromosomes sampled for each breed which leads to a reduction in the precision of allele frequency estimation. The remaining

BovineSNP50 genotypes were filtered on call rate  $\geq 85\%$  which left a total of 52,942 SNPs. Within the higher density AFFXB1P data, we required SNPs to have a call rate of  $\geq 85\%$  across all 5 breeds and a minimum call rate of 50% within each of the individual breeds. Following filtering 2,575,339 SNPs remained. These thresholds were based upon an empirical examination of the call rate distributions in the datasets (not shown) to retain the largest number of quality SNP in consideration of the within-breed sample size, which was small for the AFFXB1P genotyped individuals.

*Identification of putative selective sweep regions using BovineSNP50 data*

The BovineSNP50 data were analysed by breed to identify putative selective sweeps. Because the number of variable loci differed within each breed primarily due to the breed of origin of SNP discovery in the design of the assay [14, 15], we required a breed-specific number (Table 2.1), and at a minimum 5, contiguous SNPs spanning at least 200 kb based upon UMD3.1 coordinates for which no SNP had a minor allele frequency (MAF)  $> 0.01$  to declare a selective sweep. While, on completion, selective sweeps are characterised by the complete loss of variation within the swept region, we allowed  $MAF \leq 0.01$  to account for genotyping errors, the possibility of new mutations and assembly errors which may have erroneously assigned a variable marker to a sweep region. Determination of the number of contiguous markers within each breed with  $MAF \leq 0.01$  to define a sweep region required a trade-off between type I error and the size of the detected sweep region. For example, if 15% of SNPs are monomorphic within a breed (Table 2.1) [32, 33], the probability that N contiguous SNPs are monomorphic is  $0.15^N$ , assuming independence, and in testing 52,942 SNP on 29 autosomes we would expect to find  $0.15^N \times (52,942 - 29 \times (N - 1))$  regions in which N contiguous SNPs had fixed



**Table 2.1 Summary for genotyped individuals**

Breed <sup>1</sup>	Origin	Primary Historical Use	Contiguous		Number		Number
			BovineSNP50 Loci <sup>2</sup>	Monomorphic SNP50 Loci	BovineSNP50 Individuals	AFFXB1P Individuals	
Angus	Scotland	Beef	6	8,443	2,918	23	
Braunvieh	Switzerland	Beef	5	7,405	142		
Charolais	France	Beef	5	5,884	44		
Hanwoo	Korea	Beef, Draught	5	8,353	48	11	
Hereford	UK	Beef	6	8,154	812		
Limousin	France	Beef	6	8,430	261		
Salers	France	Beef	6	8,409	72		
Shorthorn	UK	Beef, Dairy	6	8,558	108		
Simmental	Switzerland	Beef, Dairy	6	6,599	123	6	
Brown Swiss	Switzerland	Dairy	9	12,553	74		
Finnish Ayrshire	Scotland	Dairy	5	6,185	599		
Holstein	Netherlands	Dairy	6	8,587	995		
Jersey	Jersey	Dairy	8	12,547	78		
Wagyu	Japan	Beef, Draught				10	
Brahman	USA	Beef	10	13,006	99	8	
<b>Total</b>					<b>6,373</b>	<b>58</b>	

<sup>1</sup>Brahman are indicine cattle developed in the US as a composite of Nelore, Gir, Guzerat and Indu Brazil cattle originating in India but imported primarily from Brazil. The remaining breeds are taurine.

<sup>2</sup>Number of contiguous loci spanning at least 200 kb and with a minor allele frequency  $\leq 0.01$  required to declare a selective sweep region in each breed.

alleles. For  $N = 5$  this corresponds to 4.0 false positives per breed but only 0.6 false positives per breed when  $N = 6$ . While increasing the number of contiguous monomorphic markers decreases the type I error rate, it simultaneously increases the size of the sweep region that can be detected to, on average,

$(N - 1) \times 37$  kb. To allow the identification of moderately sized sweeps, we chose an intermediate balance of these conflicting constraints based on the idea that any sweep identified in two or more breeds would almost certainly be real and likely share a common haplotype, while true sweeps found in only one breed would ultimately be independently validated by other studies. For example, assuming one false positive in each breed, there are  $52,942 - 29 \times (6 - 1) = 52,797$  possible locations for a fixed 6 SNP haplotype in each breed and in the second breed there are 11 positions where the two haplotypes may overlap by at least one SNP (ignoring the centromeres and telomeres where the number is less). Thus the probability that the two false positive fixed haplotypes overlap anywhere in the genome is  $52,797 \times (1/52,797 \times 11/52,797) = 0.0002$ .

#### *Identification of putative selective sweep regions using AFFXB1P data*

We independently analysed the AFFXB1P data by breed requiring a putative selective sweep region to harbour at least 20 contiguous SNPs spanning at least 100 kb, with no more than 5% of the SNP having a  $MAF > (2M)^{-1}$  where  $M$  was the number of individuals with genotypes for the SNP within the analysed breed. Among the variable SNPs, we further required that no more than 3 be contiguous. These thresholds were set to allow no more than one individual be heterozygous for a SNP within a selective sweep region in the event of genotyping errors and to allow for new variation to have been created within each region by mutation. These conditions again also allow for the

possibility that the SNPs may not have been correctly ordered by the UMD3.1 assembly and that variable contigs may have erroneously been included within scaffolds containing a selective sweep. For regions containing 20 contiguous monomorphic SNPs, we would expect less than  $10^{-10}$  false positives per breed if 15% of the SNPs were monomorphic within the breed. However, estimation of MAF for the loci on this assay was also influenced by the small sample size within each breed.

#### *Annotation and functional analysis*

Annotation of the genes present within all putative selective sweep regions was performed using the UCSC Genome Browser [35] and NCBI Gene database. Genes for which annotations were retrieved included any genes that were fully, or partially, contained within each region. Phenotypes known to be affected by variation in these genes were determined from a search of the literature and were assessed for their likely causality for each sweep. Functional analyses were performed for the sweeps detected within each breed using the functional annotation and clustering tools in the Database for Annotation, Visualization and Integrated Discovery (DAVID) [36].

## **Results**

### **Regions identified as harbouring selective sweeps using BovineSNP50 data**

Twenty eight genomic regions on 15 chromosomes were identified as putatively harbouring selective sweeps (Table 2.2). Selective sweeps were found in all 14 breeds; however, breed-specific selective sweeps were not identified in every breed. Twenty three predicted sweeps were breed-specific and 5 were shared among two to seven breeds. Four sweep regions were common to at least four breeds (Figure 2.1). Breed-specific sweeps averaged  $336 \pm 119$  kb and ranged in size from 207 to 702 kb but were not

different in size ( $P < 0.19$ ) to sweeps common to two or more breeds which when calculated for each breed (not the common core size of overlap only) averaged  $441 \pm 222$  kb and ranged in size from 215 to 866 kb. Common sweeps overlapped but did not have identical boundaries in all breeds, however, the haplotypes found at the core loci in each of these sweeps were identical for each of the breeds in which the sweep was detected. The average number of selective sweeps found in Charolais, Hanwoo, Salers, Brown Swiss and Jersey (all with 78 or fewer animals) was 3.8 while the average number detected in Hereford, Angus and Holstein (all with at least 812 animals) was 4.0 (Table 2.2) suggesting that variation in sample size did not play a significant role in elevating the false positive rate in the breeds with small sample sizes.

Three of the five selective sweep regions detected in two or more breeds involved both beef and dairy breeds, whereas the 358 kb region on BTA12 is common to only the Angus, Salers, Shorthorn and Simmental beef breeds. None of the five selective sweeps shared by two or more breeds are phylogenetically congruent in the sense that we might have expected the sweep to have arisen in a recent common ancestor [37]. While the large selective sweep region on chromosome 6 at  $\sim 75.9$ -76.7 Mb is shared by the closely related Salers and Brown Swiss breeds, Salers and Limousin are sister breeds [6, 7] and Limousin does not demonstrate evidence of this sweep. This suggests that the mutations were independently selected within these breeds despite the complex history of intercrossing that occurred during breed development [6, 7]. There were no putative sweeps shared in common between any of the taurine breeds with the indicine Brahman breed. The DAVID functional analysis did not yield any significant functional enrichment of

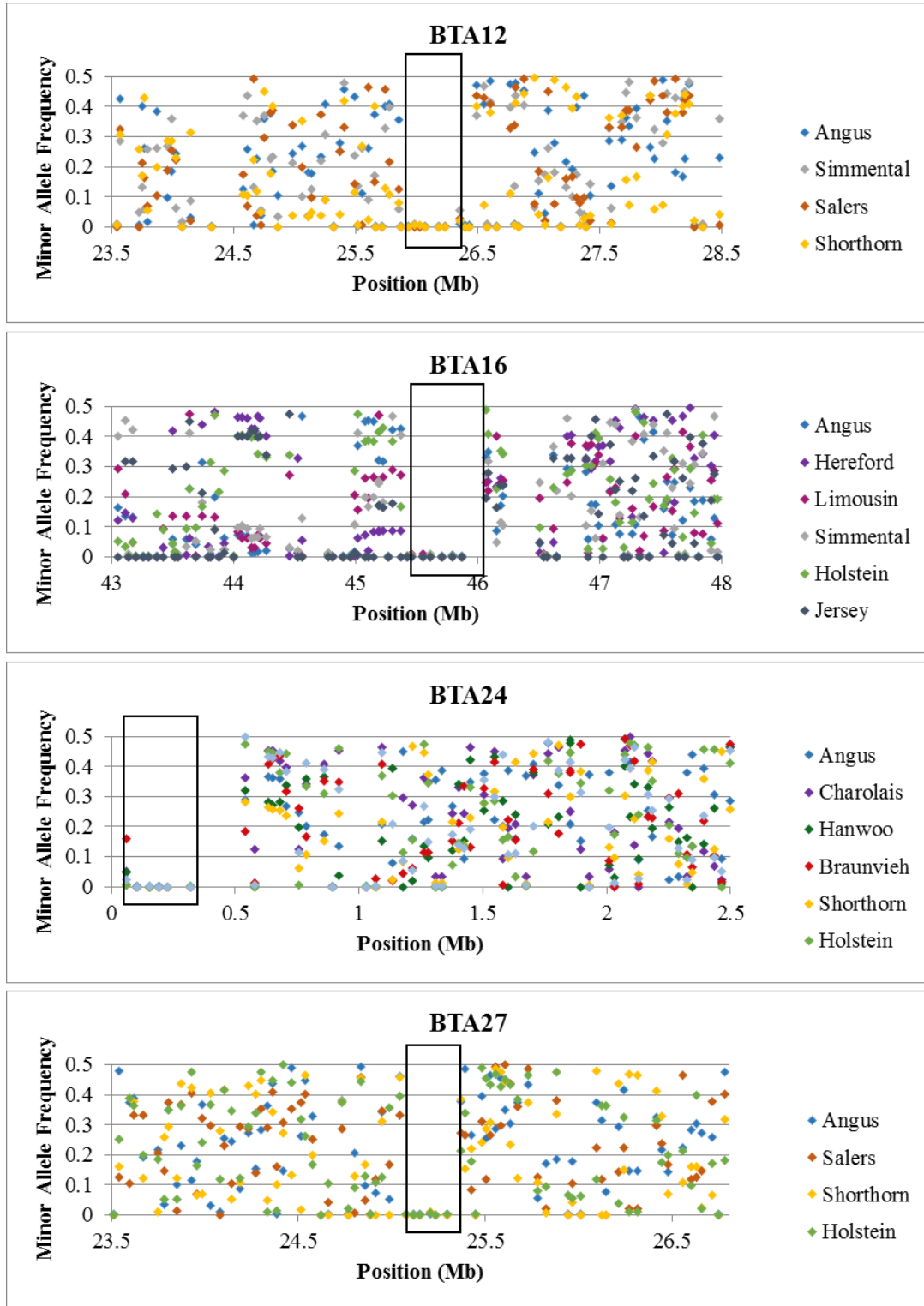
gene ontology terms for these sweep regions suggesting that each of the sweeps were based on functionally independent variants that influenced distinct phenotypes.

### **Regions identified as harbouring selective sweeps using AFFXB1P data**

A total of 85 putative selective sweep regions spanning from 200 to 846 kb and averaging  $321 \pm 132$  kb were identified on 28 of the 29 bovine autosomes in the five genotyped breeds and, of these regions, 20 were shared in two or more breeds (Tables 2.3 and 2.4). These regions harboured from 20 to 477 contiguous SNPs with no more than 5% of the SNPs being variable. Among the selective sweeps identified in two or more breeds, the number of breeds included in this analysis was too small to make inferences about the phylogenetic congruence of shared sweep regions, however, three sweeps were found only in the closely related East Asian Wagyu breeds and, of these regions, 20 were shared in two or more breeds (Tables 2.3 and 2.4). These regions harboured from 20 to 477 contiguous SNPs with no more than 5% of the SNPs being variable. Among the selective sweeps identified in two or more breeds, the number of breeds included in this analysis was too small to make inferences about the phylogenetic congruence of shared sweep regions, however, three sweeps were found only in the closely related East Asian Wagyu and Hanwoo breeds. All of the breeds that shared a common selective sweep (Table 2.4) were fixed for the same core haplotype with two exceptions. The sweep on BTA16 at 45,386,065-45,652,672 and on BTA21 at 1,727,412-2,142,823, both shared by the Angus and Simmental breeds, were fixed for haplotypes that differed at a single SNP. The allele that varied in the haplotypes that were swept to fixation on BTA16 was the 7<sup>th</sup> of 248 SNPs, whereas on BTA21 the variable allele was at the 22<sup>nd</sup> of 30 SNPs, indicating a

**Figure 2.1 Selective sweep regions discovered in the analysis of the BovineSNP50 data that were predicted to be common to two or more breeds.**

Regions identified as harbouring commonly selected haplotypes are indicated by the near-zero MAF values and are indicated by black boxes.



**Table 2.2 Putative selective sweep regions identified by analysis of Bovine SNP50 genotypes**

Breed	BTA	UMD3.1 Coordinates (bp)	# SNP	Size (bp)
Angus <sup>1</sup>	1	1,712,261-2,013,659	11	301,398
Limousin	2	5,974,885-6,344,425	7	369,540
Brown Swiss	2	73,436,684-73,978,469	10	541,785
Hanwoo <sup>1</sup>	3	19,860,064-20,175,158	5	315,094
Hanwoo <sup>1</sup>	3	112,997,892-113,219,287	5	221,395
Brown Swiss	4	61,098,696-61,376,132	9	277,436
Hereford	6	70,655,812-70,865,694	6	209,882
Brown Swiss	6 <sup>2</sup>	75,830,633-76,696,893	9	866,260
Salers		75,996,320-76,696,893	8	700,573
Jersey	6	105,390,830-105,730,372	8	339,542
Limousin	7	40,250,259-40,485,825	7	235,566
Jersey	7	45,439,468-45,828,427	10	388,959
Holstein	7	72,908,532-73,126,315	7	217,783
Brown Swiss	11	25,577,969-25,941,999	10	364,030
Jersey	12	1,113,009-1,436,093	8	323,084
Angus <sup>1</sup> , Salers, Shorthorn, Simmental <sup>1</sup>	12	25,878,820-26,236,394	7	357,574
Brahman <sup>1</sup>	12	36,287,533-36,989,957	14	702,424
Holstein	13	15,456,721-15,683,571	6	226,850
Braunvieh	14	28,674,493-28,960,475	5	285,982
Brown Swiss	14	42,739,573-43,140,076	9	400,503
Braunvieh	16	39,714,165-39,898,049	3	389,103
Salers	16	43,200,419-43,618,684	9	418,265
Jersey	16	45,376,614-45,874,144	8	497,530
Angus <sup>1</sup> , Holstein, Limousin, Simmental <sup>1</sup>		45,425,579-45,874,144	7	448,565
Hereford		45,464,423-45,874,144	6	409,721
Hanwoo <sup>1</sup>	16	52,535,473-52,742,523	4	207,050
Hanwoo <sup>1</sup>	18	14,526,709-14,847,049	5	320,340
Limousin	19	37,301,353-37,520,214	6	218,861
Brahman <sup>1</sup>	19	43,376,032-43,835,219	10	459,187
Angus <sup>1</sup> , Holstein, Shorthorn	24	63,576-320,143	6	256,567
Braunvieh, Charolais, Finnish Ayrshire, Hanwoo <sup>1</sup>		105,589-320,143	5	214,554
Angus <sup>1</sup> , Holstein, Salers, Shorthorn	27	25,075,449-25,295,935	6	220,486

<sup>1</sup>These breeds were also genotyped with the AFFXB1P assay and allow the potential validation of these putative selective sweep regions.

<sup>2</sup>Call rates for some SNP within this region did not exceed the 85% threshold that was applied to the remaining data.

**Table 2.3 Putative selective sweep regions detected in at least two breeds using AFFXB1P data**

Breeds	BTA	UMD3.1 Coordinates (bp)	Size (bp)	# SNPs Total	# SNPs Fixed	# SNPs Variable
Angus	6	5,639,799-5,993,214	353,415	95	92	3
Simmental, Wagyu, Hanwoo		5,639,799-5,993,214	353,415	95	94	1
Angus, Simmental, Wagyu	6	106,844,116-107,308,280	464,164	21	20	1
Simmental	7	4,344,221-4,677,474	333,253	22	21	1
Hanwoo		4,443,937-4,838,781	394,844	90	88	2
Angus	7	51,000,141-51,430,242	430,101	406	391	15
Simmental		51,144,109-51,788,442	644,333	477	469	8
Wagyu	14	2,098,182-2,588,143	489,961	29	29	0
Hanwoo		2,097,493-2,603,935	506,442	31	30	1
Angus	16	44,463,682-44,881,229	417,547	213	203	10
Simmental		44,693,447-45,207,060	513,613	162	158	4
Angus	16	45,386,065-45,652,672	266,607	248	238	10
Simmental		45,386,065-45,677,279	291,214	286	280	6
Simmental	16	49,400,423-49,679,673	279,250	22	21	1
Wagyu		49,400,423-49,661,831	261,408	20	20	0
Simmental	16	52,629,624-52,857,759	228,135	30	30	0
Wagyu		52,629,624-52,857,759	228,135	30	29	1
Wagyu	17	69,501,174-69,703,717	202,543	116	112	4
Wagyu		69,953,291-70,183,461	230,170	145	140	5
Hanwoo		69,705,850-70,283,199	577,349	421	402	19
Simmental	17	73,619,837-73,975,539	355,702	24	24	0
Wagyu		73,761,834-74,325,363	563,529	57	55	2
Angus	18	14,725,181-14,973,411	248,230	107	102	5
Simmental		14,725,181-14,973,411	248,230	107	104	3
Angus	18	53,342,619-53,596,733	254,114	53	51	2
Simmental		53,342,619-53,573,919	231,300	46	44	2
Angus	19	57,106,999-57,377,040	270,041	26	25	1
Simmental		57,065,941-57,377,040	311,099	28	28	0
Angus	20	71,679,859-72,012,001	332,142	37	36	1
Simmental		71,780,338-72,012,001	231,663	29	28	1
Angus	21	2,134-742,281	740,147	422	408	14
Simmental		2,134-326,342	324,208	187	179	8
Angus, Simmental	21	1,727,412-2,142,823	415,411	30	29	1
Angus	21	70,948,810-71,284,393	335,583	23	22	1
Hanwoo		71,112,766-71,575,370	462,604	32	31	1
Angus	26	21,432-718,976	697,544	242	234	8
Simmental		21,432-723,266	701,834	250	244	6
Wagyu		21,432-225,284	203,852	30	29	1
Wagyu		249,534-454,031	204,497	81	77	4
Hanwoo		21,432-454,031	432,599	121	119	2
Simmental	29	1,156-558,130	556,974	155	153	2
Wagyu		105,179-536,644	431,465	122	120	2
Hanwoo		315,440-558,130	242,690	90	86	4



**Table 2.4 Genomic regions predicted to harbour selective sweeps using BovineSNP50 data and validated by AFFXB1P data.**

<b>BTA</b>	<b>Breed</b>	<b>UMD3.1 Coordinates (bp)</b>	<b># Fixed SNP</b>	<b># Variable SNP</b>	<b>Size (bp)</b>	<b>Annotations</b>
1	Angus	1,673,108-2,024,737	313	8	351,629	Horn-polled
13	Wagyu	15,493,906-15,739,251	301	4	245,345	DGKZ
16	Angus	45,386,065-45,652,672	238	10	266,607	ENO1, RERE
	Simmental	45,386,065-45,677,279	280	6	291,214	
16	Simmental	52,629,624-52,857,759	30	0	228,135	C16H1orf159, RNF223, FIN376, AGRN, ISG15, HES4, C16H1orf170, PLEKHNI, KLHL17, NOC2L, SAMD11
	Wagyu	52,629,624-52,857,759	29	1	228,135	
18	Angus	14,725,181-14,973,411	102	5	248,230	TCF25, SPIRE2, MC1R, TUBB3, MIR220D, DEF8, CENPBD1, LOC532875, DBNDD1, GAS8, LOC100296324, LOC100336472, SHCBP1
	Simmental	14,725,181-14,973,411	104	3	248,230	

conserved core at both loci. No sweeps were found in common between the four taurine breeds and the indicine Brahman.

Since only a subset of the 14 breeds genotyped with the BovineSNP50 assay were also assayed with the AFFXB1P assay, we had the potential to validate only 11 of the putative selective sweep regions identified in Table 2.2 and five regions were confirmed (Table 2.5). Only two of the regions were confirmed in the same breeds that led to their identification using the BovineSNP50 assay and for two of the remaining regions, discovery occurred in Hanwoo and confirmation occurred in the phylogenetically similar Wagyu breed [6, 7]. However, the region on BTA13 was identified in Holstein using the BovineSNP50 data but was independently validated in Wagyu by the AFFXB1P data.

#### **Annotation and casual candidates underlying selective sweep regions**

The putative selective sweep regions were found to harbour annotated bovine protein coding regions or human orthologues, conserved sequences likely to be regulatory elements, and pseudogenes. However, relatively few regions yielded genes likely to be selection candidates based upon identifiable phenotypes (Table 2.6). Five regions were associated with breed hallmarks such as coat colour and pattern or morphological characteristics. Several regions harboured olfactory receptor-like variants, or genes associated with neurological development or behavioural disorders as well as embryo patterning, survival, and development.

#### **Discussion**

We utilised two genotyping assays to identify putative selective sweep regions within the bovine genome. The BovineSNP50 assay was employed because we have genotyped a large number of registered animals from several breeds with this assay; however, we

recognise that the assay is not ideal for this purpose due to the ascertainment of common SNPs in its design. Since the *Bos taurus taurus* breeds in Table 2.1, and Angus and Holstein in particular, were used for SNP discovery and SNPs with high minor allele frequencies in these breeds were preferentially included during the design of the assay [14, 15] it is clearly unsuited to the identification of selective sweep regions that might be common among breeds. However, SNPs were included in the design of the BovineSNP50 assay if they were found to be variable in several, but not necessarily all of these breeds. Therefore, the assay theoretically possesses the ability to identify selective sweeps that are specific to individual breeds or to a small number of breeds. However, rather than characterising sweeps that occurred during the domestication of cattle and that should therefore be common, e.g., among European taurine breeds that descended from cattle that were domesticated in the Fertile Crescent, these sweeps are much more likely to have occurred during the formation of breeds and will reflect selection to fix phenotypes such as coat colour or the absence of horns within specific breeds.

A second limitation of this assay is that of calibration relative to the size of the sweep regions. While strong sweeps in numerically small populations are expected to result in the fixation of large haplotypes, weak selection in numerically large populations will result in the fixation of only a small core haplotype which may not be detected using this assay. Thus, historic variation in the census population size among breeds may have resulted in variation in the size of the fixed haplotype and our inability to detect small haplotypes. By requiring  $N$  contiguous loci to each have a minor allele frequency (MAF)  $< \alpha$ , for small  $\alpha$ , we must choose  $N$  to be sufficiently large that it would be highly unlikely to observe  $N$  contiguous loci all with a  $MAF < \alpha$  due to chance alone and yet

sufficiently small that the targeted sweeps are not smaller than  $37 \times (N-1)$  kb, where 37 kb represents the median intermarker interval on the BovineSNP50 assay. The design of the BovineSNP50 assay also led to lower average MAF and larger numbers of monomorphic loci in breeds such as Brahman, that are phylogenetically distant from the SNP discovery breeds [14]. To adjust for this bias, we defined N separately for each breed (Table 2.1) requiring larger N for breeds with larger numbers of monomorphic and low MAF loci. The definition of  $\alpha > 0$  is also important to this discussion since in the detection of sweeps we must allow for old sweeps in which *de novo* mutations may have begun to accumulate on the fixed haplotype, genotyping errors which are locus specific but average about 0.5% for this assay, and the incorrect ordering of loci by the UMD3.1 sequence assembly. However, errors in the assembly are vastly more likely to cause false negative than false positive sweeps by incorrectly introducing a variable locus into a region of dramatically reduced variability within the genome.

We also employed the AFFXB1P which contained almost 2.8 million putative SNPs that were screened for variability in a small number of animals from several breeds prior to the design of the commercial BOS 1 Axiom assay. While we had many fewer animals genotyped with this assay which influenced the estimation of MAF, the AFFXB1P assay had over 50 $\times$  the number of SNPs present on the BovineSNP50 assay which offered considerably greater power for identifying small sweeps and the application of this assay also suffers less from ascertainment bias. While loci that have been fixed in all domesticated cattle relative to their auroch forbears will still not appear on this assay due to the requirement that the putative SNP must have been predicted to have been variable in the sequence data for at least one breed, there was much less

selection for SNPs with high MAF in numerous breeds in the design of this assay relative to the BovineSNP50. Consequently, we expected this assay to identify putative sweep regions that could not be identified by the application of the BovineSNP50 assay and to more precisely define the boundaries of sweeps that were detected by the BovineSNP50 assay and validated by the AFFXB1P assay.

One of the main focuses of this study was to identify selective sweep regions common to multiple breeds. The existence of these shared regions dramatically reduces the potential of false positive discoveries since the likelihood of extended identical haplotypes being found in multiple breeds due to chance alone is close to nil. We identified a handful of regions in each dataset that met these criteria. The putative sweep shared among six breeds on BTA24 contains four pseudogenes, of which three are olfactory receptor-like. Whether any of these pseudogenes are expressed is unclear; however, olfactory receptor loci were detected as being recently duplicated within the bovine genome [38] suggesting that they may also be under strong selection for newly evolving functions. The common sweep region on BTA12 contains neurobeachin (*NBEA*) and mab21-like 1 (*MAB21L1*) which have been implicated in human autism and psychiatric disorders, respectively [39-41]. Since these phenotypes represent extreme behaviours, it is intriguing to speculate that mutations in these genes may also predispose cattle to increased docility and more favourable temperaments when handled by humans.

Of the regions identified in both datasets, a selective sweep on BTA13 was detected in Holstein using the BovineSNP50 data and was also discovered in Wagyu using the AFFXB1P data. The region from 15.49-15.74 Mb contains diacylglycerol kinase zeta (*DGKZ*) and several bovine ESTs. *DGKZ* has been implicated as a member of

the downstream leptin signalling pathway and reduced expression or activity within the hypothalamus has been associated with obesity [42]. The Holstein and Wagyu breeds are phylogenetically distant, however, Wagyu are believed to have been influenced by several European taurine breeds, including Holstein, during the late 1800s and both breeds are known for their ability to store intramuscular fat without accumulating excessive subcutaneous fat [43]. The 351 kb selective sweep region from 1.67-2.02 Mb on BTA1 found in Angus using the BovineSNP50 and validated in Angus using the AFFXB1P data contains a fixed 321 marker haplotype which harbours the *POLL* locus [10] a hallmark of the breed which contains only polled animals.

A region on BTA18 from 14.72-14.97 Mb was detected to harbour a selective sweep in Hanwoo cattle using the BovineSNP50 data and rediscovered in Angus and Simmental using the AFFXB1P data. This 248 kb region contains several annotated genes (Table 2.5), but importantly harbours melanocortin 1 receptor (*MC1R*) in which mutations lead to the black coat colour in cattle [11]. American Angus have been strongly selected for black coat colour and almost all registered animals in the American Angus Association are now homozygous black confirming that the basis for this selective sweep in Angus was for the black coat colour allele. The American Simmental Association registers crossbred animals and many breeders have developed Simmental × Angus crossbreds to capitalize on the premium that carcasses from black coated cattle can achieve if they qualify for Angus branded products. While all of the Simmentals genotyped with the BovineSNP50 assay were full-blood, of the 6 Simmental animals genotyped with the AFFXB1P assay, one was homozygous black, four were heterozygous and one was homozygous red. Furthermore, the Hanwoo animals are all

yellow coated. The fact that the sweep in Angus was not found using the BovineSNP50 data suggests a resolution issue with the requirement that at least 6 contiguous loci spanning at least 200 kb be fixed in order to declare a sweep. On the other hand, the fact that a sweep was detected in the AFFXB1P data for Simmental that was not detected using the BovineSNP50 data suggests a sampling issue since of the 12 chromosomes representing this breed, only 6 originated in Simmental for this region of the genome. The result for Hanwoo is more interesting since the sweep was declared in Hanwoo using the BovineSNP50 data for 48 individuals but was not confirmed in the 11 individuals genotyped with the AFFXB1P data. This suggests that either the region is not correctly assembled, or that an ancient breed foundation event may have occurred in which the yellow allele was fixed in this breed, but that sufficient mutation events have occurred on this *MC1R* haplotype to cause it to fail to be detected as a sweep using the ultra-high-density data. Finally, of particular interest is the fact that no sweep was identified in this genomic region in Wagyu cattle suggesting that black coat colour in Angus and Wagyu cattle may not be allelic. Recently, a mutation within beta-defensin 103 (*CBD103*) has been shown to cause black coat colour in dogs [44]. The cattle orthologue of *CBD103* maps to 4.89 Mb on BTA27 centromeric of the sweep that was detected in Wagyu cattle (Table 2.3).

After analysing regions identified in multiple breeds, we sought to identify any potential phenotypes under selection within breed-specific regions. Within the lower density BovineSNP50 data, we identified a sweep region towards the centromere of BTA1 harbouring 11 contiguous monomorphic SNPs and spanning 301 kb in Angus (Figure 2.2). This region contains the *POLL* locus [10] for which this breed has been

**Table 2.5 Putative breed-specific selective sweeps identified using AFFXB1P data**

<b>Breed</b>	<b>BTA</b>	<b>UMD3.1 Coordinates (bp)</b>	<b>Size (bp)</b>	<b># Total SNPs</b>	<b># Fixed SNP</b>	<b># Variable SNP</b>
Angus	1	1,673,108-2,024,737	351,629	235	231	4
	1	111,659,758-111,900,241	240,483	231	221	10
	3	28,196,188-28,405,853	209,665	207	199	8
	6	6,308,164-6,696,861	388,697	74	72	2
	8	70,780,263-71,227,757	447,494	24	24	0
	8	73,538,014-73,760,303	222,289	288	282	6
	11	99,717,097-99,929,710	212,613	63	60	3
	14	3,550,689-3,885,375	334,686	21	21	0
	14	24,631,146-25,173,007	541,861	387	376	11
	18	200,903-406,270	205,367	117	114	3
	20	70,827,025-71,040,113	213,088	30	29	1
	21	1,186,567-1,489,860	303,293	26	25	1
	21	3,091,822-3,417,143	325,321	308	301	7
	25	39,262,249-39,468,067	205,818	21	20	1
	27	36,633,324-36,896,534	263,210	135	130	5
	29	49,557,166-49,922,377	365,211	70	69	1
	Brahman	5	48,679,627-48,903,409	223,782	249	237
10		24,519,718-24,794,435	274,717	20	20	0
22		10,701,509-10,963,520	262,011	88	86	2
Hanwoo	9	1,384,189-1,586,988	202,799	214	211	3
	22	60,772,158-61,015,491	243,333	24	23	1
Simmental	1	83,467,837-83,685,705	217,868	124	118	6
	2	36,576,839-36,828,533	251,694	216	206	10
	2	119,873,146-120,084,764	211,618	158	151	7
	2	121,398,894-121,698,563	299,669	196	188	8
	5	68,631,784-68,857,802	226,018	276	267	9
	7	21,101,050-21,348,345	247,295	65	63	2
	7	52,426,108-52,673,522	247,414	40	38	2
	8	70,246,437-70,540,424	293,987	163	158	5
	10	59,147,189-59,382,774	235,585	190	182	8
	10	72,811,087-73,041,092	230,005	217	210	7
	13	12,236,039-12,524,368	288,329	382	369	13
	13	66,468,866-66,708,674	239,808	153	146	7
	15	82,007,368-82,208,624	201,256	32	32	0
	16	42,555,966-42,788,613	232,647	46	44	2
16	43,767,572-44,053,904	286,332	186	181	5	
17	13,291,922-13,492,010	200,088	141	136	5	
19	34,588,859-35,435,279	846,420	59	57	2	



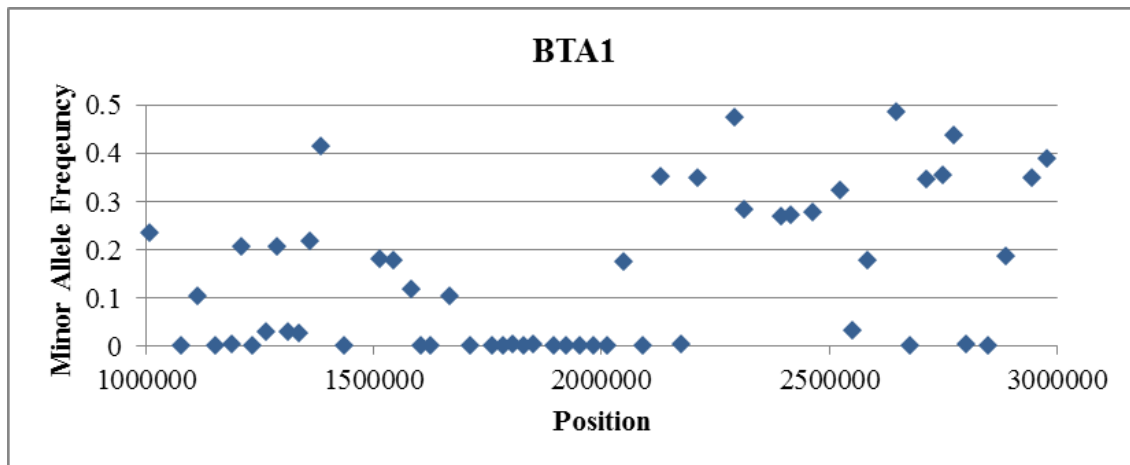
	22	2,788,408-3,001,597	213,189	216	214	2
	25	26,433,626-26,744,488	310,862	126	120	6
	25	42,202,212-42,775,697	573,485	21	21	0
	28	30,480,640-30,711,987	231,347	254	243	11
	29	39,045,837-39,396,761	350,924	93	91	2
	29	49,052,295-49,273,683	221,388	21	20	1
Wagyu	1	197,944-574,931	376,987	321	313	8
	1	12,544,028-12,840,164	296,136	335	325	10
	2	71,299,739-71,508,079	208,340	177	169	8
	2	136,434,039-136,676,537	242,498	22	21	1
	3	12,027,611-12,227,923	200,312	65	64	1
	3	113,841,208-114,151,444	310,236	189	181	8
	3	117,817,862-118,017,993	200,131	114	110	4
	4	117,846,325-118,095,397	249,072	21	20	1
	4	119,437,108-119,832,270	395,162	26	25	1
	11	33,020,266-33,264,604	244,338	307	293	14
	11	59,513,020-59,783,144	270,124	282	269	13
	11	103,333,722-103,590,033	256,311	33	32	1
	11	105,596,531-105,828,115	231,584	23	22	1
	12	23,283-377,340	354,057	137	137	0
	13	15,493,906-15,739,251	245,345	305	301	4
	13	58,016,490-58,335,951	319,461	114	110	4
	16	51,275,719-51,519,519	243,800	39	38	1
	19	53,829,568-54,065,288	235,720	25	24	1
	22	49,088,027-49,456,146	368,119	102	98	4
	24	36,584,702-36,795,576	210,874	224	218	6
	27	17,911,399-18,114,646	203,247	304	289	15

---

strongly selected for homozygosity of the *POLL* allele [45]. Hereford cattle are also homozygous for the dominant *spotted* allele at the *spotted* locus which is a candidate for the 210 kb sweep region at 70.65-70.87 Mb on BTA6 [12]. The *spotted* locus affects the white points on the face, underline, feet and tail which are a characteristic of the Hereford breed. These breed-specific sweeps are clearly examples of strong selection on loci which underlie phenotypes that are hallmarks of certain breeds and where the underlying causal mutation is known or has been mapped to a chromosomal location.

**Figure 2.2 Selective sweep surrounding the POLL locus in Angus cattle**

The selective sweep region on BTA1 in Angus is from ~1.7 Mb to 2.0 Mb and contains the *POLL* locus. The locus is contained within an extended region of reduced diversity relative to the up- and down-stream SNPs.



Among other selective sweep regions identified using the BovineSNP50 data, several contain no annotated genes. Either this reflects the incomplete annotation of the bovine genome, or the fact that the selected functional mutation within each of these regions is not located within a protein coding gene. Examining each of these regions for the alignment of human, pig and sheep mRNA orthologues failed to identify any genes. Recent work has identified ncRNAs which regulate the expression of nearby genes [46]

and may help identify candidates for the mutations in these regions that were subjected to selection. A region on BTA29 found in Hanwoo, Simmental, and Wagyu breeds was noted as containing ncRNA regulatory activity (Table 2.4).

Among the other detected breed-specific sweep regions, there were several regions linked to known phenotype to genotype associations. A region on BTA14 specific to Angus cattle harbours several genes including *PLAG1* which has been associated with variation in human height [47], the stature of cattle [48] and with vertebrae number and back elongation in swine [24]. Angus has been strongly selected for growth and frame size during the last 30 years [49] perhaps creating this selective sweep. The telomeric sweep region on BTA1 found in Angus contains several genes, including *LEKRI* and *CCNLI* in which mutations have been associated with reduced birth weight in humans [50]. Angus cattle have recently been selected to reduce birth weights to ease dystocia [37]. Another region of particular interest, and perhaps the most interesting identified within the indicine sample, was detected in Brahman and located at 48.68-48.90 Mb on BTA5 and contains the methionine sulfoxide reductase B3 (*MSRB3*) gene which has previously been identified as a candidate for a QTL affecting ear floppiness and morphology in dogs [51, 52]. Brahman cattle were developed in the US as a cross between the *Bos taurus indicus* breeds Guzerat, Nelore, Gir and Indu Brazil imported primarily from Brazil but all originating in India and the *Bos taurus taurus* Shorthorn and Hereford breeds [53]. There is considerable variation among these breeds for ear length and morphology with Indu Brazil animals having particularly large, pendulous ears. Thus, the sweep in this region may reflect strong recent selection by breeders to establish a specific Brahman ear morphology type. Key fitness traits such as

behaviour and reproduction are postulated to underlie the sweeps detected in multiple regions on chromosomes 6, 11, 12, 22, 25, and 28 (Table 2.6). The link between genes involved in psychiatric disorders poses a potential link to selection for improved temperament in cattle. Mutations in these behaviour-associated genes may confer improved temperament when cattle are handled by humans and these would have been strongly selected following breed formation to develop more manageable animals. The presence of sweeps in regions harbouring genes associated with reproductive processes may be a result of the selection for mutations which enhance reproductive rate or that result in improved calving ease.

Perhaps the most interesting and relevant trend in the data across breed-specific and common selective sweeps identified is the presence of a number of genes related to immune function and response. These regions are significant as they point to adaptive evolution within the bovine genome most possibly as a result of domestication. Human management of the species likely introduced pathogens new to the cattle population and selection occurred for individuals with beneficial alleles for host defense. Adaptive evolution of the immune system has been seen in many species such as *Drosophila*[54, 55] and humans[13], as well as plants[56]. Demonstrating a similar trend in cattle is interesting and indicative of the changes that human interaction and domestication imposed upon the species.

The abundance of olfactory receptor (OR) genes and pseudogenes within sweep regions is intriguing and suggests that olfactory loci play a major role in the domesticability of species. Olfactory receptor genes have previously been found to have been under selection in cattle [12, 25] and more recently in swine [57] and it has been

**Table 2.6 Potential causal genes underlying selective sweep regions and their associated function or phenotype**

SNP50 Discovery Breed(s)	AFFXB1P Discovery Breed(s)	BTA	Sweep Position (Mb)	Potential Causative Gene(s)	Gene or Region Type*	Associated Function or Phenotype
<b>SNP50 Specific Sweep Regions</b>						
Hereford	---	6	70.65-70.87	<i>spotted</i> <sup>1</sup>	--- <sup>1</sup>	white points on face, underline, feet & tail in cattle <sup>[12]</sup>
Angus, Salers, Shorthorn, Simmental	---	12	25.87-26.23	<i>NBEA</i> <i>MAB21L1</i>	PC PC	autism spectrum disorder in human <sup>[40, 41]</sup> psychiatric disorders in human <sup>[39]</sup>
Angus, Hanwoo, Hereford, Holstein, Limousin, Simmental	---	24	0.06-0.32	<i>LOC10013911</i> <i>LOC783097</i> <i>LOC100337172</i>	PS PS PS	olfactory receptor-like variants olfactory receptor-like variants olfactory receptor-like variants
<b>Co-Discovered Regions (SNP50 and AFFXB1P)</b>						
Angus	Angus	1	1.67-2.02	<i>POLL</i> <sup>2</sup>	--- <sup>2</sup>	presence or absence of horns in cattle <sup>[9, 10, 45]</sup>
Holstein	Wagyu	13	15.49-15.74	<i>DGKZ</i>	PC	leptin signalling – linked to human obesity <sup>[42]</sup>
Angus, Hereford, Jersey, Holstein, Limousin, Simmental	Angus, Simmental	16	45.39-45.65	<i>RERE</i>	PC	embryogenesis and embryonic survival <sup>[58-60]</sup>
Hanwoo	Simmental, Wagyu	16	52.63-52.86	<i>AGRN</i> <i>HES4</i> <i>ISG15</i>	PC PC PC	cell adhesion between embryo & maternal tissues <sup>[61]</sup> oscillation and repression of embryogenesis <sup>[62]</sup> acknowledgement & maintenance of pregnancy <sup>[63, 64]</sup>
Hanwoo	Angus, Simmental	18	14.72-14.97	<i>MC1R</i>	PC	black coat colour in cattle <sup>[11]</sup>
<b>AFFXB1P Specific Sweep Regions</b>						
---	Simmental	1	83.46-83.68	<i>HTR3E</i> <i>HTR3C</i>	PC PC	serotonin receptors – social cognition (tameness) in foxes and dogs <sup>[65-</sup>

68]						
---	Angus	1	111.66-111.90	<i>LEKRI CCNLI</i>	PC	reduced birth weight in humans <sup>[50]</sup>
					PC	reduced birth weight in humans <sup>[50]</sup>
---	Angus	3	28.19-28.40	<i>TSPAN2</i>	PC	oligodendrocyte signalling & maturation <sup>[69, 70]</sup>
---	Brahman	5	48.68-48.90	<i>MSRB3</i>	PC	ear floppiness and morphology in dogs <sup>[51, 52]</sup>
---	Angus, Hanwoo, Simmental, Wagyu	6	5.64-5.99	<i>MAD2L1</i>	PC	proper onset of anaphase during cell division
---	Angus, Simmental, Wagyu	6	106.84-107.30	<i>RGS12</i>	PC	positional candidate for ovulation rate in swine <sup>[71]</sup>
---	Wagyu	11	33.02-33.26	<i>NRXN1</i>	PC	synaptic function – linked to autism spectrum disorders <sup>[72, 73]</sup>
---	Simmental	13	66.46-66.70	<i>SAMHD1</i>	PC	innate immune response <sup>[74]</sup>
---	Angus	14	24.63-25.17	<i>PLAG1</i>	PC	variation in human height <sup>[47]</sup> and cattle stature <sup>[48]</sup>
---	Angus, Simmental	16	44.69-44.88	<i>PIK3CD</i>	PC	immune response
				<i>SPSB1</i>	PC	immune regulation <sup>[75]</sup>
---	Simmental, Wagyu	17	73.76-73.97	<i>VpreB</i>	PC	B cell development
				<i>PRODH</i>	PC	schizophrenia <sup>[76]</sup>
---	Angus, Simmental	20	71.67-72.01	<i>IgCgamma</i>	PC	immune response
				<i>PDCD6</i>	PC	T cell receptor-induced apoptosis
---	Angus, Hanwoo	21	70.11-71.28	<i>AHNAK2</i>	PC	skeletal muscle fibre organization <sup>[77]</sup>
				<i>IGHE</i>	other	immune response
				<i>CRIP1</i>	PC	immune regulation <sup>[78, 79]</sup>
				<i>CRIP2</i>	PC	immune regulation <sup>[78, 79]</sup>
---	Brahman	22	10.70-10.96	<i>TRANK1</i>	PC	bipolar disorder in humans <sup>[80]</sup>
				<i>DCLK3</i>	PC	brain development <sup>[81]</sup>
				<i>GOLGA4</i>	PC	golgi
				<i>ITGA9</i>	PC	identity/structure <sup>[82]</sup>
						gamete interaction & reduced fertilization success <sup>[83]</sup>

---	Wagyu	22	49.08-49.45	<i>TLR9</i>	PC	innate immunity
---	Simmental	25	26.43-26.74	HSA16 <sup>3</sup>	HO	autism spectrum disorders <sup>[84-86]</sup>
---	Angus, Hanwoo Simmental, Wagyu	26	0.21-0.45	<i>OR5D18</i> various loci <sup>4</sup>	PC PS	olfactory receptor <sup>[87]</sup> olfactory receptor- like variants
---	Wagyu	27	17.91-18.11	<i>CBX1</i>	PC	epigenetic reprogramming of embryo/gene expression <sup>[88-90]</sup>
---	Simmental	28	30.48-30.71	<i>ADK</i>	PC	physiological state – blood flow, neurotransmission in brain <sup>[91]</sup>
---	Hanwoo, Simmental, Wagyu	29	0.32-0.56	<i>ANKRD26P3</i> <i>CCDC144C</i>	PS PS	ncRNA regulatory activity ncRNA regulatory activity

---

\*Gene and region types are defined as: PC = protein coding, PS = pseudogene, HO = human orthologue, other = known type, unspecific.

<sup>1</sup>The causal mutation has not yet been identified but the locus has been fine mapped to this region, see reference.

<sup>2</sup>The *POLL* locus in cattle has been identified but is characterised by allelic heterogeneity among breeds.

<sup>3</sup>This region is orthologous to human chromosome 16 (HSA16).

<sup>4</sup>There are 13 loci identified as olfactory receptor-like variants which have been determined to be pseudogenes.

hypothesized that pigs rely intensely on their sense of smell for scavenging. Alterations in the need for wild animals to search for food following their domestication may result in a relaxation of the need for purifying selection acting on these genes allowing them to freely evolve to gain new functions in odorant and tastant detection. In tetrapods, anywhere from 20 to 50% of OR loci exist as pseudogenes [92] and while it is not clear if these genes were ever functional, the acquisition of trichromatic vision has been postulated as facilitating the loss of OR genes [93]. On the other hand, cattle are dichromatic and yet still possess significant numbers of OR pseudogenes [92] some of which were found to have potentially been under strong selection in this study. If these pseudogenes lack functionality, we might have expected them to have been deleted from the genome or to have been significantly disrupted by mutation. However, as many as 67% of OR pseudogenes are expressed in human olfactory epithelium [94] suggesting that similar percentages of bovine OR pseudogenes are also expressed and that many of these loci are functional and rapidly evolving in copy number [38].

We found evidence for selective sweeps in genomic regions that were detected to have diverged between breeds using integrated haplotype scores (iHS) and  $F_{ST}$  statistics in the Bovine HapMap project [15]. Putative sweep regions overlapped on chromosomes 2, 11, 12, and 14 detected either by an extreme  $F_{ST}$  or iHS value. We found evidence for only one selective sweep region for which the Bovine HapMap project [15] identified divergence between breeds using  $F_{ST}$  statistics. Using the BovineSNP50 data, we found a putative sweep in Angus, Salers, Shorthorn and Simmental cattle on BTA12 in a region harbouring *NBEA* that was found to have differentiated among taurine breeds. The fact that only one such region was detected is not surprising since regions that have been



strongly selected for a derived allele in one breed are likely to be selectively neutral in other breeds which do not possess this allele, leading to small differences in flanking SNP allele frequencies and modest  $F_{ST}$  statistics. Large  $F_{ST}$  statistics imply divergent selection for alternate alleles within different breeds suggesting that there may be several mutations in *NBEA* that have been strongly historically selected in some breeds and that are currently under selection in others. None of the putative sweep regions detected in the Holsteins were concordant with regions detected to be responding to recent selection in Israeli Holsteins [95]. Rather than reflecting differences in the origin of the founders of the US and Israeli Holstein populations [96], this more likely reflects the fact that our study focused on the identification of loci where selection had driven the desirable allele to fixation, whereas the Israeli study focused on the identification of loci currently responding to strong selection.

Other regions within the current analysis were also found in previous cattle selection studies. Two breed-specific regions in Limousin on BTA2 at 5.97 Mb and BTA7 at 40.25 Mb (UMD3.0 positions) were identified in a Bayesian scan of West African cattle[30]. There were also genes of similar function found between the studies relating to serotonin receptors. Comparing the current study to an artificial selection analysis in dairy cattle[31], the sweep region for black coat colour (*MC1R*) was co-identified as well as a region overlapping on chromosome 14 at 24.63 Mb (UMD position) which they have no identified genes noted but this study found to harbour *PLAG1* relating to cattle stature[48]. It is interesting to note that between these two studies, the current one identified platelet-derived growth factor alpha polypeptide (*PDGFA*) while the dairy cattle analysis identified the receptor for this growth factor

(PDGFRA). Our findings are also consistent with another dairy cattle study on complex traits where many reproductive and exterior traits were found to have significant evidence of selection via iHS test statistics[97]. With regions overlaps and gene similarity between these previous studies and what has been identified here, there is evidence to validate results noted herein. The types of genes underlying the previously regions also are consistent with the findings of this selective sweep analysis and show the selection pressures affecting cattle in many different geographical and biological type populations.

No putative selective sweep regions were found in common between Brahman and any of the *Bos taurus taurus* breeds which likely reflects the recent admixture that occurred in the formation of the Brahman and the fact that the breed does not share a common phenotype such as coat colour with any of the taurine breeds. Furthermore, indicine cattle are more commonly found in the southern tier of the US where they are exposed to higher temperatures and humidities and lower pasture qualities and availability than are taurine cattle which are more frequently found in the northern US. Consequently, we would not expect these breeds to have been subjected to selection for common morphological or adaptive phenotypes. Additionally, no common sweeps were detected between the cattle sub-species possibly due to the more severe ascertainment bias on MAF for BovineSNP50 loci in Brahman cattle. While SNP discovery was performed in Brahman for the development of the AFFXB1P assay, the number of indicine breeds sequenced for SNP discovery was small relative to the number of sequenced taurine breeds leading to a bias towards SNPs common in taurine cattle being included on the assay. However, the density of SNPs on this assay is so great (~1 SNP / kb) that we did not expect the reliability of sweep regions identified in Brahman to be

significantly less than those identified in taurine breeds. Our identification of a putative sweep region harbouring a previously identified QTL for ear length and floppiness in Brahman is consistent with the introduction of the undesirable allele from Indu Brazil cattle during breed formation and subsequent strong selection by breeders to remove the allele and fix a shorter ear type within the breed.

High-density assays, such as the BovineSNP50, have previously been shown to be adequate for the identification of runs of homozygosity (ROH) and for estimating inbreeding coefficients within cattle breeds [98]. However, this is not the case for the detection of selective sweeps which typically span smaller regions of the genome than ROH which are frequently due to consanguinity and which may represent as much as 12.39% of the genome [98]. We found high-density (~50,000) SNP data to be generally inadequate for the detection of selective sweeps due to the poor calibration of SNP density relative to the size of the targeted sweep regions. Relaxation of the number of contiguous SNPs with fixed alleles in order to detect smaller sweep regions leads to an elevation of type I error rate. Strong, recent selective sweeps causing the fixation of large haplotypes may be identified using high-density SNP panels, however, older sweeps which have accumulated new mutations and weak sweeps which have resulted in the fixation of relatively small haplotypes will not be detected.

Several putative selective sweeps identified using the BovineSNP50 data failed to be validated using the AFFXB1P data. While many fewer animals representing each breed were genotyped with the AFFXB1P assay, among the 50× additional SNPs within each such region, we found that at least 5% of the SNP had a  $MAF > (2M)^{-1}$  where M is the number of genotyped animals. We have previously found that the genotyping error

rate of loci on the Affymetrix Axiom BOS 1 assay is very similar to that of the Illumina BovineSNP50 assay (~0.5%, data not shown) and thus, we do not expect genotyping errors to explain this result, although it is certainly a possibility. It appears that the phenomenon is either due to errors in the genome assembly or mapping of probes for the AFFXB1P loci, or is simply due to type I errors. As a consequence, the reliability of declaration of a selective sweep is dramatically improved when sweeps are found to be common between breeds, particularly when the breeds are phylogenetically distant. We found several sweep regions that were common to two or more breeds and five sweeps predicted from the BovineSNP50 data were validated by the AFFXB1P data.

Identifying the mutations that underlie these sweep regions will be paramount to more fully understanding the effects of human interaction on the genomes of domesticated cattle. Candidates will soon become available by sequencing the genomes of individuals that are homozygous for identical SNP haplotypes within a sweep region but where some originate from the breeds predicted to have undergone a selective sweep and the others from breeds in which no sweep was detected. However, even after these mutations have been identified, our understanding of the phenotype that was created and selected to complete fixation may still be limited. The functional analysis of genes within the selected regions sheds little light on this since each mutation within these genes may lead to unpredictable phenotypes. Finally, while our sampling of breeds was small, we found little evidence for the sharing of sweeps among phylogenetically closely related breeds. This further supports our conjecture that the design of the BovineSNP50 assay to include common variation makes it primarily suitable for the detection of sweeps that have occurred following breed formation.

## Conclusions

We identified selective sweeps that primarily appear to have occurred following breed formation events. Due to the constraint that SNPs be variable in multiple breeds which was imposed during the design of both of the utilized assays, we did not identify any sweeps that were common to all breeds within the study. There were also no sweep regions predicted to be in common between breeds of taurine and indicine descent probably reflecting the different environmental and demographic forces to which these sub-species have been exposed during breed formation. For several of the detected sweep regions we were able to identify the phenotypes and genes that were subjected to selection, or to propose these based upon the results of previous mapping studies. However, for many of these regions the selected gene and phenotype are unclear. The fact that so many of the detected sweep regions harbour genes associated with behavioural characteristics, immunity, reproductive processes, or embryonic development is probably not remarkable considering the fact that strong selection acts on these fitness traits and that the time required to achieve fixation of variants of modest effect may be considerably longer than the 200 years since breed formation during which strong human selection has acted.

We demonstrate that the resolution and SNP ascertainment bias inherent in the design of the assay used to detect selective sweeps is of paramount importance and that the BovineSNP50 assay is not generally suitable for this purpose due to the high type I error rates that are likely to be encountered. SNP ascertainment bias leads to lower MAF in breeds that are phylogenetically distant from the SNP discovery breeds and an increased rate of monomorphic SNPs within these breeds. As whole genome sequencing

becomes less expensive, these problems will likely be ameliorated by sequencing a few distantly related individuals from each breed and this approach may also be used to identify the candidate mutations which underlie each sweep. However, the approach is reliant on the alignment of sequences to a draft Hereford reference assembly [38] which introduces a new set of biases unless *de novo* sequence assemblies can accurately be created for each breed.

The identification of genes and variants underlying historical selective sweeps is of interest from the perspective of understanding how human interaction with cattle has influenced the patterning of variation within the bovine genome. Perhaps of more importance, the discovery of the selected variants will lead to the identification of large effect QTL and ultimately a better understanding as to the phenotypes which are affected by variation within genes and regulatory elements.

### **List of Abbreviations**

Abbreviations are defined in the text.

### **Competing interests**

The authors declare that they have no competing interests.

### **Authors' contributions**

JFT, JED and HRR designed the experiment. HRR, RDS and JFT analysed data. SDM, MMR and RDS extracted DNA. MMR and SDM prepared samples for genotyping, SDM ran the Illumina assay, and RDS genotyped samples and managed the genotype database. HRR and JFT wrote the manuscript. All authors read and approved the final manuscript.

## **Acknowledgements**

We gratefully acknowledge the provision of semen samples from breeders and semen distributors. This project was supported by the University of Missouri, National Research Initiative grants number 2008-35205-04687 and 2008-35205-18864 from the USDA Cooperative State Research, Education and Extension Service and National Research Initiative grant number 2009-65205-05635 from the USDA National Institute of Food and Agriculture.

### **3. A survey of epistasis in growth and carcass traits of Angus steers**

**Holly R. Ramey<sup>1</sup>, Jesse L. Hoff<sup>1</sup>, Polyana C. Tizioto<sup>1,2</sup>, Robert D. Schnabel<sup>1</sup> and**

**Jeremy F. Taylor<sup>1,§</sup>**

<sup>1</sup>Division of Animal Sciences, University of Missouri, Columbia MO 65211 USA

<sup>2</sup>Embrapa Southeast Livestock, São Carlos, SP, Brazil

<sup>§</sup>Corresponding author

Email addresses:

Holly R. Ramey ([hrrxb2@missouri.edu](mailto:hrrxb2@missouri.edu))

Jesse L. Hoff ([jlh4df@missouri.edu](mailto:jlh4df@missouri.edu))

Polyana C. Tizioto ([polytizioto@yahoo.com.br](mailto:polytizioto@yahoo.com.br))

Robert D. Schnabel ([schnabelr@missouri.edu](mailto:schnabelr@missouri.edu))

Jeremy F. Taylor ([taylorjerr@missouri.edu](mailto:taylorjerr@missouri.edu))

#### **Corresponding Author:**

Dr. Jeremy F. Taylor

S135B Animal Sciences

920 East Campus Drive

University of Missouri

Columbia, MO 65211-5300

E-mail: [taylorjerr@missouri.edu](mailto:taylorjerr@missouri.edu)

Tel.: +1-573-884-4946

Fax: +1-573-882-6827



## **Abstract**

Genome-wide association studies have identified numerous loci associated with complex traits using single locus models of association. While the identified quantitative trait loci (QTL) reveal the existence of genes with additive effects on phenotype, it is unclear whether these loci may participate in epistatic interactions which may modify the effects of QTL. Of particular interest in livestock is the finding that QTL are often detected in a population-specific manner. While this may be due to differences among populations in allele frequencies at the QTL or flanking single nucleotide polymorphisms (SNPs) used to detect the QTL via association, it may also be due to epistasis where differences in allele frequencies among the populations at modifying loci attenuate the effect of QTL. We performed a genome-wide scan for pairwise interacting loci influencing 12 growth and carcass phenotypes in 3,958 commercial Angus steers using 46,925 autosomal SNPs. Significant epistatic interactions and networks were identified for seven traits – birth weight, weaning weight, weight at ultrasound, hot carcass weight, fat thickness, ultrasound intramuscular fat, and ultrasound fat thickness. Of particular interest, we found a core of three SNPs spanning 53,900 bp on BTA2 to anchor epistatic networks for both weaning and ultrasound weight. The finding that epistatic effects on animal growth persist between correlated phases of beef production provides support that the detected effects are biologically valid and not type I errors. In general, the genomic regions identified as harbouring interacting loci had previously been identified as harbouring QTL with additive effects on growth and adiposity traits, and also contained genes that were significantly enriched in pathways and networks involved in connective tissue, skeletal, muscle and lipid development, function and disorders. This study provides

preliminary evidence for the existence of large, complex epistatic networks underlying growth and carcass traits of beef cattle. The potential for using these results in selection programs is unclear although for loci with significant additive by additive interactions, it may be possible to select for optimal homozygous genotype combinations at several loci using genotyped parents.

### **Keywords**

epistasis, pairwise interactions, beef cattle, complex traits, carcass and growth traits

### **Background**

Complex trait dissection using molecular markers has been now been performed within the agriculturally important meat species for nearly 25 years. Quantitative trait locus (QTL) scans within segregating families, and more recently, genome-wide association studies (GWAS) in relatively unrelated individuals have now revealed 13,145 QTL affecting 492 traits in cattle [99]. Typically, these studies have identified that relatively few QTL of large effect influence each trait and that cumulatively these QTL explain only a small proportion of the additive genetic variance in the trait. However, this problem of the “missing heritability” may, at least, partly involve the overestimation of additive genetic variance in population-based models which ignore the presence of epistasis [4]. However, the estimation of even two-locus epistatic interactions is problematic due to the need for large sample sizes of genotyped individuals to ensure that every possible genotypic combination is observed in the data. Even more problematic is the issue of multiple testing, both from the perspective of computational requirements and type I error rates since the number of pairwise comparisons scales quadratically with the number of genotyped markers. With the development of software designed to

capitalise on parallel distributed processing architectures and the availability of high density single nucleotide polymorphism (SNP) genotyping platforms, the importance of epistasis to complex trait variation can now be more thoroughly evaluated.

The widespread use of crossbreeding within the US beef industry is designed to capitalise on heterosis which may be explained by intra-locus interactions among alleles [100] but does not exclude the possibility of significant inter-locus interactions. Understanding the nature of epistatic interactions of large effect may allow for improved estimates of heritability, a better understanding of the population specificity of QTL effects and the design of mate selection programs to maximise the value of slaughter progeny while meeting consumer demands for quality beef. The identification of genic interactions of large effect will also increase our biological understanding of the processes of growth and development leading to a deeper understanding of the genetic networks and mechanisms that underlie variation in quantitative traits.

Two-locus models of epistasis have recently been examined for carcass composition traits in Duroc x Pietrain crosses [101], fatness traits in Large White x Meishan crosses [102] and meat quality traits in Meishan, European Wild Boar and Pietrain crosses [103]. Two recent studies have identified epistatic interactions as underlying piebald pigmentation in Merino sheep [104] and skin and fibre pigmentation in Merino x Awassi crosses [105]. While two beef cattle studies have identified two-locus interactions between specific genes as creating variation in meat and carcass traits [106, 107], no genome-wide scans for epistasis have yet been performed in beef cattle. In this study, the extent of pairwise epistatic interactions and nature of epistatic networks

affecting growth and carcass phenotypes were examined genome-wide using Illumina BovineSNP50 genotype data.

## **Materials and Methods**

### *Samples and phenotyping*

This study contained samples from 4,032 steers from the Circle A Angus Ranch headquartered in Iberia, Missouri. The ranch owns both registered and commercial cattle, but the commercial females have been bred to registered Angus bulls for several generations such that the commercial steers employed in this study can be considered to be very high percentage Angus in origin. Observations were collected on 12 growth and carcass traits (Table 3.1). All observations were corrected to phenotypes by adjusting for the estimates of effects of contemporary group, age at weaning, and birth order within calving season, which were generated by fitting a mixed linear model which also included an animal effect. The covariance among observations on related animals was explained using a genomic relationship matrix and additive genetic and residual variances were estimated by REML.

### *Genotyping and SNP filtering*

DNA samples from each of the steers were genotyped using the BovineSNP50 BeadChip assay (Illumina, Inc.) either at the University of Missouri (Columbia, MO) or at Neogen (Lincoln, NE). The original 55,074 markers on the genotyping panel were filtered for minor allele frequency (MAF < 1%), Hardy-Weinberg Equilibrium chi-square statistic ( $\chi^2 > 300$ ) and genotypes for the X chromosome were removed. Any animals with an overall genotype call rate of <90% were also removed. After filtering, there were 3,958 individuals each with N = 46,925 SNP genotypes and 1.02% of missing genotypes were

imputed using fastPHASE 1.2. All SNP coordinates correspond to the UMD3.1 cattle reference assembly.

**Table 3.1 Summary of analysed traits.**

Trait	Units	Abbreviation	Number of Phenotypes	Mean	$h^2$	$V_A$	$V_E$
Average daily gain – postweaning	lb/day	POSTADG	2,692	2.83	0.26	0.0297	0.0826
Birth weight <sup>Δ</sup>	lb	BW	3,281	84.21	0.37	30.42	51.10
Fat thickness <sup>Δ</sup>	in	FAT	3,749	0.56	0.30	0.0092	0.0210
Hot carcass weight <sup>Δ</sup>	lb	HCW	3,858	759.13	0.29	1246.7	3026.5
Marbling	units	MARB	3,854	493.65	0.46	3922.7	4686.1
Ribeye muscle area	sq in	REA	3,831	12.03	0.32	0.4545	0.9718
Ultrasound fat thickness <sup>Δ</sup>	in	USFT	2,547	0.36	0.47	0.0039	0.0043
Ultrasound intramuscular fat <sup>Δ</sup>	%	USIMF	2,551	5.22	0.34	0.3922	0.7631
Ultrasound ribeye area	sq in	USREA	2,550	11.48	0.32	0.3625	0.7873
Ultrasound weight <sup>Δ</sup>	lb	USWT	2,705	1060.3	0.34	2844.1	5480.1
Weaning weight <sup>Δ</sup>	lb	WW	3,261	484.75	0.24	755.41	2376.32
Yield grade	units	YG	3,814	3.34	0.35	0.1535	0.2857

<sup>Δ</sup>Traits for which significant pairwise epistatic interactions were detected.

#### *Identification of two-locus epistatic interactions*

The EPISNPmpi software was used to perform a genome-wide analysis for two-locus interactions for each trait [108]. EPISNPmpi is a parallel computing statistical package developed to analyse SNP data for additive (A) and dominance (D) effects on a SNP-wise basis as well as interactions among all  $N*(N-1)/2 = 1,100,954,350$  possible marker pairs. The tests performed for each interaction includes I (overall interaction), AxA (additive by additive), AxD/DxA (additive by dominance according to directionality of the interaction), and DxD (dominance by dominance). Statistical significance was

determined by the package based on a least squares analysis in which the single markers and potential marker-pair interactions were fitted. Because EPISNPmpi does not calculate type I error rates for each test adjusted for the multiple comparisons that are performed and also does not output the full set of  $P$ -values for all performed tests [108], we used an overall genome-wide significance level of  $P < 10^{-8}$ , based on generally accepted genomic community standards, to consider significance within the results (Table 3.2). The program was run using the HPC resources at the University of Missouri Bioinformatics Consortium.

EPISNPmpi generates output only for the most significant,  $L$ , single locus and,  $T$ , pairwise effects. In this analysis, the most significant  $T = 100$  and  $600$  pairwise interactions and  $L = 50$  and  $300$  individual locus effects were output and used for network formation (Table 3.2). The EPINET program within EPISNPmpi was used to generate all epistatic network maps. The top 10 networks, based on number of interacting nodes, were constructed where an individual locus represented a single node in the network. Networks were formed individually for each trait.

#### *Annotation and pathway analysis*

For each trait, regions of size 200 kb flanking each SNP involved in a significant interaction were used to identify genes for inclusion in the pathway and network analyses. Any gene that was either partially or wholly located within each 400 kb region was selected for analysis. In cases where a gene was identified as being involved in multiple epistatic interactions, the smallest  $P$ -value associated with the tagging SNP was associated with genes in the 400 kb region tagged by the SNP for the purposes of network analysis. For each trait, the genes tagged by SNP as being potentially involved in epistatic

interactions were analysed for common biological functions and a pathway analysis was performed using Ingenuity Pathway Analysis (IPA; Qiagen, Redwood City, CA). The UCSC Genome Browser was used for individual locus and regional annotation [35] and CattleQTLdb was used to retrieve QTL information [99].

## **Results and Discussion**

### **Software selection**

Several alternative software packages have been developed for performing genome-wide epistasis analyses [109-111], but were deemed unsuitable for the objectives of this study. In particular, we sought to interrogate every possible combination of pairwise locus interactions in a model which fit the individual SNPs and their interaction terms. Since this increases the number of computations quadratically over single locus analyses, computational efficiency is essential; however, to achieve this some programs use a data reduction step based on retaining only those SNPs that achieve global genome-wide significance in the analysis of individual SNPs [111]. This limits the ability to detect loci that behave epistatically to only those with individually large additive or dominance effects. Other programs were not as user-friendly, did not have a well-developed reference manual, contained SNP filtering procedures, or other features that were not aligned with the goals of this study. We chose EPISNPmpi for this analysis primarily because it was developed for parallel-computing environments and because it was designed to screen all pairwise interactions in large datasets involving tens of thousands of genotyped SNPs.

### **Single locus effects**

Additive + dominance or additive effects were detected for the 12 analysed traits. Only additive marker effects were detected in the top 50 most strongly associated single locus effects across the 12 traits. The proportion of variance explained by each of these loci was relatively low. These results are not unexpected and are consistent with other studies which find that there are relatively few QTL of large effect underlying quantitative traits of livestock [112-114].

### **Interaction effects**

Epistatic associations were found for groups of growth traits (BW, WW, USWT, and HCW) or fatness-related traits (FAT, USFT, and USIMF) (Tables 3.S1-7). All pairwise interactions were considered significant (Table 3.2). There was some overlap between traits for the location of interacting pairs but overall there were no patterns in which the same interacting loci were observed across the traits. While these traits were analysed separately, weight and fatness traits tended to be genetically and phenotypically correlated and we expect to find variants that have additive effects on weight throughout life. Likewise, we would expect to find epistatic interactions that affect weight at one stage of life to likely also affect weight at other stages of life in view of the strong phenotypic correlations among these traits. The estimates of genotype interaction effects for these traits were occasionally very large (Tables 3.S1-7) and are likely to be highly biased just as the most significant QTL effects detected in a GWAS are biased for their magnitude of effect [115]. In fact, these estimates are likely to be even more highly biased than additive QTL effects simply because the number of comparisons that are performed in pairwise epistasis analyses is orders of magnitude larger than for single



locus association analyses. The interactions that were detected as being significant must be validated in independent populations to arrive at unbiased estimates of multilocus genotype effects.

**Table 3.2 Range in significance values of pairwise interactions by trait**

Trait	Smallest p-value	Largest p-value – top 100 interactions	Largest p-value - top 600 interactions	Number significant interactions ( $p < 10^{-8}$ )	Number significant interactions ( $p < 10^{-9}$ )	Number AxA	Number AxD/DxA	Number DxD
BW	1.33E-12	1.93E-09	2.42E-08	312	76	219	15	10
WW	2.89E-11	2.30E-09	2.54E-08	413	38	132	101	110
HCW	2.50E-12	2.19E-09	1.84E-08	370	55	171	55	50
USWT	1.35E-11	9.61E-10	1.59E-08	332	44	117	70	77
FAT	1.71E-11	1.63E-08	9.69E-08	264	68	73	33	41
USFT	1.82E-11	2.50E-09	2.01E-08	438	100	66	74	110
USIMF	5.48E-11	2.22E-09	1.54E-08	66	6	29	10	9

### Network formation

The EPINET program was used to form networks from the pairwise associated SNPs. Of the networks formed for significant epistatic loci detected across the traits, the networks ranged in size from 2 to 30 nodes with all nodes being significant at the  $p < 10^{-8}$  level. These networks were largely anchored by a single or a small number of nodes with limited peripheral node interaction (Figures 3.1 and 3.2). The size of the network is not as important as the underlying functionality of the annotated genes within the networks; although the importance of small networks with only 2-3 nodes is difficult to assess and the most common network size was small with  $\leq 4$  nodes in most. To investigate whether the small number of network nodes was strongly dependent upon the number of interacting pairs output by EPISNPmpi, the top networks were recreated using the 600 most significant pairs of epistatically interacting loci. The median network size for HCW,

USIMF, USWT, and WW increased in size with the smallest of the networks now containing 5 nodes. The networks with the largest number of nodes for each trait also considerably increased in size (Figures 3.S1 and 3.S2). These networks maintained the core relationships detected using the 100 most significant interacting locus pairs but were expanded to include more peripheral nodes in most instances. On the other hand, some of the smaller networks with only 2-3 nodes were not expanded by this process and their overall ranking decreased resulting in their being removed from the top 10 largest networks. This suggests that the most important interactions of the network were detected for each trait and that changing the level of significance for accepting the significance of epistatic interactions had only a minor effect on network generation.

### **Weight trait networks**

Network generation for the four weight traits using the 100 most significant pairwise interactions resulted in the development of networks for WW, USWT, and HCW. In general, only the top ranked network in terms of the number of nodes generated for these traits was complex and these networks were multinodal and interchromosomal. Nodes that were central to each network were considered to be network anchors and were assessed for annotated genes within 200 kb up- and down-stream of the SNP coordinates defining the node as well as for overlapping QTL information reported in Animal QTLdb. The networks were then compared for overlap.

#### *Overlap between traits*

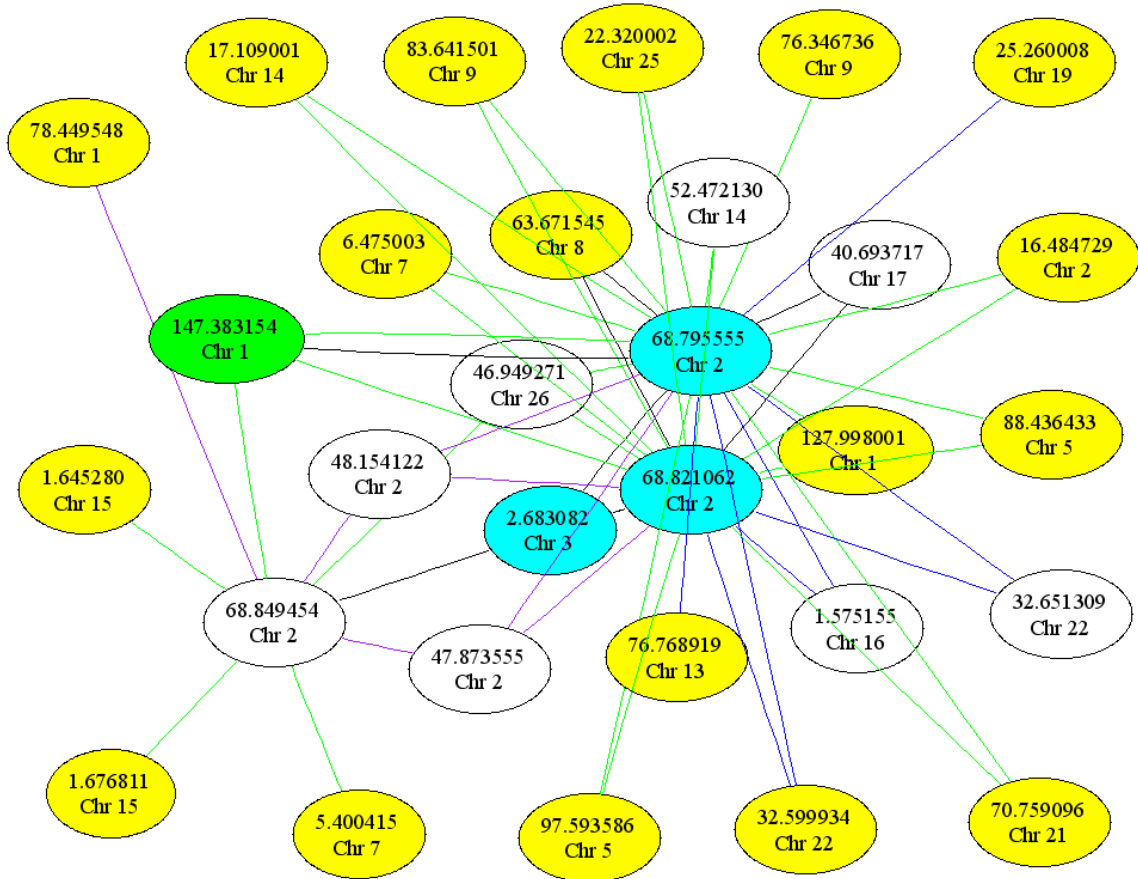
Overlap was found between the networks of epistatic genes affecting weaning weight and weight at ultrasound. The largest network for WW included 30 nodes representing regions on 17 chromosomes (Figure 3.1) while the largest USWT network had 27 nodes

containing regions on 16 chromosomes (Figure 3.3). Of the 57 nodes contained in the two networks, 5 overlap. These include BTA1 at 147,383,154 bp, BTA2 at 68,795,555, 68,821,062 and 68,849,454 bp and BTA19 at 25,260,008 bp. The SNPs on BTA2 anchor the WW and the USWT networks as the majority of pairwise interactions are specific to these three loci. The three markers are likely in strong linkage disequilibrium as they are consecutive SNPs and the two networks are therefore anchored at a single node which comprises a 53,900 bp region. Nevertheless, the networks also share two nodes on other chromosomes. This was not due to a lack of distinction among the phenotypes. The average age of animals at weaning was 195 d and at ultrasound measurement was 403 d. Thus USWT is approximately equivalent to yearling weight.

Within the three coinciding regions, meaning the node and flanking 200 kb regions for each locus, are two QTL influencing weaning weight located at 62.2-75.5 Mb on BTA2 and at 21.6-37.5 Mb on BTA19 [116]. Within the 200 kb regions flanking each of the SNPs in the network nodes we identified three cattle genes on BTA1. Five genes were identified in cow associated with the node on BTA19. The 53,900 bp region on BTA2 contains no annotated genes except for a human small nucleolar RNA (*SNORA80A*) and a Drosophila mRNA transcript variant (*CG5703*). The genes associated with the node on BTA1 are *PCBP3*, *COL6A1*, and *COL6A2*. This region contains two of the  $\alpha$ -chain collagen type VI genes, *COL6A1* and *COL6A2*, which are important within connective tissues for binding extracellular matrices to provide structural stability during tissue development and for the maintenance of homeostasis. The genes located within 200 kb of the node SNP on BTA19 include several ion channels (*P2RX1*), zinc finger

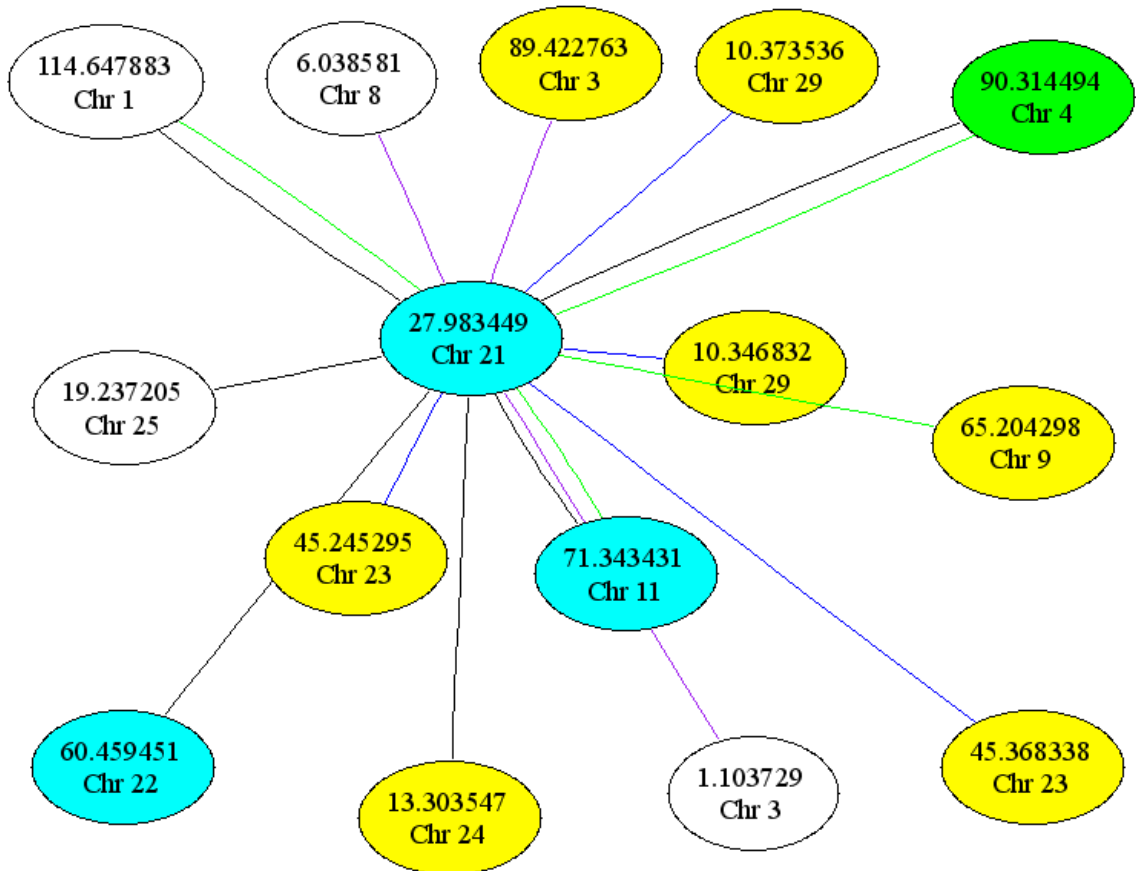
### Figure 3.1 Top epistatic network for WW

The largest epistatic network generated for weaning weight. The network contains 30 total nodes across 17 chromosomes with a set of three SNPs on BTA2 anchoring the network. These SNPs define a 53,900 bp region of chromosome 2. Connecting line colours signify the type of epistatic interaction: black = overall interaction, purple = additive x dominance, blue = dominance x additive, and green = dominance x dominance. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$ , green =  $p < 10^{-10}$ , white =  $p < 10^{-9}$ , and yellow =  $p < 10^{-8}$ .



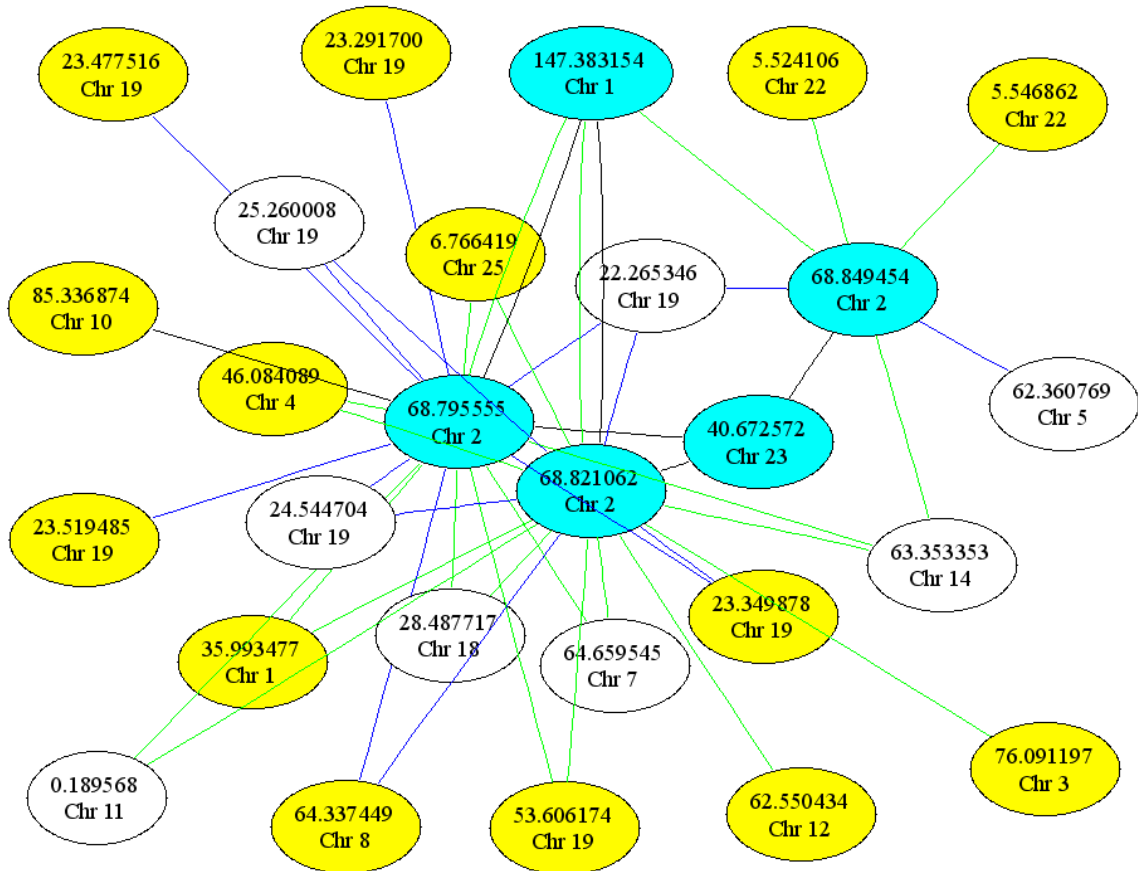
### Figure 3.2 Top epistatic network for HCW

The largest epistatic network for hot carcass weight. This network incorporates 15 nodes across 12 chromosomes. The anchoring SNP is located on BTA21 at 27,983,449 bp. Connecting line colours signify the type of epistatic interaction: black = overall interaction, purple = additive x dominance, blue = dominance x additive, and green = dominance x dominance. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$ , green =  $p < 10^{-10}$ , white =  $p < 10^{-9}$ , and yellow =  $p < 10^{-8}$ .



### Figure 3.3 Top epistatic network for USWT

The largest epistatic network for weight at ultrasound. The network is comprised of 27 nodes across 16 chromosomes. There is a set of three SNPs on BTA2 spanning 53,900 bp that anchor the network. Connecting line colours signify the type of epistatic interaction: black = overall interaction, blue = dominance x additive, and green = dominance x dominance. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$ , white =  $p < 10^{-9}$ , and yellow =  $p < 10^{-8}$ .



proteins (*ANKFY1*, *ZZEF1*), the calcium/calmodulin kinase cascade (*CAMKK1*) and *ATP2A3* which is associated with calcium sequestration for muscular excitation and contraction.

*Correlated traits with non-overlapping networks*

None of the 10 largest networks for BW or HCW shared any of the 5 core SNPs on BTA1, 2 and 19 that were common to the WW and USWT networks. However, when the 600 most significant pairwise interactions were considered, the three BTA2 SNPs were included in the third largest network for HCW. This was not the case for the BTA1 and BTA19 locus nodes with none of the five SNPs present within this network (Figure 3.S2). None of the BW networks contained any of these core loci when the networks were constructed from either the 100 or 600 most significant epistatic interactions. The networks generated for birth weight were also small with the largest network including 9 SNPs between 91,226,072 and 92,474,466 bp on BTA7 (Figure 3.S3). The remaining networks comprised two nodes involving SNPs on different chromosomes.

In contrast, the largest HCW network included 15 nodes representing loci on 12 chromosomes (Figure 3.2) and primarily comprised a mixture of AxD/DxA and DD effects. The anchor SNP is located on BTA21 at 27,983,449 bp. A region within the 200 kb flanking regions of this SNP is associated with a previously reported QTL for carcass weight (25.9-33.3 Mb) [116]. The 200 kb window flanking this locus contains four annotated genes which include microRNA mir-211 (*MIR211*), myotubularin-related protein 10 (*MTMR10*), methylmalonyl CoA epimerase (*MCEE*), and FANCD2/FANCI-associated nuclease 1 (*FANI*). As the names suggest, *MTMR10* could play a role in increased lean muscle mass of beef cattle as this gene is in a family of genes related to

those involved in myotubular development and myopathies [117]. Moreover the *MCEE* gene plays a role in fatty acid metabolism [118] and has been associated with SNPs suspected to underlie propanoate metabolism in pigs [119].

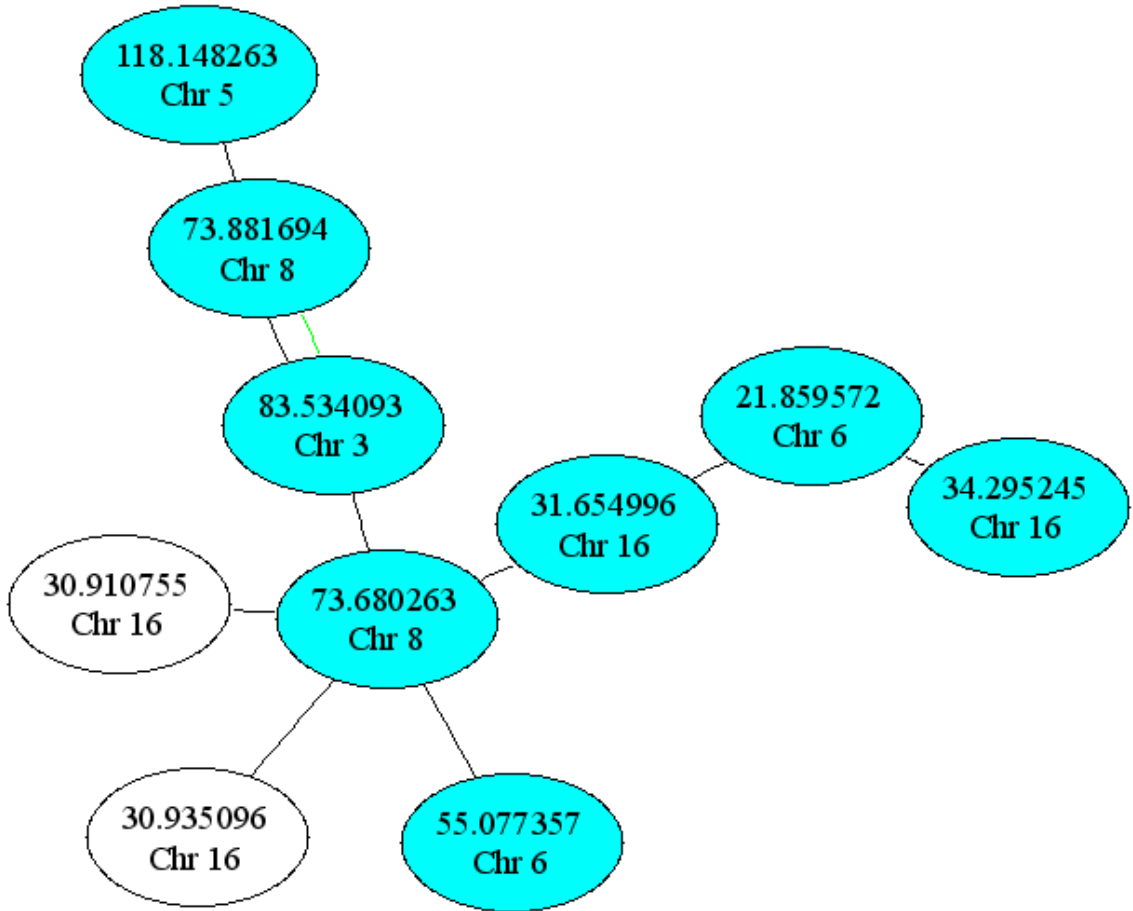
### **Fatness trait networks**

Animals either had fatness traits recorded at slaughter or they had estimates of these traits determined by ultrasound as live animals shortly before they were slaughtered. Median ages of the animals with ultrasound data (USIMF and USFT) and slaughter data (FAT) were 403 and 449 days, respectively. Pairwise epistatic interactions were found for USIMF and many had moderate to large estimated effects (Table 3.S17), however, the formed networks involved only three nodes representing loci on two chromosomes. From the 100 most significant epistatic pairwise interactions two significant networks were formed for FAT. Both networks primarily included overall interaction effects rather than interactions between A and D effects as in previous networks. However, again both networks involve interactions of moderate to large estimated effect sizes (Tables 3.S14 and 3.S16). The largest ranked network for FAT contained 10 nodes representing loci on five chromosomes (Figure 3.4), which was anchored by a SNP at 73,680,263 bp on BTA8. There are four genes within the flanking regions of this position, which are *GNRH1*, *KCTD9*, *CDCA2*, and *EBF2*. Gonadotropin-releasing hormone 1 (*GNRH1*) is luteinizing-releasing hormone which is involved in reproduction and may be associated with the fat thickness of steers as castrated males have a more extensive fat covering and differ in body size to their intact male counterparts [120, 121]. Early B-cell factor 2 (*EBF2*) has been implicated in osteoblast and osteoclast development, which have major roles in bone development and maintenance [122]. Relationships between osteogenesis,



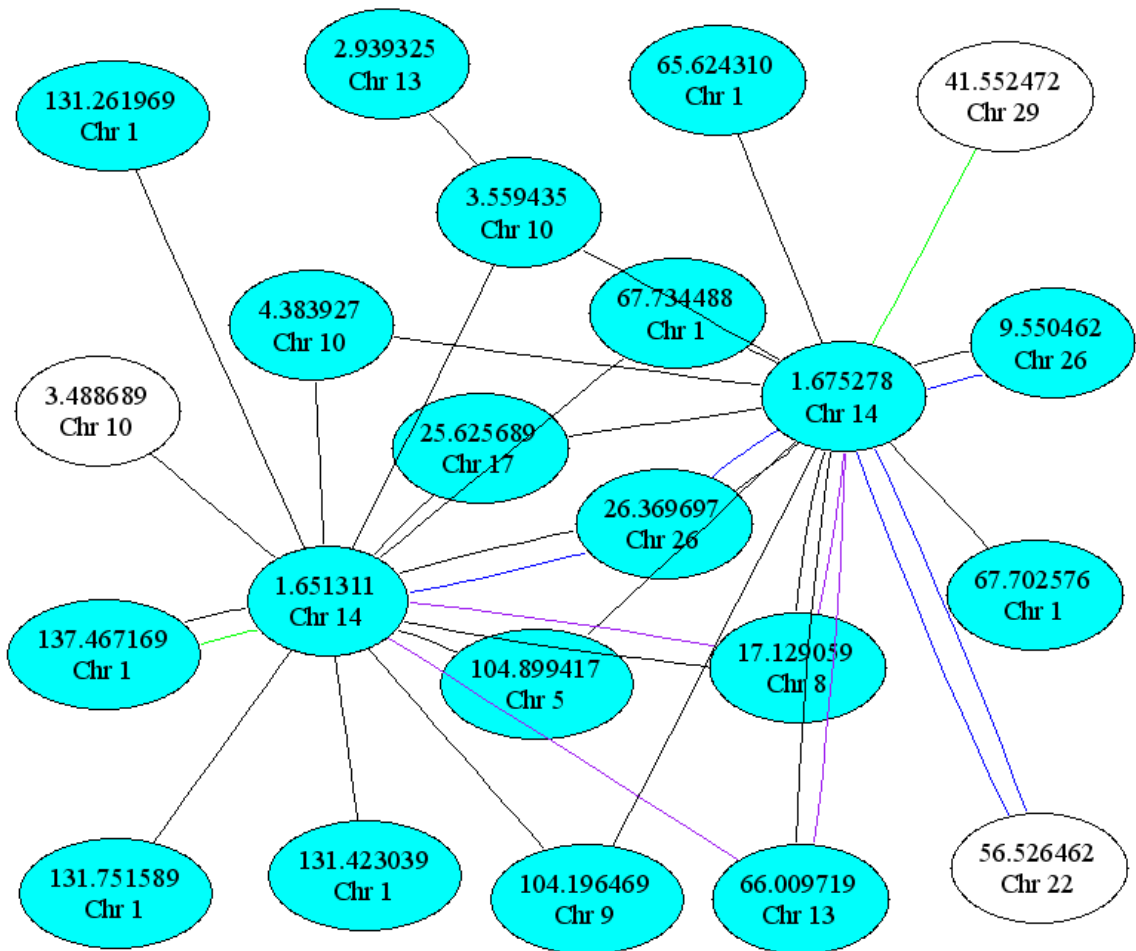
**Figure 3.4 Top epistatic network for FAT**

The largest epistatic network for fat thickness. There are 10 nodes across five chromosomes in this network. The anchor SNP is located at 73,680,263 bp on BTA8. Connecting line colours signify the type of epistatic interaction: black = overall interaction and green = dominance x dominance. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$  and white =  $p < 10^{-9}$ .



### Figure 3.5 Top epistatic network for USFT

The largest epistatic network for fat thickness at ultrasound. This is a large network spanning 22 nodes and 11 chromosomes. Two markers on BTA14 anchor this network at 1,651,311 and 1,675,278 bp representing a 23,968 bp segment of this chromosome. Connecting line colours signify the type of epistatic interaction: black = overall interaction, purple = additive x dominance, blue = dominance x additive, and green = dominance x dominance. Node colour indicates p-value significance: cyan =  $p < 10^{-11}$  and white =  $p < 10^{-9}$ .



haematopoiesis and adipogenesis also exist presumably because cell types of all three lineages are derived from multipotent cells originating in adult bone marrow [123, 124]. Furthermore, the adiponectin-mediated suppression of adipogenesis suggests a role for subcutaneous fat in haematopoiesis [125]. The second largest FAT network contains eight nodes representing loci on five chromosomes (Figure 3.S4) with an anchor node at 49,562,974 bp on BTA26. Genes in the 200 kb regions flanking the SNP include only glutaredoxin 3 (*GLRX3*). While *GLRX3* is not located in region of the genome for which bovine QTL have been discovered, porcine *GLRX3* may underlie a belly weight QTL in the Duroc and Pietrain breeds [126].

Four epistatic networks were found for USFT (Figures 3.5, 3.S5-7). The largest network included 22 nodes with loci on 11 chromosomes. Two SNPs on BTA14 at 1,651,311 and 1,675,278 bp, defining a 23,968 bp segment of BTA14, anchor this network and the region contains 25 genes. The most interesting of these is diacylglycerol O-acyltransferase 1 (*DGATI*) which is associated with milk fat and protein contents and has been suggested as a positional candidate for an intramuscular fat QTL in German Holstein and Charolais cattle [127] and meat quality and carcass fatness QTL in commercial Chinese cattle [128]. Within the 24 genes is a group (*CYHR1-VPS28-DGATI*) that has been exposed to strong artificial selection within the Holstein dairy breed [129]. Within this cluster, *DGATI* is responsible for variation in milk composition [130] and possibly pregnancy rate within dairy breeds [131]. Variation within *DGATI* also affects milk composition in *Bos indicus* (Zebu cattle) and *Bubalus bubalis* (riverine buffalo) [132, 133]. There is evidence of QTL overlapping *DGATI* that are associated

with subcutaneous fat, carcass weight, rump fat thickness, body condition score and intramuscular fat within the CattleQTLdb [99].

### **Pathway analysis**

Pathway analysis was performed for the genes within the 200 kb flanking regions of every locus present within the pairwise association results. Both weight and fatness traits had lipid and carbohydrate metabolism (HCW, USWT, USFT and USFT, WW, respectively) within the top five molecular and cellular functions (Table 3.3). Tissue morphology and development, connective tissue development and function and skeletal and muscular system development and function were enriched terms in the epistatic networks generated for WW, USWT, HCW, FAT, USFT, and USIMF (Table 3.3). Considerable overlap was detected when comparing the five largest networks for each of the traits (Table 3.4). Examination of the terms enriched in the networks reveals that every trait contains genes within epistatic interactions that are involved in connective tissue disorders or developmental, skeletal and muscular disorders. Four of the seven traits also involve genes in epistatic interactions that are involved in lipid or carbohydrate metabolism. Overlaps between connective tissue disorders and lipid metabolism (USWT, HCW, and FAT), skeletal muscle disorders and development (WW or BW) and lipid and carbohydrate metabolism and tissue morphology (USFT) were also found. Among the other largest networks there was a strong enrichment of genes involved in cellular signalling, assembly and organization and function and maintenance, endocrine disorders and development, small molecule biochemistry and hereditary disorders.

The phosphatidylinositol 3-kinase (PI3k)/Akt signalling pathway was the highest ranked canonical pathway found for WW ( $p = 0.00127$ ) and USFT ( $p = 0.00269$ ) and is

**Table 3.3 Functional enrichment of genes underlying pairwise epistasis.**

<b>Enrichment Description</b>	<b>BW</b>	<b>WW</b>	<b>USWT</b>	<b>HCW</b>	<b>FAT</b>	<b>USFT</b>	<b>USIMF</b>
<b><i>Top Diseases and Disorders</i></b>							
skeletal & muscular disorders	5				4	2	
connective tissue disorders	2						
immunological disease							5
<b><i>Top Molecular &amp; Cellular Functions</i></b>							
lipid metabolism			4	3		5	
carbohydrate metabolism		1				4	
<b><i>Top Physiological System Development and Functions</i></b>							
connective tissue development and function				3		1	
tissue morphology		5			5		4
tissue development		4	5		3		5
skeletal and muscular system development and function			1				
immune cell trafficking	4						2
cell-mediated immune response	2						
organismal functions	1						
organismal development		3	3			5	
organ morphology	5	2		2			
organ development				5		4	
embryonic development			2	4	1	2	
nervous system development and function			4		2	3	3
haematological system development and function	3						1
cardiovascular system development and function					4		
endocrine system development and function		1					
auditory and vestibular system development and function				1			

Numbers indicate the ranking of each term on a scale of 1 to 5. Blank entries represent a trait without the enrichment description. All entries are significant at the level of  $p < 0.03$ .

**Table 3.4 Top five networks identified by trait**

Trait	Rank	Network Description
<b>BW</b>	1	cell morphology, cellular function and maintenance, DNA replication, recombination and repair
	2	cell cycle, cellular assembly and organization, cellular function and maintenance
	3	cell morphology, <b>skeletal and muscular system development and function</b> , cell-to-cell signalling and interaction
	4	<b>connective tissue disorders</b> , nervous system development and function, dermatological diseases and conditions
	5	cell-to-cell signalling and interaction, nervous system development and function, cell morphology
<b>WW</b>	1	small molecule biochemistry, cell morphology, cellular assembly and organization
	2	cancer, endocrine system disorders, endocrine system development and function
	3	hereditary disorder, <b>skeletal and muscular disorders</b> , <b>connective tissue disorders</b>
	4	hereditary disorders, <b>skeletal and muscular disorders</b> , cell cycle
	5	cell-to-cell signalling and interaction cellular assembly and organization, cellular function and maintenance
<b>USWT</b>	1	cellular movement, cell-to-cell signalling and interaction, cellular function and maintenance
	2	<b>lipid metabolism</b> , molecular transport, small molecule biochemistry
	3	cardiovascular disease, hereditary disorder, metabolic disease
	4	organismal development, <b>connective tissue development and function</b> , embryonic development
	5	cell-to-cell signalling and interaction, cellular assembly and organization, cellular function and maintenance
<b>HCW</b>	1	cell-to-cell signalling and interaction, haematological system development and function, <b>immune cell trafficking</b>
	2	<b>connective tissue disorders</b> , dental disease, developmental disorder
	3	hereditary disorder, neurological disease, <b>connective tissue disorders</b>
	4	endocrine system development and function, small molecule biochemistry, <b>lipid metabolism</b>
	5	<b>lipid metabolism</b> , small molecule biochemistry, endocrine system disorders
<b>FAT</b>	1	developmental disorder, hereditary disorder, ophthalmic disease
	2	DNA replication, recombination, and repair, metabolic disease, organismal injury and abnormalities
	3	endocrine system disorders, <b>lipid metabolism</b> , molecular transport
	4	<b>connective tissue disorders</b> , inflammatory disease, neurological disease
	5	cell morphology, cellular assembly and organization, cell death and survival
<b>USFT</b>	1	<b>connective tissue disorders</b> , dental disease, developmental disorder
	2	post-translational modification, haematological system development and function, <b>tissue morphology</b>
	3	<b>carbohydrate metabolism</b> , <b>lipid metabolism</b> , small molecule biochemistry
	4	cell cycle, cellular assembly and organization, DNA replication, recombination, and repair
	5	<b>lipid metabolism</b> , small molecule biochemistry, <b>carbohydrate metabolism</b>
<b>USIMF</b>	1	DNA replication, recombination, and repair, gene expression, neurological disease
	2	cell morphology, <b>connective tissue development and function</b> , cellular growth and proliferation
	3	cell signalling, cardiovascular disease, cell morphology
	4	cellular function and maintenance, haematological system development and function, cell signalling
	5	hereditary disorder, neurological disease, ophthalmic disease

Descriptions in **bold** are those that directly relate to the analysed traits.

potentially involved in the regulation of FAT, where PI3K signalling in B lymphocytes was the fourth ranked canonical pathway ( $p = 0.00114$ ). The PI3K/Akt signalling pathway is a large signalling system affecting cell growth, cell cyclicity, cell survival, and glucose metabolism [134]. It is integral to protein synthesis and subsequent skeletal muscle hypertrophy driven by IGF-1 signalling [135, 136]. Akt signalling has also been shown to affect both hypertrophy and atrophy in humans [137]. The D-myo-inositol-5-phosphate metabolism and inositol phosphate compound superpathways were ranked first and second for both USIMF ( $p = 0.00482$  and  $p = 0.00895$ ) and USWT ( $p = 0.00078$  and  $p = 0.00172$ ). Inositol phosphate compounds are involved in cellular signalling and could also be related to PI3K/Akt signalling.

Enrichment of genes involved in connective tissue, skeletal and muscular development and disorders as well as lipid metabolism suggests a strong biological underpinning for the existence of pairwise interactions and epistatic networks generated from these interacting loci. The overlap of genes enriched in lipid and muscular pathways among both the weight and fatness trait networks also adds to this evidence. The deposition of muscle and fat are biologically interconnected and these traits are strongly physiologically related at maturity in cattle.

### **Data comparisons and adjustments**

A large number of QTL for weight traits were found by McClure et al. [116] who analysed 390 microsatellites, 11 SNPs and one duplication locus in 1,622 half-sib steers drawn from this population. Many of these QTL overlapped the locations of the node SNPs found to be involved in epistatic interactions in this study. However, in this analysis we analysed high density SNP genotypes in the complete set of animals

regardless of the pedigree relationships existing among the steers. The identification of coincident additive effect QTL with the locations of epistatic loci provides additional support that the identified epistatic loci are, in fact, biologically meaningful and therefore that the QTL may be context dependent.

Correcting p-values for multiple testing was not trivial in this study. EPISNPmpi makes no attempt to correct the p-values and does not allow the output of the entire distribution of p-values for all tested locus pairs which would allow the empirical establishment of the threshold for an appropriate false discovery rate. Consequently, a Bonferroni correction was adopted. With more than 1.1 billion pairwise comparisons, the Bonferroni correction corresponding to an experiment-wise  $p \leq 0.05$  is a test-wise significance of  $p = 4.54 \times 10^{-11}$  which is exceptionally conservative. In fact with only 2,547 observations included in the analysis (Table 3.1), only 2,547 of the 46,925 SNP effects are estimable and the remaining SNP effects are linear combinations of the 2,547 estimable effects. This suggests that an appropriate Bonferroni correction factor would be that corresponding to  $2,547 \times (2,546/2) = 3,242,331$  tests (presuming these are all estimable) which for an experiment-wise  $p \leq 0.05$  requires a test-wise significance of  $p = 1.54 \times 10^{-8}$ . Accordingly, we used a test-wise threshold of  $p \leq 10^{-8}$  to identify significant pairwise interactions.

## **Conclusions**

We find the existence of pairwise epistasis for seven growth and fatness traits of beef cattle and that the interacting locus pairs define complex interchromosomal networks. Among these, the most interesting is a 53,900 bp region on BTA2 defined by 3 contiguous SNPs which anchors the largest networks for the weight traits. Genes



associated with connective tissue, skeletal and muscular development and lipid metabolism are enriched within the genomic regions flanking the SNPs that are involved in the two-way interactions.

In several cases we also find evidence for QTL with large additive effects on the analysed weight and fatness traits. These results are interesting and suggest that some QTL effects may be context dependent. That is, the additive effect at a QTL locus may be dependent on genotype status or allele frequencies at a second interacting locus. Thus a variant which acts as a QTL in one population may not be a QTL in a second population where genotype status at an interacting locus differs. The extent to which epistasis explains phenotypic variation within these traits deserves further consideration to arrive at an understanding of the potential magnitude of this effect.

### **List of Abbreviations**

Abbreviations are defined in the text.

### **Competing interests**

The authors declare that they have no competing interests.

### **Authors' contributions**

JFT and HRR designed the experiment. HRR, RDS, PCT and JFT analysed data. JLH aided in scripting for data formatting and analysis. RDS genotyped samples and managed the genotype database. HRR and JFT wrote the manuscript. All authors read and approved the final manuscript.

### **Acknowledgements**

We gratefully acknowledge the donation of samples and phenotypes by the Circle A Ranch of Iberia, MO. This project was supported by the University of Missouri and

National Research Initiative grants number 2008-35205-04687 and 2008-35205-18864 from the USDA Cooperative State Research, Education and Extension Service and National Research Initiative grants number 2009-65205-05635, 2011-68004-30214 and 2013-68004-20364 from the USDA National Institute of Food and Agriculture.

**Table 3.S1 Pairwise interaction results – BW**

100 MOST SIGNIFICANT RESULTS FROM THE PAIRWISE EPISTASIS ANALYSES

I = interaction effect between the two SNPs

AA = additive x additive effect, AD = additive x dominance effect

DA = dominance x additive effect, DD = dominance x dominance effect

Chr1	Locus1 (Mb)	Chr2	Locus2 (Mb)	Trait	Test	P_value				
1	7.405025	8	50.250254	BW_Phen	AA	0.693E-09				
	AxA	1_2	2_1	1_1	2_2					
	Estimate	1.029	0.202	-0.229	-1.049					
1	7.405025	8	50.278766	BW_Phen	AA	0.101E-08				
	AxA	1_2	2_1	1_1	2_2					
	Estimate	1.024	0.198	-0.225	-1.049					
1	7.405025	8	50.310875	BW_Phen	AA	0.967E-09				
	AxA	1_1	2_2	1_2	2_1					
	Estimate	1.027	0.198	-0.225	-1.052					
1	68.403441	17	20.634031	BW_Phen	AA	0.132E-08				
	AxA	2_2	1_1	1_2	2_1					
	Estimate	1.847	0.122	-0.379	-0.467					
1	151.379106	9	105.238273	BW_Phen	I	0.000E+00				
1	151.379106	9	105.238273	BW_Phen	AA	0.135E-09				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	1.050	0.454	-0.083	-3.171					
1	154.506866	14	52.508673	BW_Phen	AA	0.117E-09				
	AxA	2_2	1_1	2_1	1_2					
	Estimate	0.520	0.407	-0.445	-0.506					
1	154.506866	14	52.508673	BW_Phen	I	0.158E-08				
1	154.529450	14	52.508673	BW_Phen	AA	0.117E-09				
	AxA	2_2	1_1	2_1	1_2					
	Estimate	0.520	0.407	-0.445	-0.506					
1	154.529450	14	52.508673	BW_Phen	I	0.158E-08				
2	21.448855	27	23.598119	BW_Phen	I	0.105E-08				
2	46.557061	6	93.486356	BW_Phen	DD	0.758E-09				
	DxD	22_22	22_11	12_12	11_22	11_11	11_12	12_11	12_22	22_12
	Estimate	13.330	7.405	0.903	0.191	0.050	-0.091	-0.521	-2.044	-12.524
2	49.914972	3	31.775531	BW_Phen	AA	0.928E-09				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.849	0.304	-0.181	-1.156					
2	63.355675	26	42.425741	BW_Phen	AA	0.326E-09				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.549	0.540	-0.241	-0.844					
2	65.301812	4	26.518653	BW_Phen	AA	0.108E-08				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	1.009	0.212	-0.459	-0.471					
2	87.843486	21	19.267278	BW_Phen	I	0.438E-09				
2	109.839219	2	135.926465	BW_Phen	AA	0.253E-09				
	AxA	1_1	2_2	1_2	2_1					
	Estimate	1.553	0.135	-0.240	-0.951					
3	18.549536	20	39.658427	BW_Phen	AA	0.123E-08				
	AxA	1_2	2_1	2_2	1_1					
	Estimate	0.929	0.288	-0.089	-2.510					
3	22.624014	28	4.189551	BW_Phen	AA	0.189E-08				
	AxA	2_1	1_2	1_1	2_2					
	Estimate	1.313	0.627	-0.123	-1.789					
3	28.978174	9	49.556408	BW_Phen	AA	0.532E-09				
	AxA	2_2	1_1	2_1	1_2					
	Estimate	2.392	0.124	-0.111	-2.132					
3	30.194236	17	0.918620	BW_Phen	AA	0.238E-09				
	AxA	1_2	2_1	1_1	2_2					
	Estimate	0.891	0.252	-0.147	-1.821					
3	30.194236	17	0.918620	BW_Phen	I	0.838E-09				
3	31.013438	11	18.390236	BW_Phen	AA	0.148E-08				
	AxA	1_2	2_1	1_1	2_2					
	Estimate	0.661	0.359	-0.246	-0.833					
3	58.003570	7	36.127497	BW_Phen	I	0.000E+00				
3	58.329027	7	36.127497	BW_Phen	AA	0.211E-09				
	AxA	2_1	1_2	1_1	2_2					
	Estimate	0.781	0.378	-0.327	-0.546					
3	58.434544	7	36.127497	BW_Phen	AA	0.214E-09				

	AxA	2_1	1_2	1_1	2_2								
	Estimate	0.639	0.403	-0.347	-0.495								
3	68.757647	7	5.352670			BW_Phen	AA	0.715E-09					
	AxA	1_2	2_1	1_1	2_2								
	Estimate	2.662	0.093	-0.357	-0.509								
3	74.234478	7	33.338008			BW_Phen	AA	0.146E-08					
	AxA	2_2	1_1	1_2	2_1								
	Estimate	0.778	0.196	-0.123	-1.891								
3	100.758921	22	34.416644			BW_Phen	I	0.000E+00					
4	38.435198	20	50.265235			BW_Phen	AA	0.138E-08					
	AxA	1_2	2_1	2_2	1_1								
	Estimate	0.561	0.304	-0.348	-0.608								
4	38.435198	20	50.407500			BW_Phen	AA	0.929E-09					
	AxA	1_1	2_2	2_1	1_2								
	Estimate	0.565	0.308	-0.352	-0.617								
4	38.435198	20	50.407500			BW_Phen	I	0.108E-08					
4	94.065239	28	44.118793			BW_Phen	AA	0.864E-09					
	AxA	1_1	2_2	2_1	1_2								
	Estimate	1.092	0.225	-0.210	-0.901								
4	94.115101	28	44.118793			BW_Phen	AA	0.885E-09					
	AxA	2_1	1_2	1_1	2_2								
	Estimate	1.094	0.224	-0.210	-0.902								
5	13.659330	12	16.400906			BW_Phen	AA	0.965E-10					
	AxA	2_1	1_2	2_2	1_1								
	Estimate	1.845	0.131	-0.295	-0.643								
5	13.659330	12	16.518473			BW_Phen	AA	0.878E-09					
	AxA	2_1	1_2	2_2	1_1								
	Estimate	1.326	0.169	-0.349	-0.453								
6	4.217935	21	40.837384			BW_Phen	I	0.200E-09					
6	4.255214	21	38.076668			BW_Phen	AA	0.584E-09					
	AxA	1_2	2_1	2_2	1_1								
	Estimate	1.559	0.283	-0.107	-1.757								
6	12.627595	7	71.545383			BW_Phen	I	0.770E-10					
6	46.533015	11	31.224438			BW_Phen	I	0.104E-08					
6	62.408025	9	8.906316			BW_Phen	DD	0.191E-08					
	DxD	11_12	12_11	12_22	22_12	22_22	22_11	12_12	11_22	11_11			
	Estimate	2.326	2.279	0.425	0.340	-0.153	-0.776	-0.889	-1.070	-4.679			
6	106.758901	28	24.900950			BW_Phen	AA	0.128E-09					
	AxA	2_1	1_2	2_2	1_1								
	Estimate	0.657	0.351	-0.403	-0.594								
6	108.423481	8	108.094943			BW_Phen	I	0.000E+00					
6	108.423481	8	108.094943			BW_Phen	AA	0.827E-09					
	AxA	1_1	2_2	1_2	2_1								
	Estimate	0.903	0.251	-0.309	-0.546								
7	39.679687	7	41.466011			BW_Phen	I	0.278E-09					
7	91.226072	7	91.483570			BW_Phen	AA	0.222E-09					
	AxA	1_2	2_1	2_2	1_1								
	Estimate	0.388	0.377	-0.738	-0.860								
7	91.226072	7	91.974966			BW_Phen	AA	0.553E-09					
	AxA	1_2	2_1	1_1	2_2								
	Estimate	0.399	0.339	-0.708	-0.761								
7	91.483570	7	92.293717			BW_Phen	I	0.000E+00					
7	91.483570	7	92.293717			BW_Phen	AA	0.415E-10					
	AxA	2_1	1_2	2_2	1_1								
	Estimate	0.419	0.364	-0.976	-1.364								
7	91.483570	7	92.474466			BW_Phen	I	0.250E-10					
7	91.483570	7	92.474466			BW_Phen	AA	0.136E-09					
	AxA	2_2	1_1	2_1	1_2								
	Estimate	0.419	0.307	-0.977	-1.385								
7	91.839217	7	92.347964			BW_Phen	AA	0.116E-10					
	AxA	1_2	2_1	1_1	2_2								
	Estimate	0.550	0.293	-0.793	-0.841								
7	91.839217	7	92.347964			BW_Phen	I	0.636E-09					
7	91.839217	7	92.474466			BW_Phen	AA	0.882E-09					
	AxA	1_2	2_1	2_2	1_1								
	Estimate	0.444	0.332	-0.490	-0.767								
7	91.946384	7	92.293717			BW_Phen	AA	0.779E-09					
	AxA	2_1	1_2	2_2	1_1								
	Estimate	0.412	0.349	-0.825	-0.865								
7	91.946384	7	92.474466			BW_Phen	AA	0.141E-08					
	AxA	2_2	1_1	2_1	1_2								



15	Estimate	1.239	0.204	-0.310	-0.655			
	76.166896	29	30.397108			BW_Phen	AA	0.193E-08
	AxA	1_1	2_2	1_2	2_1			
	Estimate	1.449	0.175	-0.126	-1.476			
18	52.685927	20	8.164878			BW_Phen	I	0.000E+00
18	52.685927	20	8.164878			BW_Phen	DA	0.966E-11
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	1.577	0.745	0.311	-0.269	-0.862	-1.571	
19	24.995142	24	60.745394			BW_Phen	AA	0.246E-09
	AxA	2_2	1_1	1_2	2_1			
	Estimate	3.310	0.075	-0.316	-0.874			
19	32.604521	24	60.745394			BW_Phen	AA	0.132E-08
	AxA	1_2	2_1	2_2	1_1			
	Estimate	3.198	0.071	-0.300	-0.928			
19	32.741678	24	60.745394			BW_Phen	AA	0.181E-08
	AxA	1_2	2_1	2_2	1_1			
	Estimate	3.211	0.069	-0.292	-0.908			
19	35.851319	23	40.951388			BW_Phen	AA	0.360E-09
	AxA	2_1	1_2	1_1	2_2			
	Estimate	1.780	0.182	-0.157	-1.312			
20	6.534735	20	7.007017			BW_Phen	AA	0.955E-10
	AxA	2_1	1_2	2_2	1_1			
	Estimate	0.474	0.428	-0.285	-0.981			
20	19.757303	20	21.606396			BW_Phen	AA	0.136E-08
	AxA	2_2	1_1	2_1	1_2			
	Estimate	0.538	0.300	-0.333	-0.795			
20	19.757303	20	21.650376			BW_Phen	AA	0.151E-08
	AxA	2_1	1_2	2_2	1_1			
	Estimate	0.521	0.298	-0.337	-0.788			
20	68.509279	28	26.972581			BW_Phen	AA	0.138E-08
	AxA	2_2	1_1	1_2	2_1			
	Estimate	0.813	0.241	-0.437	-0.484			

---

**Table 3.S2 Pairwise interaction results – FAT**

100 MOST SIGNIFICANT RESULTS FROM THE PAIRWISE EPISTASIS ANALYSES

I = interaction effect between the two SNPs

AA = additive x additive effect, AD = additive x dominance effect

DA = dominance x additive effect, DD = dominance x dominance effect

Chr1	Locus1 (Mb)	Chr2	Locus2 (Mb)	Trait	Test	P_value				
1	34.934516	1	43.471154	FAT_Phen	I	0.000E+00				
1	40.830066	3	47.820211	FAT_Phen	AA	0.199E-09				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.014	0.005	-0.003	-0.029					
1	40.854754	3	47.820211	FAT_Phen	AA	0.366E-09				
	AxA	2_1	1_2	1_1	2_2					
	Estimate	0.014	0.005	-0.003	-0.029					
1	46.201490	15	15.127363	FAT_Phen	AA	0.171E-10				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.014	0.006	-0.003	-0.024					
1	46.201490	15	15.127363	FAT_Phen	I	0.333E-10				
1	46.201490	15	15.162470	FAT_Phen	AA	0.280E-10				
	AxA	1_2	2_1	2_2	1_1					
	Estimate	0.014	0.006	-0.003	-0.024					
1	46.201490	15	15.162470	FAT_Phen	I	0.653E-09				
1	49.480638	11	32.066197	FAT_Phen	I	0.000E+00				
1	49.480638	13	44.467210	FAT_Phen	I	0.000E+00				
1	69.386583	16	8.179652	FAT_Phen	I	0.000E+00				
1	131.580342	21	46.560345	FAT_Phen	I	0.734E-09				
1	138.692649	3	47.519010	FAT_Phen	DA	0.126E-08				
	DxA	11_1	12_2	13_1	22_2	12_1	11_2			
	Estimate	0.013	0.012	0.011	-0.009	-0.012	-0.015			
2	15.709188	6	63.532785	FAT_Phen	I	0.132E-08				
2	21.595945	22	28.569757	FAT_Phen	I	0.144E-08				
2	73.412080	16	48.048566	FAT_Phen	AA	0.139E-08				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.008	0.007	-0.009	-0.009					
2	87.108690	26	36.938045	FAT_Phen	I	0.000E+00				
2	87.108690	26	36.988650	FAT_Phen	I	0.000E+00				
2	113.174994	17	21.352266	FAT_Phen	I	0.000E+00				
2	113.174994	17	21.488452	FAT_Phen	I	0.000E+00				
2	113.174994	26	51.376213	FAT_Phen	I	0.000E+00				
2	133.599634	7	57.641183	FAT_Phen	I	0.000E+00				
3	13.474988	18	37.927789	FAT_Phen	DA	0.161E-08				
	DxA	22_2	11_2	12_1	11_1	13_1	12_2			
	Estimate	0.154	0.007	0.005	-0.001	-0.017	-0.031			
3	39.841272	26	49.562974	FAT_Phen	I	0.000E+00				
3	39.978795	26	49.562974	FAT_Phen	I	0.000E+00				
3	40.678490	26	49.562974	FAT_Phen	I	0.164E-09				
3	73.465083	20	9.965585	FAT_Phen	I	0.169E-08				
3	83.534093	8	73.680263	FAT_Phen	I	0.000E+00				
3	83.534093	8	73.881694	FAT_Phen	I	0.000E+00				
3	83.534093	8	73.881694	FAT_Phen	DD	0.110E-08				
	DxD	11_11	22_11	12_12	11_22	22_22	12_22	22_12	11_12	12_11
	Estimate	0.490	0.026	0.008	0.004	0.001	-0.001	-0.004	-0.021	-0.079
3	105.654383	7	59.766632	FAT_Phen	I	0.971E-09				
4	36.616352	24	2.466721	FAT_Phen	I	0.000E+00				
4	73.669041	8	79.817203	FAT_Phen	DA	0.702E-09				
	DxA	11_1	12_2	13_1	22_2	12_1	11_2			
	Estimate	0.339	0.004	0.004	0.000	-0.042	-0.067			
4	75.666435	12	77.992661	FAT_Phen	I	0.975E-09				
4	87.744523	6	109.835444	FAT_Phen	I	0.258E-09				
5	13.362021	21	14.915708	FAT_Phen	DD	0.129E-09				
	DxD	22_11	11_11	12_12	22_22	11_22	12_22	11_12	22_12	12_11
	Estimate	0.574	0.071	0.014	0.003	0.001	-0.001	-0.008	-0.032	-0.119
5	29.839619	29	14.116468	FAT_Phen	AA	0.153E-08				
	AxA	2_1	1_2	1_1	2_2					
	Estimate	0.044	0.002	-0.010	-0.017					
5	63.527807	29	41.133252	FAT_Phen	DD	0.500E-09				
	DxD	11_22	11_11	12_12	22_22	22_11	22_12	12_11	12_22	11_12
	Estimate	0.078	0.020	0.017	0.013	0.004	-0.008	-0.009	-0.025	-0.035
5	73.713772	5	78.909424	FAT_Phen	AA	0.656E-09				

	AxA	1_1	2_2	1_2	2_1				
5	Estimate	0.043	0.002	-0.005	-0.011	FAT_Phen	AA	0.105E-08	
	73.713772	5	78.936196						
	AxA	1_2	2_1	1_1	2_2				
5	Estimate	0.043	0.002	-0.005	-0.011	FAT_Phen	AA	0.100E-08	
	74.170131	5	78.909424						
	AxA	1_1	2_2	1_2	2_1				
5	Estimate	0.041	0.002	-0.004	-0.012	FAT_Phen	AA	0.106E-08	
	74.170131	5	78.936196						
	AxA	1_2	2_1	1_1	2_2				
5	Estimate	0.041	0.002	-0.004	-0.012	FAT_Phen	AA	0.233E-09	
	74.191590	5	78.909424						
	AxA	2_1	1_2	2_2	1_1				
5	Estimate	0.038	0.002	-0.004	-0.013	FAT_Phen	AA	0.505E-10	
	74.191590	5	78.936196						
	AxA	2_2	1_1	2_1	1_2				
5	Estimate	0.041	0.002	-0.004	-0.013	FAT_Phen	I	0.653E-09	
5	74.191590	5	78.936196			FAT_Phen	AA	0.670E-09	
	74.191590	5	78.956945						
	AxA	2_1	1_2	2_2	1_1				
5	Estimate	0.039	0.002	-0.004	-0.014	FAT_Phen	AA	0.171E-08	
	76.385743	19	39.126656						
	AxA	1_2	2_1	2_2	1_1				
5	Estimate	0.015	0.005	-0.005	-0.013	FAT_Phen	I	0.000E+00	
6	118.148263	8	73.881694			FAT_Phen	I	0.000E+00	
6	21.859572	16	31.654996			FAT_Phen	I	0.000E+00	
6	21.859572	16	34.295245			FAT_Phen	I	0.000E+00	
6	31.275687	21	13.633071			FAT_Phen	AA	0.744E-09	
	AxA	1_1	2_2	2_1	1_2				
6	Estimate	0.058	0.001	-0.004	-0.014	FAT_Phen	I	0.170E-08	
6	31.275687	21	13.633071			FAT_Phen	I	0.153E-08	
6	33.509143	10	57.701989			FAT_Phen	I	0.175E-08	
6	33.542318	10	57.701989			FAT_Phen	I	0.000E+00	
6	49.724627	6	107.921117			FAT_Phen	I	0.000E+00	
6	55.077357	8	73.680263			FAT_Phen	I	0.000E+00	
6	61.555559	18	26.844474			FAT_Phen	I	0.180E-08	
7	3.530805	10	79.778640			FAT_Phen	I	0.123E-08	
8	10.488006	16	9.211977			FAT_Phen	I	0.927E-09	
8	15.975916	12	88.586618			FAT_Phen	I	0.111E-08	
8	15.975916	17	32.911852			FAT_Phen	I	0.209E-08	
8	29.831209	11	6.458630			FAT_Phen	I	0.000E+00	
8	47.598802	13	78.115143			FAT_Phen	I	0.000E+00	
8	73.680263	16	30.910755			FAT_Phen	I	0.992E-09	
8	73.680263	16	30.935096			FAT_Phen	I	0.749E-09	
8	73.680263	16	31.654996			FAT_Phen	I	0.000E+00	
9	31.971498	9	50.869600			FAT_Phen	I	0.687E-09	
9	32.004342	9	50.869600			FAT_Phen	I	0.177E-08	
9	99.229165	26	49.562974			FAT_Phen	I	0.000E+00	
9	102.996425	9	105.263583			FAT_Phen	AA	0.981E-10	
	AxA	2_1	1_2	2_2	1_1				
10	Estimate	0.046	0.002	-0.004	-0.010	FAT_Phen	I	0.000E+00	
10	54.082665	25	36.234087			FAT_Phen	AA	0.820E-09	
	62.792747	26	10.375915						
	AxA	1_2	2_1	1_1	2_2				
10	Estimate	0.014	0.005	-0.007	-0.010	FAT_Phen	AA	0.210E-08	
	80.602211	12	48.885042						
	AxA	1_1	2_2	1_2	2_1				
11	Estimate	0.013	0.005	-0.007	-0.010	FAT_Phen	AA	0.963E-09	
	9.665006	15	36.997902						
	AxA	2_1	1_2	2_2	1_1				
11	Estimate	0.029	0.003	-0.002	-0.029	FAT_Phen	AA	0.266E-09	
	71.117880	20	57.781878						
	AxA	2_2	1_1	2_1	1_2				
11	Estimate	0.012	0.006	-0.004	-0.018	FAT_Phen	DD	0.686E-09	
	94.494230	22	17.857858						
	DxD	11_22	22_22	12_12	11_11	22_11	12_11	22_12	11_12
12	Estimate	0.751	0.259	0.013	0.003	0.001	-0.001	-0.006	-0.023
	68.060288	15	71.772056			FAT_Phen	DD	0.196E-08	
	DxD	11_11	11_22	12_12	22_11	22_22	22_12	12_22	12_11
12	Estimate	0.849	0.047	0.011	0.003	0.001	-0.002	-0.005	-0.029
	79.310379	27	25.424395			FAT_Phen	I	0.000E+00	



13	38.568755	17	42.046346		FAT_Phen	DD	0.350E-09
	DxD	11_22	11_11 12_12	22_22	22_11 22_12	12_11	12_22 11_12
	Estimate	0.310	0.018 0.013	0.011	0.001 -0.003	-0.005	-0.036 -0.058
13	82.418960	17	54.025463		FAT_Phen	I	0.120E-08
14	2.826632	26	49.562974		FAT_Phen	I	0.171E-08
14	69.665346	25	39.756424		FAT_Phen	DA	0.824E-09
	DxA	11_1	12_2 13_1	22_2	12_1 11_2		
	Estimate	0.081	0.012 0.001	-0.001	-0.010 -0.077		
14	69.665346	25	39.756424		FAT_Phen	I	0.228E-08
15	24.632628	22	17.857858		FAT_Phen	I	0.979E-09
15	81.869228	16	69.795545		FAT_Phen	I	0.128E-08
16	8.486467	22	56.235584		FAT_Phen	AA	0.117E-08
	AxA	1_1	2_2 2_1	1_2			
	Estimate	0.029	0.002 -0.007	-0.013			
16	9.211977	19	37.035051		FAT_Phen	I	0.000E+00
17	44.992963	26	5.387056		FAT_Phen	AD	0.512E-09
	AxD	1_12	2_11 2_22	2_12	1_22 1_11		
	Estimate	0.030	0.013 0.001	-0.004	-0.013 -0.067		
18	18.863162	26	48.838979		FAT_Phen	I	0.000E+00
18	18.863162	26	49.562974		FAT_Phen	I	0.000E+00
18	25.587174	25	34.770404		FAT_Phen	I	0.230E-08
18	34.126956	19	45.501147		FAT_Phen	AA	0.452E-09
	AxA	2_2	1_1 2_1	1_2			
	Estimate	0.011	0.006 -0.004	-0.019			
18	40.002413	26	39.319670		FAT_Phen	DD	0.550E-09
	DxD	11_11	11_22 12_12	22_11	22_22 22_12	12_22	12_11 11_12
	Estimate	0.433	0.042 0.011	0.005	0.000 -0.002	-0.004	-0.033 -0.100
19	17.189453	20	0.593947		FAT_Phen	I	0.798E-09
19	17.189453	20	0.829627		FAT_Phen	I	0.198E-08
20	2.159884	20	4.989460		FAT_Phen	I	0.189E-08
20	11.228271	22	12.081026		FAT_Phen	I	0.152E-08
22	17.857858	29	43.652252		FAT_Phen	I	0.883E-09
24	40.260020	28	5.073546		FAT_Phen	AD	0.113E-08
	AxD	1_22	2_12 1_11	2_11	1_12 2_22		
	Estimate	0.026	0.011 0.006	-0.006	-0.011 -0.020		
24	62.079669	29	28.849610		FAT_Phen	I	0.000E+00
24	62.079669	29	28.849610		FAT_Phen	AD	0.394E-09
	AxD	1_11	2_12 1_22	2_22	1_12 2_11		
	Estimate	0.202	0.004 0.004	-0.001	-0.032 -0.037		

**Table 3.S3 Pairwise interaction results – HCW**

100 MOST SIGNIFICANT RESULTS FROM THE PAIRWISE EPISTASIS ANALYSES

I = interaction effect between the two SNPs

AA = additive x additive effect, AD = additive x dominance effect

DA = dominance x additive effect, DD = dominance x dominance effect

Chr1	Locus1 (Mb)	Chr2	Locus2 (Mb)	Trait	Test	P_value
1	7.782816	10	103.048273	HCW_Phen	AA	0.176E-08
	AxA	1_2	2_1 2_2	1_1		
	Estimate	2_608	2.218 -0.406	-20.635		
1	52.607381	10	17.867003	HCW_Phen	AD	0.723E-09
	AxD	2_12	1_22 1_11	1_12		
	Estimate	11.697	6.362 0.529	-1.683 -2.735 -41.721		
1	57.692535	14	10.013686	HCW_Phen	AA	0.101E-08
	AxA	1_2	2_1 1_1	2_2		
	Estimate	3.669	3.096 -0.656	-14.693		
1	57.890954	14	10.013686	HCW_Phen	AA	0.987E-09
	AxA	1_2	2_1 1_1	2_2		
	Estimate	3.696	3.100 -0.660	-14.539		
1	57.941195	14	10.013686	HCW_Phen	AA	0.111E-08
	AxA	2_2	1_1 2_1	1_2		
	Estimate	3.679	3.105 -0.657	-14.543		
1	58.402388	14	10.013686	HCW_Phen	AA	0.169E-08
	AxA	2_2	1_1 2_1	1_2		
	Estimate	3.672	3.093 -0.656	-14.308		
1	58.444431	14	10.013686	HCW_Phen	AA	0.894E-09
	AxA	1_2	2_1 1_1	2_2		
	Estimate	7.072	1.538 -1.214	-7.475		
1	65.769328	16	29.890723	HCW_Phen	AA	0.801E-09
	AxA	2_2	1_1 1_2	2_1		
	Estimate	3.601	2.047 -0.978	-10.581		
1	89.877873	21	21.065240	HCW_Phen	AA	0.183E-08
	AxA	1_2	2_1 2_2	1_1		
	Estimate	3.208	2.872 -0.548	-16.519		
1	97.906442	21	4.638691	HCW_Phen	I	0.172E-08
1	97.906442	21	5.478138	HCW_Phen	I	0.820E-09
1	100.432385	3	46.721234	HCW_Phen	I	0.193E-08
1	114.647883	21	27.983449	HCW_Phen	DD	0.560E-09
	DxD	12_22	11_12 22_12	12_11 22_11 11_11	12_12 22_22 11_22	
	Estimate	97.000	14.965 4.043	0.266 -0.127 -0.613 -8.404 -79.862	*****	
1	114.647883	21	27.983449	HCW_Phen	I	0.219E-08
1	116.472073	2	69.511580	HCW_Phen	AA	0.105E-08
	AxA	1_2	2_1 2_2	1_1		
	Estimate	8.142	1.190 -1.431	-6.266		
1	116.472073	2	69.662990	HCW_Phen	AA	0.637E-09
	AxA	1_2	2_1 2_2	1_1		
	Estimate	8.366	1.218 -1.452	-6.298		
1	116.472073	2	69.702569	HCW_Phen	AA	0.408E-09
	AxA	1_2	2_1 2_2	1_1		
	Estimate	8.386	1.229 -1.462	-6.397		
1	116.472073	2	70.978502	HCW_Phen	AA	0.154E-08
	AxA	1_1	2_2 2_1	1_2		
	Estimate	9.948	0.847 -2.016	-5.149		
1	118.002920	2	69.511580	HCW_Phen	AA	0.279E-09
	AxA	1_2	2_1 2_2	1_1		
	Estimate	9.220	1.186 -1.419	-6.354		
1	118.002920	2	69.662990	HCW_Phen	AA	0.240E-09
	AxA	1_2	2_1 2_2	1_1		
	Estimate	9.329	1.200 -1.427	-6.374		
1	118.002920	2	69.702569	HCW_Phen	AA	0.170E-09
	AxA	1_2	2_1 2_2	1_1		
	Estimate	9.376	1.210 -1.436	-6.435		
1	119.421373	14	24.326513	HCW_Phen	I	0.000E+00
1	119.544478	10	91.155650	HCW_Phen	I	0.250E-11
2	68.795555	10	84.792134	HCW_Phen	DA	0.572E-09
	DxA	22_2	12_1 11_2	11_1 12_2 13_1		
	Estimate	81.108	3.530 0.252	-0.341 -3.017 -69.553		
3	1.103729	21	27.983449	HCW_Phen	AD	0.526E-09
	AxD	1_22	2_12 1_11	2_11 1_12 2_22		

	Estimate	56.548	5.464	0.266	-0.245	-7.104	*****			
3	89.422763	21	27.983449				HCW_Phen	AD	0.155E-08	
	AxD	2_22	1_12	2_11	1_11	2_12	1_22			
	Estimate	54.887	9.636	0.143	-0.329	-3.923	*****			
3	91.433659	14	35.663156				HCW_Phen	I	0.506E-09	
3	95.120026	15	35.437164				HCW_Phen	DD	0.118E-09	
	DxD	12_22	11_12	12_11	22_12	22_11	22_22	12_12	11_11	11_22
	Estimate	18.895	17.412	1.182	0.922	-0.296	-3.545	-4.530	-5.192	*****
3	95.120026	15	35.537347				HCW_Phen	DD	0.522E-10	
	DxD	12_22	11_12	12_11	22_12	22_11	22_22	12_12	11_11	11_22
	Estimate	20.779	17.762	1.223	0.952	-0.298	-3.676	-4.656	-5.466	*****
4	22.093546	15	26.660750				HCW_Phen	AA	0.673E-09	
	AxA	2_1	1_2	1_1	2_2					
	Estimate	6.596	1.358	-2.411	-4.537					
4	24.183919	21	32.544201				HCW_Phen	I	0.200E-08	
4	32.508042	27	4.577749				HCW_Phen	AA	0.103E-08	
	AxA	2_2	1_1	2_1	1_2					
	Estimate	4.521	3.280	-0.227	-53.859					
4	78.143883	20	23.593803				HCW_Phen	I	0.000E+00	
4	78.143883	20	23.593803				HCW_Phen	AA	0.257E-09	
	AxA	1_1	2_2	2_1	1_2					
	Estimate	4.321	2.085	-3.062	-3.652					
4	90.314494	21	27.983449				HCW_Phen	DD	0.477E-10	
	DxD	12_22	11_12	22_12	12_11	22_11	11_11	12_12	11_22	22_22
	Estimate	134.206	13.545	4.298	0.286	0.057	-0.518	-5.790	-57.823	*****
4	90.314494	21	27.983449				HCW_Phen	I	0.170E-08	
4	95.336471	8	85.786900				HCW_Phen	AA	0.846E-09	
	AxA	1_1	2_2	1_2	2_1					
	Estimate	3.818	2.313	-2.489	-3.922					
5	26.986116	24	12.775555				HCW_Phen	AA	0.136E-08	
	AxA	1_1	2_2	1_2	2_1					
	Estimate	5.444	2.285	-0.956	-9.434					
5	115.279103	13	15.178253				HCW_Phen	AA	0.208E-08	
	AxA	2_2	1_1	2_1	1_2					
	Estimate	6.905	1.586	-0.595	-14.921					
6	8.478503	26	10.854941				HCW_Phen	AD	0.210E-08	
	AxD	1_22	1_11	2_12	2_11	2_22	1_12			
	Estimate	8.616	4.699	2.763	-2.187	-3.634	-6.351			
6	23.738304	12	69.345857				HCW_Phen	I	0.507E-10	
6	40.893067	10	50.353345				HCW_Phen	AA	0.232E-09	
	AxA	1_2	2_1	2_2	1_1					
	Estimate	4.530	2.002	-2.317	-4.666					
6	86.237705	11	82.282516				HCW_Phen	AA	0.182E-09	
	AxA	2_2	1_1	2_1	1_2					
	Estimate	7.872	1.306	-2.621	-3.928					
6	86.237705	11	82.282516				HCW_Phen	I	0.137E-08	
6	112.098405	17	32.943354				HCW_Phen	I	0.000E+00	
7	1.838926	8	84.178715				HCW_Phen	I	0.354E-09	
7	1.838926	8	84.178715				HCW_Phen	AA	0.814E-09	
	AxA	2_2	1_1	2_1	1_2					
	Estimate	2.907	2.815	-2.053	-5.262					
7	6.206051	18	8.729030				HCW_Phen	AA	0.107E-08	
	AxA	1_2	2_1	1_1	2_2					
	Estimate	4.054	2.422	-2.266	-4.187					
7	6.206051	18	8.765998				HCW_Phen	AA	0.443E-10	
	AxA	1_2	2_1	1_1	2_2					
	Estimate	4.196	2.711	-2.550	-4.416					
7	6.248226	18	8.729030				HCW_Phen	AA	0.215E-09	
	AxA	1_2	2_1	1_1	2_2					
	Estimate	4.248	2.105	-2.455	-4.669					
7	6.248226	18	8.765998				HCW_Phen	AA	0.476E-10	
	AxA	1_2	2_1	1_1	2_2					
	Estimate	4.229	2.293	-2.700	-4.664					
7	6.248226	18	8.765998				HCW_Phen	I	0.102E-08	
7	78.987067	20	63.332241				HCW_Phen	I	0.000E+00	
7	78.987067	20	63.332241				HCW_Phen	AA	0.417E-11	
	AxA	1_1	2_2	1_2	2_1					
	Estimate	4.919	2.394	-0.388	-29.979					
7	85.645433	7	93.289032				HCW_Phen	AA	0.155E-08	
	AxA	2_2	1_1	2_1	1_2					
	Estimate	9.332	0.237	-2.180	-5.829					



	Estimate	29.633	7.258	4.226	1.545	-1.065	-1.575	-5.934	-24.101	-33.279
13	54.663649	16	20.232360			HCW_Phen		DD	0.177E-08	
	DxD	12_11	22_12	11_12	12_22	11_22	22_22	12_12	11_11	22_11
	Estimate	98.832	8.207	3.825	0.369	-0.241	-0.533	-5.942	-73.764	*****
14	68.304416	15	35.437164			HCW_Phen		DD	0.110E-08	
	DxD	22_12	12_22	12_11	11_12	11_11	11_22	12_12	22_11	22_22
	Estimate	100.075	36.519	2.107	0.448	-0.115	-1.718	-8.166	-28.178	*****
14	68.304416	15	35.537347			HCW_Phen		DD	0.444E-09	
	DxD	22_12	12_22	12_11	11_12	11_11	11_22	12_12	22_11	22_22
	Estimate	101.470	40.164	2.124	0.455	-0.116	-1.774	-8.217	-28.506	*****
15	48.068620	29	41.939696			HCW_Phen		AA	0.168E-08	
	AxA	2_2	1_1	1_2	2_1					
	Estimate	3.471	2.361	-1.350	-7.514					
16	20.232360	27	31.426001			HCW_Phen		DD	0.688E-09	
	DxD	11_22	11_11	12_12	22_22	22_11	22_12	12_11	12_22	11_12
	Estimate	202.629	24.406	4.548	1.447	-0.022	-0.139	-0.878	-24.735	*****
16	37.024624	16	42.386711			HCW_Phen		I	0.294E-09	
18	9.060482	18	14.058919			HCW_Phen		AA	0.147E-08	
	AxA	2_2	1_1	1_2	2_1					
	Estimate	5.829	1.498	-1.764	-5.493					
18	9.060482	18	15.069718			HCW_Phen		AA	0.154E-08	
	AxA	2_1	1_2	1_1	2_2					
	Estimate	7.231	0.848	-3.636	-4.570					
21	27.983449	22	60.459451			HCW_Phen		I	0.000E+00	
21	27.983449	23	45.245295			HCW_Phen		DA	0.210E-08	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2			
	Estimate	60.283	7.799	0.165	-0.398	-3.669	-95.203			
21	27.983449	23	45.368338			HCW_Phen		DA	0.118E-08	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2			
	Estimate	62.258	7.643	0.129	-0.363	-3.049	-99.049			
21	27.983449	24	13.303547			HCW_Phen		I	0.188E-08	
21	27.983449	25	19.237205			HCW_Phen		I	0.692E-09	
21	27.983449	29	10.346832			HCW_Phen		DA	0.142E-08	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2			
	Estimate	76.976	10.969	0.130	-0.439	-3.184	-96.779			
21	27.983449	29	10.373536			HCW_Phen		DA	0.155E-08	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1			
	Estimate	76.911	10.900	0.129	-0.444	-3.163	-96.558			

**Table 3.S4 Pairwise interaction results – USFT**

100 MOST SIGNIFICANT RESULTS FROM THE PAIRWISE EPISTASIS ANALYSES

I = interaction effect between the two SNPs

AA = additive x additive effect, AD = additive x dominance effect

DA = dominance x additive effect, DD = dominance x dominance effect

Chr1	Locus1 (Mb)	Chr2	Locus2 (Mb)	Trait	Test	P_value				
1	28.622946	19	47.525998	USFT_Phen	AA	0.556E-09				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.005	0.005	-0.005	-0.006					
1	31.659179	19	47.525998	USFT_Phen	AA	0.441E-09				
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.007	0.003	-0.003	-0.009					
1	42.890259	5	2.586338	USFT_Phen	I	0.000E+00				
1	42.890259	5	2.586338	USFT_Phen	AA	0.451E-09				
	AxA	1_2	2_1	2_2	1_1					
	Estimate	0.014	0.002	-0.002	-0.015					
1	52.607381	8	86.780525	USFT_Phen	I	0.742E-10				
1	52.607381	8	86.810388	USFT_Phen	I	0.000E+00				
1	52.607381	27	0.128800	USFT_Phen	DD	0.264E-09				
	DxD	22_22	22_11	12_12	11_22	11_11	11_12	12_11	12_22	22_12
	Estimate	0.400	0.016	0.008	0.008	0.000	-0.002	-0.002	-0.053	-0.068
1	62.765705	4	9.404614	USFT_Phen	I	0.000E+00				
1	65.624310	14	1.675278	USFT_Phen	I	0.000E+00				
1	67.702576	14	1.675278	USFT_Phen	I	0.000E+00				
1	67.734488	14	1.651311	USFT_Phen	I	0.000E+00				
1	67.734488	14	1.675278	USFT_Phen	I	0.000E+00				
1	87.100663	16	61.221637	USFT_Phen	AA	0.877E-09				
	AxA	1_2	2_1	1_1	2_2					
	Estimate	0.027	0.001	-0.005	-0.014					
1	131.261969	14	1.651311	USFT_Phen	I	0.000E+00				
1	131.423039	14	1.651311	USFT_Phen	I	0.000E+00				
1	131.751589	14	1.651311	USFT_Phen	I	0.000E+00				
1	137.467169	14	1.651311	USFT_Phen	I	0.000E+00				
1	137.467169	14	1.651311	USFT_Phen	DD	0.170E-09				
	DxD	11_22	22_22	12_12	11_11	22_11	12_11	22_12	11_12	12_22
	Estimate	0.187	0.033	0.010	0.005	0.001	-0.002	-0.004	-0.039	-0.075
1	140.364786	19	10.367765	USFT_Phen	I	0.000E+00				
1	140.364786	19	11.199299	USFT_Phen	I	0.538E-09				
1	140.364786	19	11.238727	USFT_Phen	I	0.000E+00				
1	140.364786	19	11.435269	USFT_Phen	I	0.000E+00				
1	140.364786	29	28.908928	USFT_Phen	I	0.000E+00				
1	141.531071	9	75.896385	USFT_Phen	I	0.000E+00				
2	19.259873	18	34.817149	USFT_Phen	I	0.000E+00				
2	80.569319	12	4.614286	USFT_Phen	DA	0.510E-09				
	DxA	13_1	11_1	12_2	11_2	22_2	12_1			
	Estimate	0.026	0.012	0.003	-0.002	-0.004	-0.020			
3	47.849415	16	73.711818	USFT_Phen	I	0.528E-09				
3	110.242807	6	21.473715	USFT_Phen	I	0.961E-09				
4	19.855620	22	21.547900	USFT_Phen	AA	0.370E-09				
	AxA	2_1	1_2	2_2	1_1					
	Estimate	0.009	0.004	-0.002	-0.011					
4	96.034920	13	2.924417	USFT_Phen	I	0.943E-09				
4	112.944762	20	47.776960	USFT_Phen	AA	0.363E-10				
	AxA	2_2	1_1	2_1	1_2					
	Estimate	0.037	0.001	-0.004	-0.006					
5	0.864491	18	58.260521	USFT_Phen	DD	0.655E-09				
	DxD	12_11	11_12	22_12	12_22	22_22	11_22	12_12	22_11	11_11
	Estimate	0.036	0.016	0.003	0.001	-0.001	-0.003	-0.006	-0.017	-0.440
5	17.507982	15	71.467340	USFT_Phen	DD	0.493E-09				
	DxD	22_12	12_22	12_11	11_12	11_11	11_22	12_12	22_11	22_22
	Estimate	0.019	0.012	0.010	0.006	-0.005	-0.006	-0.011	-0.015	-0.016
5	104.899417	14	1.651311	USFT_Phen	I	0.000E+00				
5	104.899417	14	1.675278	USFT_Phen	I	0.000E+00				
6	21.473715	20	67.576765	USFT_Phen	I	0.000E+00				
6	55.077357	16	80.527514	USFT_Phen	AA	0.593E-09				
	AxA	1_2	2_1	1_1	2_2					
	Estimate	0.011	0.002	-0.004	-0.007					
6	55.103665	16	80.527514	USFT_Phen	AA	0.106E-09				

	AxA	2_2	1_1	2_1	1_2							
	Estimate	0.012	0.002	-0.004	-0.007							
6	55.103665	16	80.527514			USFT_Phen	I	0.468E-09				
6	55.124345	16	80.527514			USFT_Phen	AA	0.142E-09				
	AxA	1_2	2_1	1_1	2_2							
	Estimate	0.012	0.002	-0.004	-0.007							
6	62.206143	12	51.461607			USFT_Phen	AD	0.579E-09				
	AxD	2_11	1_12	2_22	1_22	2_12	1_11					
	Estimate	0.018	0.008	0.002	-0.003	-0.006	-0.029					
6	99.128925	22	24.112377			USFT_Phen	AA	0.231E-09				
	AxA	1_2	2_1	1_1	2_2							
	Estimate	0.012	0.002	-0.003	-0.013							
6	99.128925	22	24.112377			USFT_Phen	I	0.666E-09				
7	37.983978	19	27.444684			USFT_Phen	I	0.000E+00				
7	60.219706	9	75.896385			USFT_Phen	I	0.488E-09				
7	60.219706	9	75.896385			USFT_Phen	DD	0.557E-09				
	DxD	22_22	11_22	12_12	22_11	11_11	12_11	11_12	22_12	12_22		
	Estimate	0.321	0.012	0.005	0.002	0.001	-0.001	-0.002	-0.014	-0.037		
7	70.483529	16	76.091078			USFT_Phen	I	0.640E-09				
7	70.483529	20	67.576765			USFT_Phen	I	0.000E+00				
7	84.145466	22	19.668749			USFT_Phen	I	0.138E-09				
7	84.145466	22	19.668749			USFT_Phen	DD	0.671E-09				
	DxD	11_11	11_22	12_12	22_11	22_22	22_12	12_22	12_11	11_12		
	Estimate	0.132	0.026	0.010	0.006	0.001	-0.003	-0.005	-0.025	-0.046		
7	107.186560	21	71.109676			USFT_Phen	AA	0.156E-09				
	AxA	2_2	1_1	1_2	2_1							
	Estimate	0.010	0.003	-0.005	-0.006							
8	17.129059	14	1.651311			USFT_Phen	I	0.000E+00				
8	17.129059	14	1.651311			USFT_Phen	AD	0.227E-09				
	AxD	2_22	1_12	2_11	1_11	2_12	1_22					
	Estimate	0.037	0.008	0.001	-0.001	-0.005	-0.064					
8	17.129059	14	1.675278			USFT_Phen	I	0.000E+00				
8	17.129059	14	1.675278			USFT_Phen	AD	0.416E-10				
	AxD	2_22	1_12	2_11	1_11	2_12	1_22					
	Estimate	0.039	0.008	0.001	-0.001	-0.005	-0.069					
8	45.549903	28	26.362076			USFT_Phen	I	0.000E+00				
9	2.867786	28	20.096355			USFT_Phen	I	0.640E-09				
9	7.602607	26	44.141331			USFT_Phen	I	0.580E-09				
9	75.896385	11	6.758495			USFT_Phen	I	0.000E+00				
9	75.896385	15	76.960713			USFT_Phen	I	0.000E+00				
9	104.196469	14	1.651311			USFT_Phen	I	0.000E+00				
9	104.196469	14	1.675278			USFT_Phen	I	0.000E+00				
10	3.488689	14	1.651311			USFT_Phen	I	0.530E-09				
10	3.559435	13	2.939325			USFT_Phen	I	0.000E+00				
10	3.559435	14	1.651311			USFT_Phen	I	0.951E-10				
10	3.559435	14	1.675278			USFT_Phen	I	0.000E+00				
10	4.383927	14	1.651311			USFT_Phen	I	0.000E+00				
10	4.383927	14	1.675278			USFT_Phen	I	0.000E+00				
10	60.527567	13	3.427049			USFT_Phen	AD	0.846E-09				
	AxD	2_22	2_11	1_12	1_11	1_22	2_12					
	Estimate	0.010	0.008	0.006	-0.006	-0.007	-0.008					
11	101.452171	21	48.686907			USFT_Phen	DD	0.293E-09				
	DxD	12_22	22_12	11_12	12_11	11_11	22_11	12_12	11_22	22_22		
	Estimate	0.051	0.026	0.002	0.001	0.000	-0.004	-0.006	-0.011	-0.440		
12	24.696637	20	67.576765			USFT_Phen	I	0.151E-09				
12	46.711181	28	28.185694			USFT_Phen	DA	0.421E-09				
	DxA	13_1	12_2	11_1	11_2	12_1	22_2					
	Estimate	0.047	0.005	0.003	-0.001	-0.010	-0.024					
13	2.686675	20	67.576765			USFT_Phen	DD	0.660E-09				
	DxD	11_11	22_11	12_12	11_22	22_22	12_22	22_12	11_12	12_11		
	Estimate	0.258	0.016	0.006	0.004	0.001	-0.001	-0.003	-0.036	-0.056		
13	2.924417	18	49.611296			USFT_Phen	I	0.129E-09				
13	22.418071	13	73.600600			USFT_Phen	AA	0.515E-09				
	AxA	1_2	2_1	1_1	2_2							
	Estimate	0.007	0.003	-0.003	-0.012							
13	66.009719	14	1.651311			USFT_Phen	AD	0.953E-09				
	AxD	1_22	2_12	1_11	2_11	1_12	2_22					
	Estimate	0.101	0.004	0.002	0.000	-0.010	-0.028					
13	66.009719	14	1.675278			USFT_Phen	I	0.000E+00				
13	66.009719	14	1.675278			USFT_Phen	AD	0.182E-10				
	AxD	1_22	2_12	1_11	2_11	1_12	2_22					

	Estimate	0.124	0.004	0.002	-0.001	-0.011	-0.030			
14	1.651311	17		25.625689		USFT_Phen		I	0.000E+00	
14	1.651311	26		26.369697		USFT_Phen		I	0.000E+00	
14	1.651311	26		26.369697		USFT_Phen		DA	0.562E-10	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1			
	Estimate	0.110	0.004	0.002	-0.001	-0.010	-0.030			
14	1.675278	17		25.625689		USFT_Phen		I	0.000E+00	
14	1.675278	22		56.526462		USFT_Phen		DA	0.748E-09	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1			
	Estimate	0.050	0.009	0.001	-0.002	-0.005	-0.061			
14	1.675278	22		56.526462		USFT_Phen		DA	0.748E-09	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1			
	Estimate	0.050	0.009	0.001	-0.002	-0.005	-0.061			
14	1.675278	26		9.550462		USFT_Phen		I	0.000E+00	
14	1.675278	26		9.550462		USFT_Phen		DA	0.271E-09	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1			
	Estimate	0.110	0.003	0.002	0.000	-0.015	-0.026			
14	1.675278	26		26.369697		USFT_Phen		I	0.000E+00	
14	1.675278	26		26.369697		USFT_Phen		DA	0.759E-09	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1			
	Estimate	0.097	0.004	0.001	-0.001	-0.009	-0.030			
14	1.675278	29		41.552472		USFT_Phen		DD	0.573E-09	
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22	12_11	22_12
	Estimate	0.190	0.026	0.009	0.002	0.001	-0.001	-0.004	-0.019	-0.088
14	2.194228	17		64.142204		USFT_Phen		DA	0.812E-09	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2			
	Estimate	0.217	0.003	0.002	0.000	-0.025	-0.035			
15	18.605768	18		65.871032		USFT_Phen		I	0.000E+00	
16	41.283716	19		11.238727		USFT_Phen		I	0.000E+00	
16	41.345308	19		11.238727		USFT_Phen		I	0.000E+00	
17	65.748385	24		28.400451		USFT_Phen		AD	0.227E-09	
	AxD	1_11	1_22	2_12	2_22	2_11	1_12			
	Estimate	0.220	0.004	0.001	0.000	-0.007	-0.020			
17	65.748385	24		28.425789		USFT_Phen		AD	0.241E-09	
	AxD	1_22	1_11	2_12	2_11	2_22	1_12			
	Estimate	0.220	0.004	0.001	0.000	-0.007	-0.020			
18	13.197861	18		65.871032		USFT_Phen		AD	0.708E-09	
	AxD	2_11	1_12	2_22	1_22	2_12	1_11			
	Estimate	0.045	0.011	0.000	-0.001	-0.004	-0.125			
19	56.072306	26		34.020541		USFT_Phen		I	0.000E+00	
21	46.877052	24		61.262694		USFT_Phen		I	0.287E-09	
21	46.877052	24		61.262694		USFT_Phen		DD	0.873E-09	
	DxD	12_22	11_12	12_11	22_12	22_11	11_11	12_12	22_22	11_22
	Estimate	0.034	0.027	0.003	0.003	-0.001	-0.008	-0.008	-0.009	-0.221
27	12.900677	29		23.142122		USFT_Phen		AA	0.130E-09	
	AxA	1_1	2_2	1_2	2_1					
	Estimate	0.018	0.001	-0.002	-0.014					



**Table 3.S5 Pairwise interaction results – USIMF**

100 MOST SIGNIFICANT RESULTS FROM THE PAIRWISE EPISTASIS ANALYSES

I = interaction effect between the two SNPs

AA = additive x additive effect, AD = additive x dominance effect

DA = dominance x additive effect, DD = dominance x dominance effect

Chr1	Locus1 (Mb)	Chr2	Locus2 (Mb)	Trait	Test	P_value	
1	53.630189	1	54.032055	USIMF_Phen	AA	0.574E-08	
	AxA	1_1	2_2	2_1	1_2		
	Estimate	0.052	0.033	-0.092	-0.137		
1	64.990681	10	51.657226	USIMF_Phen	I	0.785E-08	
1	105.387594	2	13.343685	USIMF_Phen	I	0.376E-08	
1	112.863292	12	85.240392	USIMF_Phen	AD	0.623E-08	
	AxD	1_22	2_12	1_11	2_11	1_12	2_22
	Estimate	1.122	0.050	0.016	-0.006	-0.126	-0.602
2	9.040720	2	14.336331	USIMF_Phen	AA	0.804E-10	
	AxA	1_1	2_2	1_2	2_1		
	Estimate	0.276	0.020	-0.030	-0.126		
2	9.040720	2	14.336331	USIMF_Phen	I	0.212E-08	
2	21.427069	7	7.822111	USIMF_Phen	AA	0.102E-07	
	AxA	2_2	1_1	2_1	1_2		
	Estimate	0.095	0.036	-0.018	-0.183		
2	33.966115	6	63.137386	USIMF_Phen	AD	0.754E-08	
	AxD	1_12	2_22	2_11	2_12	1_11	1_22
	Estimate	0.086	0.081	0.073	-0.073	-0.096	-0.100
2	37.332808	27	17.007677	USIMF_Phen	AA	0.584E-08	
	AxA	1_2	2_1	2_2	1_1		
	Estimate	0.164	0.024	-0.013	-0.293		
2	37.358592	27	17.007677	USIMF_Phen	AA	0.541E-08	
	AxA	2_2	1_1	1_2	2_1		
	Estimate	0.165	0.024	-0.013	-0.293		
2	44.193968	12	54.304795	USIMF_Phen	AA	0.114E-08	
	AxA	2_1	1_2	2_2	1_1		
	Estimate	0.113	0.036	-0.034	-0.095		
2	44.193968	12	54.891162	USIMF_Phen	AA	0.142E-07	
	AxA	2_1	1_2	2_2	1_1		
	Estimate	0.141	0.025	-0.025	-0.126		
2	45.379047	8	9.427134	USIMF_Phen	I	0.188E-08	
2	53.431931	21	54.783091	USIMF_Phen	I	0.108E-07	
2	92.891116	17	18.937238	USIMF_Phen	AA	0.140E-07	
	AxA	2_1	1_2	2_2	1_1		
	Estimate	0.072	0.049	-0.023	-0.158		
2	92.923678	23	47.627822	USIMF_Phen	AA	0.548E-10	
	AxA	1_1	2_2	1_2	2_1		
	Estimate	0.161	0.029	-0.057	-0.075		
2	92.923678	23	47.627822	USIMF_Phen	I	0.269E-08	
2	98.764258	5	83.013004	USIMF_Phen	AA	0.619E-08	
	AxA	1_2	2_1	2_2	1_1		
	Estimate	0.321	0.011	-0.052	-0.061		
3	0.932938	27	25.424395	USIMF_Phen	AA	0.138E-07	
	AxA	2_2	1_1	1_2	2_1		
	Estimate	0.064	0.051	-0.020	-0.158		
3	0.932938	27	25.555089	USIMF_Phen	AA	0.468E-08	
	AxA	2_1	1_2	1_1	2_2		
	Estimate	0.073	0.051	-0.022	-0.147		
3	61.384165	4	112.815065	USIMF_Phen	AA	0.101E-07	
	AxA	1_2	2_1	2_2	1_1		
	Estimate	0.104	0.039	-0.025	-0.120		
3	68.702409	4	17.978432	USIMF_Phen	AA	0.553E-08	
	AxA	1_2	2_1	2_2	1_1		
	Estimate	0.112	0.036	-0.042	-0.076		
3	121.228888	9	8.355654	USIMF_Phen	DA	0.126E-07	
	DxA	22_2	11_2	12_1	11_1	13_1	12_2
	Estimate	0.224	0.160	0.038	-0.032	-0.045	-0.174
4	15.110552	8	69.431109	USIMF_Phen	AA	0.254E-08	
	AxA	1_1	2_2	2_1	1_2		
	Estimate	0.061	0.060	-0.058	-0.067		
4	15.110552	8	69.455202	USIMF_Phen	AA	0.790E-08	
	AxA	1_2	2_1	2_2	1_1		

	Estimate	0.061	0.058	-0.059	-0.065					
4	45.079239	27	9.916357			USIMF_Phen	I	0.146E-07		
4	46.205276	8	74.997682			USIMF_Phen	AA	0.332E-08		
	AxA	2_1	1_2	2_2	1_1					
	Estimate	0.161	0.019	-0.050	-0.107					
4	50.417818	17	7.472614			USIMF_Phen	AD	0.161E-07		
	AxD	1_11	2_12	1_22	2_22	1_12	2_11			
	Estimate	0.177	0.100	0.024	-0.032	-0.064	-0.303			
4	69.831612	11	73.976212			USIMF_Phen	AA	0.724E-08		
	AxA	1_2	2_1	2_2	1_1					
	Estimate	0.074	0.052	-0.052	-0.064					
4	82.615365	22	60.877108			USIMF_Phen	AA	0.143E-07		
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.130	0.028	-0.033	-0.092					
4	107.806828	20	37.596667			USIMF_Phen	AA	0.111E-07		
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.178	0.025	-0.016	-0.184					
4	116.693828	15	13.519590			USIMF_Phen	DD	0.316E-08		
	DxD	12_11	11_12	22_12	12_22	22_22	11_22	12_12	22_11	11_11
	Estimate	0.204	0.147	0.106	0.065	-0.068	-0.090	-0.118	-0.178	-0.249
4	117.955439	5	88.978964			USIMF_Phen	AD	0.120E-07		
	AxD	1_12	2_22	2_11	2_12	1_11	1_22			
	Estimate	0.100	0.080	0.053	-0.065	-0.089	-0.133			
5	49.999826	7	58.333017			USIMF_Phen	I	0.623E-08		
5	72.082602	15	14.062514			USIMF_Phen	AA	0.549E-08		
	AxA	2_2	1_1	2_1	1_2					
	Estimate	0.155	0.027	-0.021	-0.150					
5	72.390588	21	7.311519			USIMF_Phen	AA	0.156E-07		
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.057	0.056	-0.031	-0.108					
5	89.098980	26	43.506424			USIMF_Phen	AA	0.127E-07		
	AxA	2_1	1_2	1_1	2_2					
	Estimate	0.164	0.019	-0.053	-0.070					
5	89.098980	26	43.528093			USIMF_Phen	AA	0.115E-07		
	AxA	2_1	1_2	1_1	2_2					
	Estimate	0.165	0.020	-0.052	-0.070					
6	4.746604	15	57.313281			USIMF_Phen	AA	0.428E-09		
	AxA	1_2	2_1	1_1	2_2					
	Estimate	0.144	0.034	-0.053	-0.060					
6	4.746604	15	57.313281			USIMF_Phen	I	0.163E-07		
6	39.529973	6	81.767374			USIMF_Phen	AA	0.109E-07		
	AxA	2_1	1_2	2_2	1_1					
	Estimate	0.089	0.029	-0.049	-0.105					
6	49.233626	13	6.822805			USIMF_Phen	DD	0.158E-07		
	DxD	22_22	11_22	12_12	22_11	11_11	12_11	11_12	22_12	12_22
	Estimate	0.178	0.156	0.111	0.098	0.081	-0.077	-0.108	-0.122	-0.173
6	49.233626	13	6.853227			USIMF_Phen	DD	0.921E-08		
	DxD	22_22	11_22	12_12	22_11	11_11	12_11	11_12	22_12	12_22
	Estimate	0.178	0.160	0.113	0.100	0.083	-0.079	-0.109	-0.125	-0.174
6	52.394411	17	32.788648			USIMF_Phen	I	0.746E-08		
6	65.728229	15	38.379083			USIMF_Phen	DA	0.809E-08		
	DxA	12_1	11_2	22_2	12_2	13_1	11_1			
	Estimate	0.193	0.071	0.018	-0.038	-0.106	-0.325			
6	81.169373	20	7.204354			USIMF_Phen	AA	0.139E-07		
	AxA	1_2	2_1	1_1	2_2					
	Estimate	0.181	0.020	-0.041	-0.066					
6	99.028913	22	30.598794			USIMF_Phen	AD	0.280E-08		
	AxD	2_11	1_12	2_22	1_22	2_12	1_11			
	Estimate	0.311	0.075	0.026	-0.023	-0.087	-0.270			
7	2.714979	25	10.073572			USIMF_Phen	AA	0.813E-08		
	AxA	1_1	2_2	1_2	2_1					
	Estimate	0.146	0.020	-0.029	-0.140					
7	14.372254	7	44.658442			USIMF_Phen	AD	0.141E-07		
	AxD	1_12	2_22	2_11	2_12	1_11	1_22			
	Estimate	0.179	0.099	0.014	-0.040	-0.061	-0.493			
7	43.224755	28	35.504637			USIMF_Phen	AA	0.726E-08		
	AxA	1_1	2_2	1_2	2_1					
	Estimate	0.066	0.057	-0.045	-0.088					
7	43.966998	11	79.889994			USIMF_Phen	AA	0.666E-08		
	AxA	2_1	1_2	1_1	2_2					
	Estimate	0.112	0.030	-0.029	-0.128					

7	92.347964	24	33.367651		USIMF_Phen	AA	0.202E-08		
	AxA	1_2	2_1	1_1	2_2				
	Estimate	0.106	0.036	-0.023	-0.134				
8	4.998120	12	29.031334		USIMF_Phen	I	0.945E-08		
8	4.998120	12	29.031334		USIMF_Phen	DA	0.108E-07		
	DxA	11_1	12_2	13_1	22_2	12_1	11_2		
	Estimate	0.201	0.131	0.013	-0.027	-0.054	-0.487		
8	8.010905	20	49.090857		USIMF_Phen	I	0.788E-08		
8	8.010905	20	49.190161		USIMF_Phen	I	0.125E-07		
8	8.010905	20	49.251437		USIMF_Phen	I	0.780E-08		
8	21.161582	16	14.934067		USIMF_Phen	AA	0.151E-07		
	AxA	2_1	1_2	2_2	1_1				
	Estimate	0.102	0.034	-0.041	-0.072				
8	21.267518	16	14.934067		USIMF_Phen	AA	0.108E-07		
	AxA	2_1	1_2	2_2	1_1				
	Estimate	0.103	0.035	-0.042	-0.073				
8	49.283003	21	54.879224		USIMF_Phen	DD	0.581E-08		
	DxD	11_22	11_11	12_12	22_22	22_11	22_12	12_11	12_22
	Estimate	0.626	0.245	0.099	0.052	0.030	-0.039	-0.079	-0.147
9	24.949214	25	2.152451		USIMF_Phen	I	0.560E-08		
9	33.378478	22	16.661066		USIMF_Phen	DA	0.840E-08		
	DxA	11_1	12_2	13_1	22_2	12_1	11_2		
	Estimate	0.213	0.082	0.033	-0.036	-0.073	-0.185		
9	103.822440	26	16.888208		USIMF_Phen	I	0.000E+00		
10	7.943252	16	14.500474		USIMF_Phen	DA	0.714E-08		
	DxA	11_2	12_1	22_2	12_2	13_1	11_1		
	Estimate	0.131	0.128	0.029	-0.059	-0.061	-0.262		
10	29.836917	21	37.522232		USIMF_Phen	DD	0.145E-07		
	DxD	11_12	12_11	12_22	22_12	22_22	12_12	22_11	11_22
	Estimate	1.943	1.011	0.012	0.005	-0.001	-0.026	-0.043	-0.174
10	38.986604	21	37.026202		USIMF_Phen	DD	0.106E-07		
	DxD	11_22	22_22	12_12	11_11	22_11	12_11	22_12	11_12
	Estimate	0.827	0.193	0.115	0.084	0.017	-0.032	-0.055	-0.251
10	41.389231	16	4.966340		USIMF_Phen	AA	0.259E-08		
	AxA	2_2	1_1	1_2	2_1				
	Estimate	0.162	0.024	-0.021	-0.161				
10	94.030802	20	46.500354		USIMF_Phen	DD	0.134E-07		
	DxD	11_12	12_11	22_12	12_22	22_22	11_22	12_12	22_11
	Estimate	0.346	0.329	0.037	0.034	-0.013	-0.099	-0.114	-0.129
11	5.593781	12	44.800140		USIMF_Phen	DD	0.596E-08		
	DxD	11_12	12_11	12_22	22_12	22_22	12_12	22_11	11_22
	Estimate	0.243	0.222	0.051	0.042	-0.020	-0.093	-0.103	-0.132
11	5.964756	29	11.099809		USIMF_Phen	I	0.588E-08		
11	11.634312	25	24.661275		USIMF_Phen	I	0.834E-08		
11	23.235749	13	31.750419		USIMF_Phen	AA	0.337E-08		
	AxA	2_2	1_1	1_2	2_1				
	Estimate	0.327	0.013	-0.014	-0.240				
11	52.045971	29	29.952558		USIMF_Phen	AA	0.753E-09		
	AxA	1_2	2_1	1_1	2_2				
	Estimate	0.237	0.015	-0.024	-0.219				
11	52.155944	29	29.952558		USIMF_Phen	AA	0.123E-08		
	AxA	1_2	2_1	1_1	2_2				
	Estimate	0.235	0.015	-0.023	-0.214				
11	95.026013	21	33.570843		USIMF_Phen	I	0.907E-08		
12	51.973484	16	61.531686		USIMF_Phen	DA	0.143E-07		
	DxA	11_2	12_1	22_2	13_1	12_2	11_1		
	Estimate	0.137	0.101	0.030	-0.049	-0.062	-0.233		
12	70.260457	15	70.284345		USIMF_Phen	DD	0.119E-07		
	DxD	22_22	22_11	12_12	11_22	11_11	11_12	12_11	22_12
	Estimate	1.439	0.118	0.105	0.098	0.014	-0.038	-0.042	-0.272
13	2.416151	16	45.083406		USIMF_Phen	DD	0.267E-08		
	DxD	22_22	22_11	12_12	11_22	11_11	11_12	12_11	22_22
	Estimate	1.967	1.143	0.088	0.014	0.012	-0.012	-0.080	-0.104
13	2.416151	16	45.150817		USIMF_Phen	DD	0.296E-08		
	DxD	22_22	22_11	12_12	11_22	11_11	11_12	12_11	22_22
	Estimate	1.965	1.139	0.088	0.014	0.013	-0.012	-0.080	-0.102
13	2.416151	16	45.309651		USIMF_Phen	DD	0.779E-08		
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22	12_11
	Estimate	1.915	1.125	0.086	0.015	0.010	-0.012	-0.071	-0.108
13	2.416151	16	45.376614		USIMF_Phen	DD	0.802E-08		
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22	12_11

	Estimate	1.919	1.120	0.085	0.015	0.011	-0.012	-0.073	-0.105	-0.691
13	56.642417	21	33.401904			USIMF_Phen		I	0.314E-08	
14	47.733142	24	52.728523			USIMF_Phen		AD	0.415E-08	
	AxD	1_11	2_12	1_22	2_22	1_12	2_11			
	Estimate	0.153	0.073	0.061	-0.038	-0.090	-0.143			
14	68.026718	14	72.959368			USIMF_Phen		AA	0.161E-07	
	AxA	2_1	1_2	1_1	2_2					
	Estimate	0.242	0.013	-0.036	-0.057					
15	29.113608	20	60.722662			USIMF_Phen		I	0.749E-08	
17	1.715231	23	13.185570			USIMF_Phen		I	0.106E-07	
17	2.209997	29	9.850630			USIMF_Phen		AD	0.518E-08	
	AxD	1_11	2_12	1_22	2_22	1_12	2_11			
	Estimate	0.847	0.061	0.009	-0.006	-0.091	-0.485			
17	4.675045	23	23.283512			USIMF_Phen		AA	0.939E-08	
	AxA	1_1	2_2	2_1	1_2					
	Estimate	0.107	0.025	-0.022	-0.168					
17	9.472973	27	31.080330			USIMF_Phen		DD	0.940E-08	
	DxD	22_11	11_11	12_12	22_22	11_22	12_22	11_12	12_11	22_12
	Estimate	1.566	0.181	0.125	0.075	0.009	-0.029	-0.038	-0.513	-0.550
19	40.081544	20	6.638385			USIMF_Phen		DA	0.996E-08	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1			
	Estimate	0.601	0.047	0.024	-0.009	-0.126	-0.245			
21	25.876602	28	40.093345			USIMF_Phen		I	0.000E+00	
21	25.876602	28	40.093345			USIMF_Phen		DA	0.370E-08	
	DxA	12_2	11_1	13_1	12_1	22_2	11_2			
	Estimate	0.172	0.100	0.019	-0.041	-0.073	-0.458			
22	4.443783	23	8.513266			USIMF_Phen		AA	0.735E-08	
	AxA	2_2	1_1	2_1	1_2					
	Estimate	0.066	0.055	-0.029	-0.112					
22	5.579646	26	41.041883			USIMF_Phen		AA	0.152E-08	
	AxA	1_1	2_2	1_2	2_1					
	Estimate	0.117	0.039	-0.034	-0.091					
22	5.629829	26	41.041883			USIMF_Phen		AA	0.613E-08	
	AxA	2_1	1_2	2_2	1_1					
	Estimate	0.081	0.050	-0.026	-0.130					
22	53.724697	28	45.700407			USIMF_Phen		AA	0.121E-08	
	AxA	2_2	1_1	2_1	1_2					
	Estimate	0.122	0.028	-0.020	-0.173					
23	3.787479	28	39.165936			USIMF_Phen		AA	0.294E-08	
	AxA	1_1	2_2	1_2	2_1					
	Estimate	0.183	0.020	-0.057	-0.070					
24	41.603109	28	42.049080			USIMF_Phen		I	0.109E-07	
24	52.165674	27	25.480002			USIMF_Phen		DD	0.148E-07	
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22	12_11	22_12
	Estimate	0.438	0.335	0.106	0.043	0.028	-0.037	-0.092	-0.163	-0.462
24	52.545933	24	54.252151			USIMF_Phen		AA	0.144E-07	
	AxA	1_1	2_2	1_2	2_1					
	Estimate	0.216	0.011	-0.012	-0.334					

**Table 3.S6 Pairwise interaction results – USWT**

100 MOST SIGNIFICANT RESULTS FROM THE PAIRWISE EPISTASIS ANALYSES

I = interaction effect between the two SNPs

AA = additive x additive effect, AD = additive x dominance effect

DA = dominance x additive effect, DD = dominance x dominance effect

Chr1	Locus1 (Mb)	Chr2	Locus2 (Mb)	Trait	Test	P_value
1	15.071831	5	94.597691	USWT_Phen	AA	0.169E-08
	AxA	1_2	2_1	1_1	2_2	
	Estimate	5.893	3.094	-2.131	-16.582	
1	35.993477	2	68.795555	USWT_Phen	DD	0.221E-08
	DxD	11_22	22_22	12_12	11_11	22_11 12_11 22_12 11_12 12_22
	Estimate	117.683	111.784	7.537	0.909	0.461 -0.542 -6.538 -7.403 *****
1	35.993477	2	68.821062	USWT_Phen	DD	0.216E-08
	DxD	11_22	22_22	12_12	11_11	22_11 12_11 22_12 11_12 12_22
	Estimate	118.313	111.893	7.506	0.945	0.463 -0.538 -6.505 -7.010 *****
1	44.752791	4	18.689580	USWT_Phen	DD	0.180E-09
	DxD	12_11	11_12	22_12	12_22	22_22 11_22 12_12 22_11 11_11
	Estimate	44.259	13.787	5.958	1.254	-0.734 -1.726 -8.441 -37.448 *****
1	67.212088	14	47.480687	USWT_Phen	AA	0.163E-08
	AxA	2_1	1_2	1_1	2_2	
	Estimate	22.756	1.091	-4.071	-9.138	
1	79.629427	20	58.449212	USWT_Phen	I	0.208E-08
1	119.421373	12	78.540440	USWT_Phen	DD	0.137E-08
	DxD	12_11	11_12	22_12	12_22	22_22 11_22 12_12 22_11 11_11
	Estimate	42.827	36.366	2.308	1.756	-0.356 -6.791 -10.082 -23.095 *****
1	120.752725	27	32.303544	USWT_Phen	I	0.408E-09
1	124.196948	27	30.510177	USWT_Phen	AA	0.174E-08
	AxA	1_2	2_1	2_2	1_1	
	Estimate	24.861	1.106	-2.568	-8.129	
1	147.383154	2	68.795555	USWT_Phen	I	0.000E+00
1	147.383154	2	68.795555	USWT_Phen	DD	0.270E-10
	DxD	12_22	11_12	22_12	12_11	22_11 11_11 12_12 22_22 11_22
	Estimate	129.882	15.908	4.455	0.725	-0.367 -1.328 -8.927 ***** *****
1	147.383154	2	68.821062	USWT_Phen	I	0.000E+00
1	147.383154	2	68.821062	USWT_Phen	DD	0.402E-10
	DxD	12_22	11_12	22_12	12_11	22_11 11_11 12_12 22_22 11_22
	Estimate	128.393	15.808	4.346	0.717	-0.357 -1.318 -8.566 ***** *****
1	147.383154	2	68.849454	USWT_Phen	DD	0.169E-08
	DxD	12_11	11_12	22_12	12_22	22_22 11_22 12_12 22_11 11_11
	Estimate	138.058	16.494	4.691	0.641	-0.328 -1.158 -9.135 ***** *****
2	46.700438	19	2.321350	USWT_Phen	I	0.199E-08
2	52.069973	9	104.071394	USWT_Phen	I	0.127E-08
2	52.392645	2	53.951533	USWT_Phen	AA	0.111E-08
	AxA	1_2	2_1	1_1	2_2	
	Estimate	19.270	0.977	-8.156	-14.257	
2	53.431931	28	28.342445	USWT_Phen	AA	0.209E-09
	AxA	1_2	2_1	2_2	1_1	
	Estimate	10.340	2.759	-2.050	-13.449	
2	58.745694	16	35.272426	USWT_Phen	I	0.000E+00
2	60.308526	23	42.357362	USWT_Phen	AA	0.367E-09
	AxA	2_1	1_2	1_1	2_2	
	Estimate	4.729	3.624	-2.323	-12.871	
2	68.795555	4	46.084089	USWT_Phen	DD	0.152E-08
	DxD	22_22	22_11	12_12	11_11	11_22 11_12 12_22 12_11 22_12
	Estimate	120.705	105.419	8.307	1.345	0.328 -0.684 -4.273 -16.453 *****
2	68.795555	7	64.659545	USWT_Phen	DD	0.422E-09
	DxD	22_11	22_22	12_12	11_11	11_22 11_12 12_22 12_11 22_12
	Estimate	177.766	88.071	7.656	1.255	0.345 -0.611 -3.992 -16.481 *****
2	68.795555	8	64.337449	USWT_Phen	DA	0.197E-08
	DxA	22_2	12_1	11_2	11_1	12_2 13_1
	Estimate	127.029	6.629	0.607	-0.483	-7.014 *****
2	68.795555	10	85.336874	USWT_Phen	I	0.227E-08
2	68.795555	11	0.189568	USWT_Phen	DD	0.103E-08
	DxD	22_11	22_22	12_12	11_22	11_11 11_12 12_11 12_22 22_12
	Estimate	125.749	107.595	8.115	1.125	0.438 -0.631 -5.485 -11.554 *****
2	68.795555	14	63.353353	USWT_Phen	DD	0.292E-09
	DxD	22_11	22_22	12_12	11_11	11_22 11_12 12_22 12_11 22_12
	Estimate	348.238	54.273	11.037	4.173	0.161 -0.858 -1.900 -45.420 *****

2	68.795555	18	28.487717		USWT_Phen	DD	0.543E-09	
	DxD	22_12	12_22	12_11	11_12	11_11	11_22	12_12
	Estimate	117.887	12.797	4.824	0.676	-0.207	-0.996	-7.086
								*****
2	68.795555	19	22.265346		USWT_Phen	DA	0.249E-09	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1	
	Estimate	97.565	15.497	0.395	-1.049	-4.485	*****	
2	68.795555	19	23.291700		USWT_Phen	DA	0.196E-08	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	95.148	22.901	0.294	-1.516	-3.593	*****	
2	68.795555	19	23.349878		USWT_Phen	DA	0.156E-08	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1	
	Estimate	95.411	23.213	0.298	-1.554	-3.625	*****	
2	68.795555	19	23.477516		USWT_Phen	DA	0.227E-08	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	94.365	16.765	0.304	-1.250	-3.757	*****	
2	68.795555	19	23.519485		USWT_Phen	DA	0.207E-08	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	94.437	17.253	0.302	-1.268	-3.734	*****	
2	68.795555	19	24.544704		USWT_Phen	DA	0.982E-09	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	95.377	17.374	0.328	-1.188	-3.933	*****	
2	68.795555	19	25.260008		USWT_Phen	DA	0.275E-09	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	104.186	11.432	0.448	-0.829	-4.986	*****	
2	68.795555	19	53.606174		USWT_Phen	DD	0.125E-08	
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22
	Estimate	168.679	83.455	8.585	1.931	0.245	-0.770	-3.515
								-22.765
								*****
2	68.795555	23	40.672572		USWT_Phen	I	0.000E+00	
2	68.795555	25	6.766419		USWT_Phen	DD	0.135E-08	
	DxD	22_22	22_11	12_12	11_11	11_22	11_12	12_22
	Estimate	201.102	68.593	6.950	0.719	0.623	-0.683	-8.496
								-10.097
								*****
2	68.821062	3	76.091197		USWT_Phen	DD	0.231E-08	
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22
	Estimate	125.013	104.831	7.542	0.693	0.440	-0.599	-6.187
								-6.773
								*****
2	68.821062	4	46.084089		USWT_Phen	DD	0.218E-08	
	DxD	22_22	22_11	12_12	11_11	11_22	11_12	12_22
	Estimate	120.067	102.187	8.099	1.337	0.330	-0.665	-4.094
								-16.368
								*****
2	68.821062	7	64.659545		USWT_Phen	DD	0.591E-09	
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22
	Estimate	175.010	87.398	7.454	1.240	0.341	-0.603	-3.779
								-16.422
								*****
2	68.821062	8	64.337449		USWT_Phen	DA	0.189E-08	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1	
	Estimate	127.287	6.432	0.612	-0.486	-6.973	*****	
2	68.821062	11	0.189568		USWT_Phen	DD	0.837E-09	
	DxD	22_11	22_22	12_12	11_22	11_11	11_12	12_11
	Estimate	126.535	108.994	8.131	1.145	0.444	-0.639	-5.353
								-11.494
								*****
2	68.821062	12	62.550434		USWT_Phen	DD	0.201E-08	
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22
	Estimate	179.539	78.747	8.561	1.549	0.289	-0.619	-3.206
								-14.672
								*****
2	68.821062	14	63.353353		USWT_Phen	DD	0.272E-09	
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22
	Estimate	350.787	54.239	10.795	4.160	0.164	-0.870	-1.891
								-45.162
								*****
2	68.821062	18	28.487717		USWT_Phen	DD	0.295E-09	
	DxD	22_12	12_22	12_11	11_12	11_11	11_22	12_12
	Estimate	120.446	13.094	4.596	0.697	-0.177	-1.066	-7.123
								*****
2	68.821062	19	22.265346		USWT_Phen	DA	0.305E-09	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1	
	Estimate	97.435	14.609	0.382	-1.020	-4.334	*****	
2	68.821062	19	23.349878		USWT_Phen	DA	0.250E-08	
	DxA	22_2	12_1	11_2	11_1	12_2	13_1	
	Estimate	94.973	22.526	0.288	-1.504	-3.484	*****	
2	68.821062	19	24.544704		USWT_Phen	DA	0.137E-08	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	94.974	16.850	0.319	-1.156	-3.792	*****	
2	68.821062	19	25.260008		USWT_Phen	DA	0.357E-09	
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	103.800	10.936	0.437	-0.811	-4.832	*****	
2	68.821062	19	53.606174		USWT_Phen	DD	0.140E-08	
	DxD	22_11	22_22	12_12	11_11	11_22	11_12	12_22
	Estimate	169.564	83.332	8.513	1.919	0.245	-0.779	-3.507
								-20.301
								*****
2	68.821062	23	40.672572		USWT_Phen	I	0.000E+00	

2	68.821062	25	6.766419			USWT_Phen	DD	0.127E-08
	DxD	22_22	22_11	12_12	11_11	11_22	11_12	12_22
	Estimate	201.689	68.796	6.791	0.721	0.625	-0.691	-8.381
2	68.849454	5	62.360769			USWT_Phen	DA	0.326E-09
	DxA	11_2	12_1	22_2	13_1	12_2	11_1	
	Estimate	126.688	8.755	0.538	-0.581	-7.800	*****	
2	68.849454	14	63.353353			USWT_Phen	DD	0.888E-09
	DxD	11_11	11_22	12_12	22_11	22_22	22_12	12_22
	Estimate	328.746	53.713	12.619	4.003	0.151	-0.806	-2.112
2	68.849454	19	22.265346			USWT_Phen	DA	0.780E-09
	DxA	11_2	12_1	22_2	13_1	12_2	11_1	
	Estimate	94.203	16.926	0.364	-0.977	-4.936	*****	
2	68.849454	22	5.524106			USWT_Phen	DD	0.107E-08
	DxD	11_22	11_11	12_12	22_11	22_22	22_12	12_22
	Estimate	118.994	117.868	8.244	0.687	0.580	-0.607	-5.947
2	68.849454	22	5.546862			USWT_Phen	DD	0.106E-08
	DxD	11_11	11_22	12_12	22_22	22_11	22_12	12_11
	Estimate	119.145	117.803	8.241	0.686	0.574	-0.614	-5.886
2	68.849454	23	40.672572			USWT_Phen	I	0.000E+00
4	43.600871	10	57.193699			USWT_Phen	AA	0.113E-08
	AxA	1_2	2_1	1_1	2_2			
	Estimate	13.488	1.405	-4.337	-13.171			
4	43.763107	10	57.193699			USWT_Phen	AA	0.129E-08
	AxA	2_2	1_1	2_1	1_2			
	Estimate	13.432	1.403	-4.310	-13.134			
4	62.416138	29	43.059581			USWT_Phen	AA	0.250E-08
	AxA	1_1	2_2	1_2	2_1			
	Estimate	6.048	3.007	-0.902	-29.670			
4	65.346870	5	120.016258			USWT_Phen	AA	0.144E-08
	AxA	2_1	1_2	1_1	2_2			
	Estimate	5.874	5.047	-2.047	-11.205			
4	69.500860	20	0.770272			USWT_Phen	AA	0.648E-09
	AxA	2_2	1_1	1_2	2_1			
	Estimate	5.020	4.469	-2.800	-10.853			
4	75.379515	14	46.959846			USWT_Phen	AA	0.943E-09
	AxA	2_1	1_2	2_2	1_1			
	Estimate	8.095	2.284	-2.619	-16.314			
4	87.437547	26	19.554675			USWT_Phen	AD	0.179E-08
	AxD	2_12	1_11	1_22	1_12	2_22	2_11	
	Estimate	13.718	7.133	2.047	-3.653	-6.147	-29.636	
4	92.678128	27	0.189179			USWT_Phen	I	0.296E-10
4	92.678128	27	0.212609			USWT_Phen	I	0.123E-08
4	92.678128	27	0.232640			USWT_Phen	I	0.000E+00
4	94.153292	25	32.725579			USWT_Phen	DA	0.567E-09
	DxA	11_2	22_2	12_1	13_1	12_2	11_1	
	Estimate	16.000	5.529	5.247	-3.822	-8.439	-11.604	
4	94.176209	25	32.725579			USWT_Phen	I	0.000E+00
4	94.176209	25	32.725579			USWT_Phen	DA	0.135E-10
	DxA	22_2	12_1	11_2	11_1	12_2	13_1	
	Estimate	18.143	5.827	5.488	-3.799	-9.432	-13.802	
4	95.336471	22	31.309419			USWT_Phen	I	0.790E-09
4	99.381115	5	78.115795			USWT_Phen	I	0.154E-09
5	6.242179	17	57.572420			USWT_Phen	I	0.215E-08
5	7.733715	14	23.737337			USWT_Phen	I	0.158E-08
5	10.788597	21	27.506924			USWT_Phen	I	0.155E-08
5	60.929878	10	98.420963			USWT_Phen	DD	0.179E-08
	DxD	12_22	22_12	11_12	12_11	11_11	22_11	12_12
	Estimate	25.902	20.252	3.290	2.175	-1.110	-4.486	-7.088
6	19.576414	14	32.153693			USWT_Phen	AA	0.118E-08
	AxA	2_1	1_2	2_2	1_1			
	Estimate	4.709	4.133	-1.397	-18.496			
6	114.426181	10	86.276682			USWT_Phen	DD	0.177E-08
	DxD	11_22	22_22	12_12	11_11	22_11	12_11	22_12
	Estimate	36.170	23.891	8.706	4.000	2.058	-2.959	-5.844
8	21.073477	11	28.082671			USWT_Phen	AA	0.125E-08
	AxA	2_2	1_1	2_1	1_2			
	Estimate	16.678	1.499	-1.321	-18.167			
8	31.721931	13	78.078786			USWT_Phen	AA	0.136E-08
	AxA	1_1	2_2	2_1	1_2			
	Estimate	7.110	3.112	-3.435	-8.250			
8	40.020068	15	41.455372			USWT_Phen	AD	0.119E-08

	AxD	2_11	1_12	2_22	1_22	2_12	1_11			
	Estimate	18.777	5.890	4.143	-2.754	-9.152	-11.524			
8	59.954021	8	108.257394			USWT_Phen		AA	0.140E-08	
	AxA	1_1	2_2	1_2	2_1					
	Estimate	4.499	3.440	-1.038	-23.437					
8	59.954021	8	108.288140			USWT_Phen		AA	0.527E-10	
	AxA	1_2	2_1	1_1	2_2					
	Estimate	5.253	3.381	-1.018	-27.538					
8	59.954021	8	108.288140			USWT_Phen		I	0.143E-08	
9	11.972126	11	70.016398			USWT_Phen		I	0.705E-10	
9	11.972126	11	70.016398			USWT_Phen		DD	0.895E-09	
	DxD	11_12	12_22	12_11	22_12	22_11	11_11	12_12	22_22	11_22
	Estimate	31.690	25.413	2.654	2.471	-0.894	-6.232	-7.842	-9.447	*****
9	91.405210	25	21.290132			USWT_Phen		I	0.107E-08	
10	4.888243	12	27.954424			USWT_Phen		I	0.129E-08	
12	30.751569	18	17.988032			USWT_Phen		AA	0.121E-08	
	AxA	1_1	2_2	2_1	1_2					
	Estimate	5.411	4.261	-3.930	-7.421					
13	71.489545	19	2.722336			USWT_Phen		AA	0.160E-08	
	AxA	2_1	1_2	1_1	2_2					
	Estimate	5.976	4.614	-1.149	-22.189					
13	76.910874	21	60.252111			USWT_Phen		AA	0.133E-08	
	AxA	2_1	1_2	2_2	1_1					
	Estimate	7.512	2.378	-0.422	-71.822					
14	18.208347	26	22.648404			USWT_Phen		AA	0.304E-09	
	AxA	2_2	1_1	1_2	2_1					
	Estimate	6.123	4.820	-3.526	-6.871					
14	18.208347	26	22.913016			USWT_Phen		AA	0.141E-09	
	AxA	2_1	1_2	1_1	2_2					
	Estimate	6.462	4.748	-3.744	-6.866					
14	18.331919	26	22.648404			USWT_Phen		AA	0.436E-09	
	AxA	2_2	1_1	1_2	2_1					
	Estimate	6.345	4.503	-3.454	-7.279					
14	18.331919	26	22.913016			USWT_Phen		AA	0.241E-10	
	AxA	2_1	1_2	1_1	2_2					
	Estimate	7.010	4.631	-3.831	-7.585					
16	18.401696	22	56.235584			USWT_Phen		I	0.106E-08	
16	76.275809	16	76.522826			USWT_Phen		AA	0.711E-09	
	AxA	2_1	1_2	2_2	1_1					
	Estimate	6.822	1.594	-1.266	-34.338					
17	3.599671	24	43.099447			USWT_Phen		AA	0.134E-08	
	AxA	2_2	1_1	1_2	2_1					
	Estimate	14.073	2.729	-1.258	-18.145					
18	8.234025	25	10.896067			USWT_Phen		DD	0.238E-08	
	DxD	22_22	11_22	22_11	12_12	11_11	11_12	12_11	12_22	22_12
	Estimate	32.289	10.407	10.061	9.302	3.207	-5.626	-5.630	-17.019	-18.339



**Table 3.S7 Pairwise interaction results – WW**

100 MOST SIGNIFICANT RESULTS FROM THE PAIRWISE EPISTASIS ANALYSES

I = interaction effect between the two SNPs

AA = additive x additive effect, AD = additive x dominance effect

DA = dominance x additive effect, DD = dominance x dominance effect

Chr1	Locus1 (Mb)	Chr2	Locus2 (Mb)	Trait	Test	P_value
1	21.538209	25	16.048464	WW_Phen	I	0.818E-09
1	62.184184	11	44.207605	WW_Phen	AA	0.110E-08
	AxA	2_1	1_2	2_2	1_1	
	Estimate	13.598	0.563	-1.111	-7.836	
1	69.520131	7	74.709675	WW_Phen	I	0.000E+00
1	78.449548	2	68.849454	WW_Phen	AD	0.112E-08
	AxD	1_11	2_12	1_22	2_22	1_12
	Estimate	43.210	3.588	0.197	-0.234	-2.626 *****
1	123.403379	4	110.420705	WW_Phen	AA	0.145E-08
	AxA	1_1	2_2	2_1	1_2	
	Estimate	6.789	1.296	-1.118	-6.160	
1	127.998001	2	68.795555	WW_Phen	DD	0.129E-08
	DxD	11_22	22_22	12_12	11_11	22_11
	Estimate	184.167	21.484	3.388	0.641	0.145 -0.271 -1.558 -6.904 *****
1	127.998001	2	68.821062	WW_Phen	DD	0.104E-08
	DxD	11_22	22_22	12_12	11_11	22_11
	Estimate	185.681	21.802	3.386	0.650	0.147 -0.280 -1.543 -6.853 *****
1	147.383154	2	68.795555	WW_Phen	DD	0.361E-10
	DxD	12_22	11_12	22_12	12_11	22_11
	Estimate	78.609	8.285	2.302	0.402	-0.201 -0.732 -4.794 -72.750 *****
1	147.383154	2	68.795555	WW_Phen	I	0.165E-08
1	147.383154	2	68.821062	WW_Phen	DD	0.460E-10
	DxD	12_22	11_12	22_12	12_11	22_11
	Estimate	78.083	8.288	2.226	0.398	-0.196 -0.731 -4.654 -72.522 *****
1	147.383154	2	68.849454	WW_Phen	DD	0.137E-09
	DxD	12_11	11_12	22_12	12_22	22_11
	Estimate	90.427	8.541	2.532	0.367	-0.187 -0.667 -5.172 -54.332 *****
2	16.484729	2	68.795555	WW_Phen	DD	0.151E-08
	DxD	12_22	11_12	22_12	12_11	22_11
	Estimate	58.339	3.420	3.200	0.290	-0.265 -0.396 -3.625 -94.426 *****
2	16.484729	2	68.821062	WW_Phen	DD	0.179E-08
	DxD	12_22	11_12	22_12	12_11	22_11
	Estimate	57.870	3.302	3.101	0.284	-0.257 -0.407 -3.540 -94.143 *****
2	47.873555	2	68.795555	WW_Phen	AD	0.990E-09
	AxD	2_22	1_12	2_11	1_11	2_12
	Estimate	43.588	7.355	0.168	-0.563	-1.794 *****
2	47.873555	2	68.821062	WW_Phen	AD	0.108E-08
	AxD	2_22	1_12	2_11	1_11	2_12
	Estimate	43.588	6.860	0.161	-0.539	-1.734 *****
2	47.873555	2	68.849454	WW_Phen	AD	0.222E-08
	AxD	2_11	1_12	2_22	1_22	2_11
	Estimate	42.034	8.848	0.176	-0.587	-2.115 *****
2	48.154122	2	68.795555	WW_Phen	AD	0.729E-09
	AxD	2_22	1_12	2_11	1_11	2_12
	Estimate	43.969	6.510	0.168	-0.498	-1.824 *****
2	48.154122	2	68.821062	WW_Phen	AD	0.805E-09
	AxD	2_22	1_12	2_11	1_11	2_12
	Estimate	43.941	6.146	0.162	-0.481	-1.769 *****
2	48.154122	2	68.849454	WW_Phen	AD	0.147E-08
	AxD	2_11	1_12	2_22	1_22	2_11
	Estimate	42.597	7.802	0.175	-0.519	-2.155 *****
2	52.392645	2	53.951533	WW_Phen	AA	0.160E-08
	AxA	1_2	2_1	1_1	2_2	
	Estimate	10.250	0.546	-5.045	-7.442	
2	68.119332	23	3.581582	WW_Phen	I	0.199E-08
2	68.795555	3	2.683082	WW_Phen	I	0.000E+00
2	68.795555	5	88.436433	WW_Phen	DD	0.203E-08
	DxD	22_11	22_22	12_12	11_11	11_22
	Estimate	106.635	27.181	3.217	0.405	0.205 -0.302 -2.521 -5.040 *****
2	68.795555	5	97.593586	WW_Phen	DD	0.173E-08
	DxD	22_12	12_22	12_11	11_12	11_11
	Estimate	69.784	5.088	2.714	0.303	-0.191 -0.482 -3.409 -66.735 *****

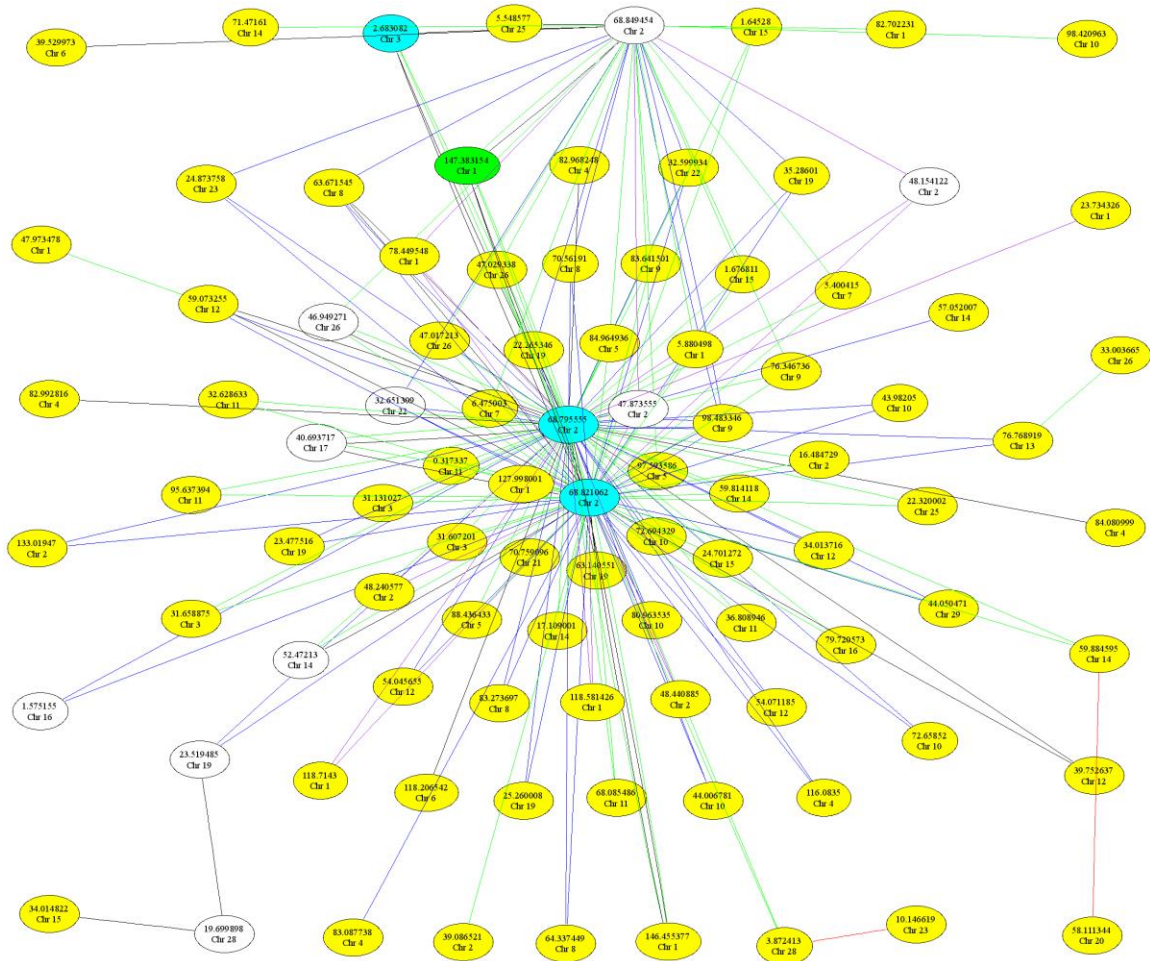
2	68.795555	7	6.475003		WW_Phen	DD	0.153E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	98.646	24.857	1.601	0.602	-0.139	-2.946
2	68.795555	8	63.671545		WW_Phen	I	0.164E-08
2	68.795555	9	76.346736		WW_Phen	DD	0.203E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	82.672	11.907	1.763	0.404	-0.127	-1.521
2	68.795555	9	83.641501		WW_Phen	DD	0.178E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	84.020	7.927	2.094	0.307	-0.168	-0.693
2	68.795555	13	76.768919		WW_Phen	DA	0.222E-08
	DxA	13_1	12_2	11_1	11_2	12_1	22_2
	Estimate	39.407	6.819	0.126	-0.585	-1.589	*****
2	68.795555	14	17.109001		WW_Phen	DD	0.160E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	84.564	6.013	2.475	0.348	-0.172	-0.591
2	68.795555	14	52.472130		WW_Phen	DD	0.416E-09
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	108.175	5.486	2.162	0.280	-0.195	-0.493
2	68.795555	16	1.575155		WW_Phen	DA	0.971E-09
	DxA	22_2	12_1	11_2	11_1	12_2	13_1
	Estimate	48.262	3.865	0.172	-0.328	-2.064	-89.954
2	68.795555	17	40.693717		WW_Phen	I	0.207E-09
2	68.795555	19	25.260008		WW_Phen	DA	0.212E-08
	DxA	13_1	12_2	11_1	11_2	12_1	22_2
	Estimate	57.101	5.082	0.207	-0.381	-2.263	*****
2	68.795555	21	70.759096		WW_Phen	DD	0.139E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	71.649	4.003	3.095	0.289	-0.254	-0.320
2	68.795555	22	32.599934		WW_Phen	DA	0.137E-08
	DxA	22_2	12_1	11_2	11_1	12_2	13_1
	Estimate	43.023	5.138	0.151	-0.421	-1.822	*****
2	68.795555	22	32.651309		WW_Phen	DA	0.966E-09
	DxA	13_1	12_2	11_1	11_2	12_1	22_2
	Estimate	43.417	5.028	0.153	-0.417	-1.859	*****
2	68.795555	25	22.320002		WW_Phen	DD	0.153E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	71.195	4.158	2.851	0.294	-0.254	-0.320
2	68.795555	26	46.949271		WW_Phen	DD	0.446E-09
	DxD	22_12	12_11	12_22	11_12	11_11	11_22
	Estimate	109.975	3.356	3.194	0.291	-0.250	-0.296
2	68.821062	3	2.683082		WW_Phen	I	0.000E+00
2	68.821062	5	88.436433		WW_Phen	DD	0.209E-08
	DxD	22_11	22_22	12_12	11_11	11_22	11_12
	Estimate	106.528	27.235	3.147	0.414	0.202	-0.302
2	68.821062	5	97.593586		WW_Phen	DD	0.156E-08
	DxD	22_12	12_22	12_11	11_12	11_11	11_22
	Estimate	70.072	5.005	2.682	0.305	-0.190	-0.489
2	68.821062	7	6.475003		WW_Phen	DD	0.137E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	98.902	24.981	1.574	0.605	-0.139	-2.949
2	68.821062	8	63.671545		WW_Phen	I	0.132E-08
2	68.821062	9	83.641501		WW_Phen	DD	0.205E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	83.764	7.906	2.045	0.305	-0.167	-0.685
2	68.821062	14	17.109001		WW_Phen	DD	0.176E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	84.366	6.009	2.386	0.344	-0.166	-0.601
2	68.821062	14	52.472130		WW_Phen	DD	0.408E-09
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	108.223	5.517	2.073	0.280	-0.195	-0.502
2	68.821062	16	1.575155		WW_Phen	DA	0.122E-08
	DxA	22_2	12_1	11_2	11_1	12_2	13_1
	Estimate	48.009	3.793	0.171	-0.328	-2.013	-89.409
2	68.821062	17	40.693717		WW_Phen	I	0.656E-09
2	68.821062	21	70.759096		WW_Phen	DD	0.127E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11
	Estimate	71.945	4.030	2.942	0.290	-0.256	-0.330
2	68.821062	22	32.599934		WW_Phen	DA	0.132E-08
	DxA	22_2	12_1	11_2	11_1	12_2	13_1
	Estimate	43.117	5.065	0.152	-0.423	-1.792	*****

2	68.821062	22	32.651309			WW_Phen	DA	0.758E-09
	DxA	13_1	12_2	11_1	11_2	12_1	22_2	
	Estimate	43.748	4.956	0.155	-0.420	-1.844	*****	
2	68.821062	25	22.320002			WW_Phen	DD	0.141E-08
	DxD	22_12	12_11	12_22	11_12	11_22	11_11	12_12
	Estimate	71.465	4.015	2.951	0.294	-0.267	-0.313	-3.129
2	68.821062	26	46.949271			WW_Phen	DD	0.573E-09
	DxD	22_12	12_11	12_22	11_12	11_11	11_22	12_12
	Estimate	109.278	3.292	3.103	0.288	-0.255	-0.287	-3.107
2	68.849454	3	2.683082			WW_Phen	I	0.814E-09
2	68.849454	7	5.400415			WW_Phen	DD	0.172E-08
	DxD	11_12	12_11	12_22	22_12	22_22	22_11	12_12
	Estimate	83.870	8.974	2.411	0.338	-0.176	-0.656	-4.599
2	68.849454	15	1.645280			WW_Phen	DD	0.194E-08
	DxD	11_12	12_22	12_11	22_12	22_11	22_22	12_12
	Estimate	133.403	5.742	2.191	0.287	-0.159	-0.563	-4.267
2	68.849454	15	1.676811			WW_Phen	DD	0.197E-08
	DxD	11_12	12_22	12_11	22_12	22_11	22_22	12_12
	Estimate	133.355	5.734	2.189	0.287	-0.159	-0.561	-4.264
2	68.849454	26	46.949271			WW_Phen	DD	0.629E-09
	DxD	11_12	12_11	12_22	22_12	22_11	22_22	12_12
	Estimate	87.982	3.904	3.640	0.289	-0.274	-0.279	-4.067
2	110.642302	4	106.321900			WW_Phen	I	0.000E+00
2	128.437731	23	8.965776			WW_Phen	AA	0.107E-08
	AxA	2_2	1_1	1_2	2_1			
	Estimate	4.090	1.688	-2.854	-2.881			
3	93.543745	22	24.112377			WW_Phen	AA	0.119E-08
	AxA	1_1	2_2	2_1	1_2			
	Estimate	15.590	2.226	-0.367	-21.675			
3	95.853703	12	50.505079			WW_Phen	AA	0.178E-08
	AxA	1_2	2_1	2_2	1_1			
	Estimate	25.601	0.584	-0.445	-13.440			
4	38.189188	28	5.506637			WW_Phen	I	0.190E-08
4	59.539393	26	20.100276			WW_Phen	AA	0.213E-08
	AxA	2_2	1_1	1_2	2_1			
	Estimate	4.615	1.352	-1.205	-6.900			
4	64.386271	13	44.086463			WW_Phen	DA	0.121E-09
	DxA	22_2	12_1	11_2	11_1	12_2	13_1	
	Estimate	17.914	3.382	1.662	-0.946	-5.276	-11.874	
4	106.383456	6	62.788712			WW_Phen	AA	0.149E-08
	AxA	1_1	2_2	2_1	1_2			
	Estimate	6.899	1.131	-1.996	-3.966			
4	112.934990	9	91.672738			WW_Phen	AA	0.719E-09
	AxA	2_1	1_2	2_2	1_1			
	Estimate	7.266	0.998	-2.548	-3.015			
4	118.717994	7	24.594731			WW_Phen	I	0.705E-09
5	10.862100	7	100.226236			WW_Phen	AA	0.671E-10
	AxA	1_2	2_1	1_1	2_2			
	Estimate	6.757	2.123	-0.813	-10.633			
6	30.812012	6	70.999090			WW_Phen	AA	0.179E-08
	AxA	2_1	1_2	1_1	2_2			
	Estimate	6.271	1.397	-2.160	-2.760			
6	104.789108	11	44.207605			WW_Phen	AA	0.163E-08
	AxA	2_1	1_2	2_2	1_1			
	Estimate	15.211	0.615	-0.925	-6.704			
7	26.110506	17	29.018684			WW_Phen	AA	0.131E-08
	AxA	1_1	2_2	1_2	2_1			
	Estimate	6.741	0.999	-2.535	-3.787			
7	36.884206	22	10.811371			WW_Phen	AA	0.347E-09
	AxA	1_1	2_2	1_2	2_1			
	Estimate	7.570	2.595	-0.599	-15.064			
7	36.884206	22	10.917454			WW_Phen	AA	0.692E-10
	AxA	1_1	2_2	1_2	2_1			
	Estimate	7.142	2.823	-0.646	-14.934			
7	68.516924	29	43.826144			WW_Phen	AA	0.147E-08
	AxA	1_2	2_1	2_2	1_1			
	Estimate	4.226	2.185	-0.788	-9.303			
7	101.153932	20	60.040554			WW_Phen	AA	0.631E-09
	AxA	2_1	1_2	1_1	2_2			
	Estimate	32.734	0.307	-1.723	-5.315			
8	62.891563	17	11.029290			WW_Phen	AA	0.158E-08

	AxA	1_2	2_1	2_2	1_1			
	Estimate	5.968	1.019	-3.956	-5.049			
9	13.418475	17	19.384978			WW_Phen	I	0.000E+00
9	13.418475	17	19.384978			WW_Phen	AA	0.154E-08
	AxA	1_2	2_1	2_2	1_1			
	Estimate	16.502	0.578	-0.747	-8.476			
9	35.187993	9	80.669302			WW_Phen	AA	0.139E-08
	AxA	2_1	1_2	1_1	2_2			
	Estimate	2.387	2.161	-0.863	-9.952			
10	73.585685	14	54.240653			WW_Phen	AA	0.135E-08
	AxA	1_2	2_1	2_2	1_1			
	Estimate	5.304	1.409	-1.902	-4.177			
10	73.655817	14	54.240653			WW_Phen	AA	0.666E-09
	AxA	1_2	2_1	2_2	1_1			
	Estimate	5.616	1.376	-1.858	-4.440			
10	73.655817	14	54.356138			WW_Phen	AA	0.205E-08
	AxA	1_1	2_2	2_1	1_2			
	Estimate	5.756	1.260	-1.907	-4.061			
10	100.472362	13	55.973589			WW_Phen	AA	0.190E-09
	AxA	2_2	1_1	1_2	2_1			
	Estimate	3.946	2.746	-0.700	-12.119			
11	41.833490	14	19.488209			WW_Phen	AA	0.215E-08
	AxA	1_1	2_2	1_2	2_1			
	Estimate	6.538	1.151	-1.615	-4.621			
11	44.122489	15	10.860778			WW_Phen	I	0.000E+00
13	1.966648	21	36.554417			WW_Phen	AA	0.214E-08
	AxA	1_1	2_2	2_1	1_2			
	Estimate	2.644	2.515	-1.894	-4.043			
14	29.942659	14	50.815043			WW_Phen	AA	0.853E-09
	AxA	1_2	2_1	1_1	2_2			
	Estimate	5.048	1.867	-1.050	-6.247			
14	49.385228	23	17.811197			WW_Phen	I	0.186E-08
14	74.753039	16	49.946542			WW_Phen	AA	0.101E-08
	AxA	2_1	1_2	1_1	2_2			
	Estimate	5.551	1.471	-1.061	-7.447			
14	74.753039	16	50.046534			WW_Phen	AA	0.225E-09
	AxA	2_1	1_2	1_1	2_2			
	Estimate	3.649	2.948	-0.688	-13.034			
14	74.753039	16	50.076028			WW_Phen	AA	0.289E-10
	AxA	2_1	1_2	1_1	2_2			
	Estimate	3.707	3.156	-0.700	-14.020			
14	74.753039	16	50.076028			WW_Phen	I	0.170E-08
15	2.378099	15	6.914643			WW_Phen	AA	0.206E-08
	AxA	2_2	1_1	1_2	2_1			
	Estimate	3.777	2.452	-2.081	-2.898			
15	36.917900	15	42.294852			WW_Phen	AA	0.216E-08
	AxA	1_2	2_1	2_2	1_1			
	Estimate	3.879	1.790	-2.333	-3.824			
16	32.063762	20	54.586441			WW_Phen	AA	0.838E-09
	AxA	2_2	1_1	2_1	1_2			
	Estimate	11.249	0.801	-1.703	-3.935			
17	69.111646	27	25.728096			WW_Phen	AA	0.118E-08
	AxA	1_1	2_2	1_2	2_1			
	Estimate	2.725	2.619	-2.387	-3.433			
19	23.519485	28	19.699898			WW_Phen	I	0.360E-09
21	53.525183	23	12.027791			WW_Phen	AA	0.187E-08
	AxA	2_2	1_1	1_2	2_1			
	Estimate	9.906	0.722	-0.957	-8.340			

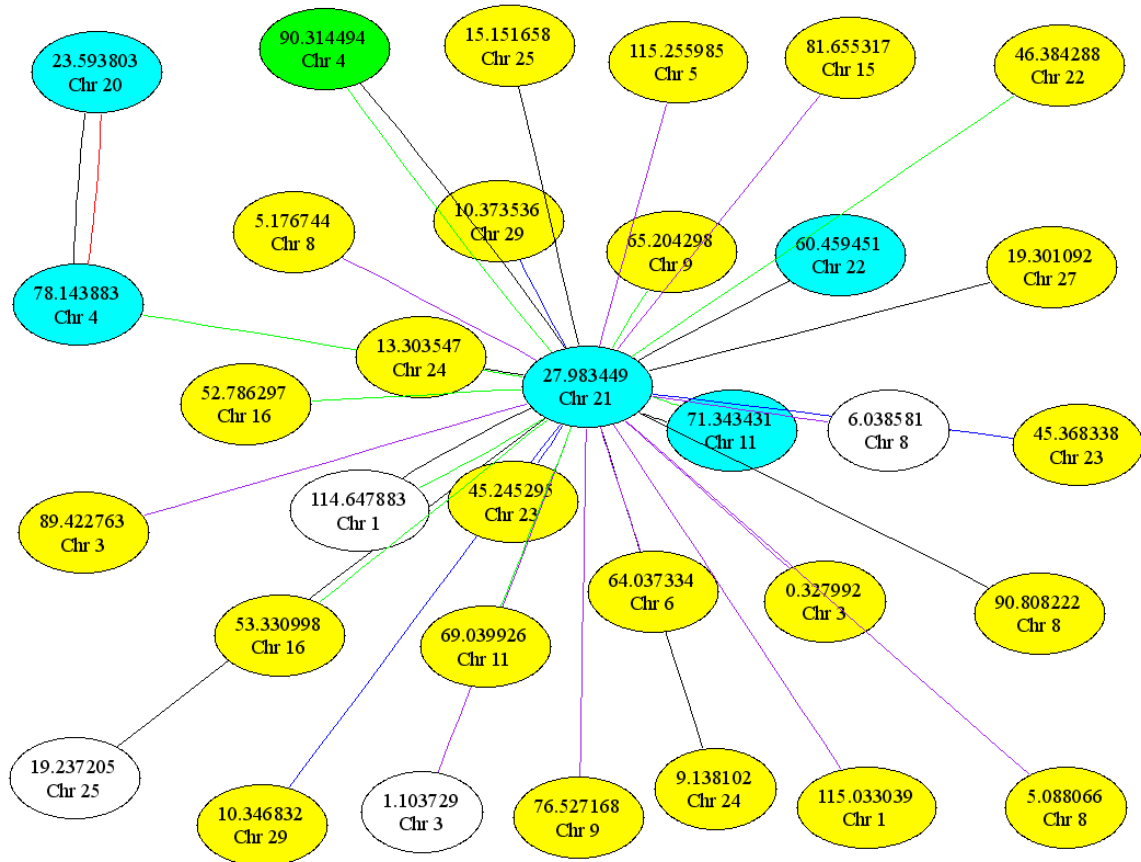
### Figure 3.S1 Top epistatic network for WW based on top 600 pairwise interactions

The largest epistatic network generated for weaning weight based on the top 600 pairwise interactions. Connecting line colours signify the type of epistatic interaction: black = overall interaction, purple = additive x dominance, blue = dominance x additive, and green = dominance x dominance. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$ , green =  $p < 10^{-10}$ , white =  $p < 10^{-9}$ , and yellow =  $p < 10^{-8}$ .



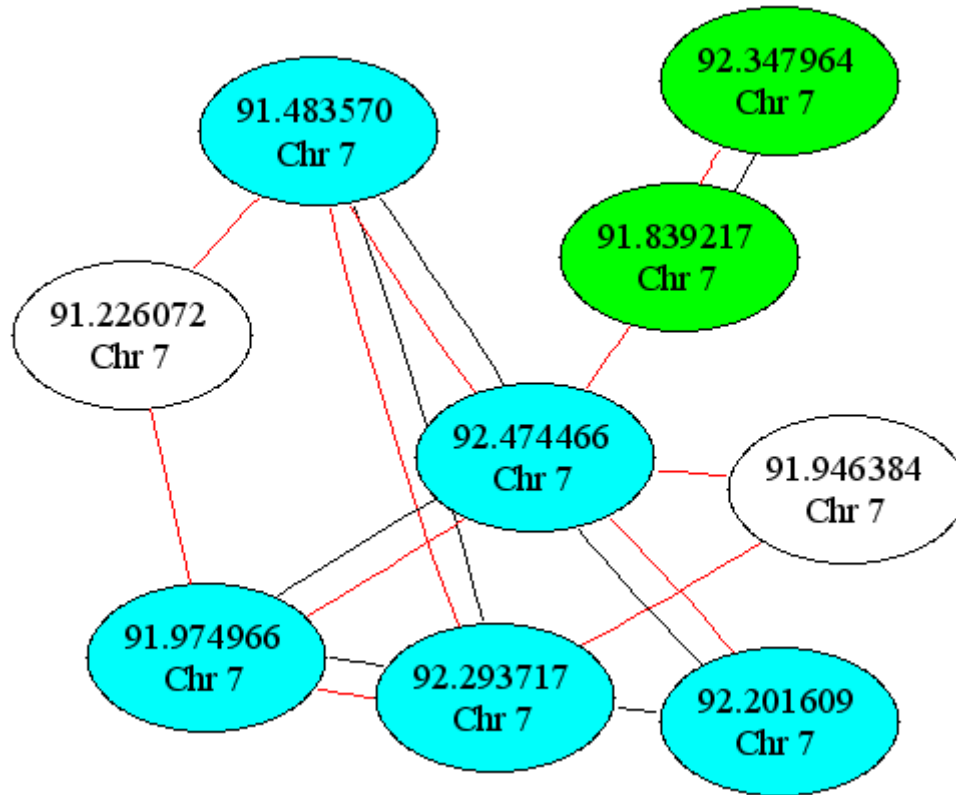
### Figure 3.S2 Top epistatic network for HCW based on top 600 pairwise interactions

The largest epistatic network generated for hot carcass weight based on the top 600 pairwise interactions. Connecting line colours signify the type of epistatic interaction: black = overall interaction, red = additive x additive, purple = additive x dominance, blue = dominance x additive, and green = dominance x dominance. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$ , green =  $p < 10^{-10}$ , white =  $p < 10^{-9}$ , and yellow =  $p < 10^{-8}$ .



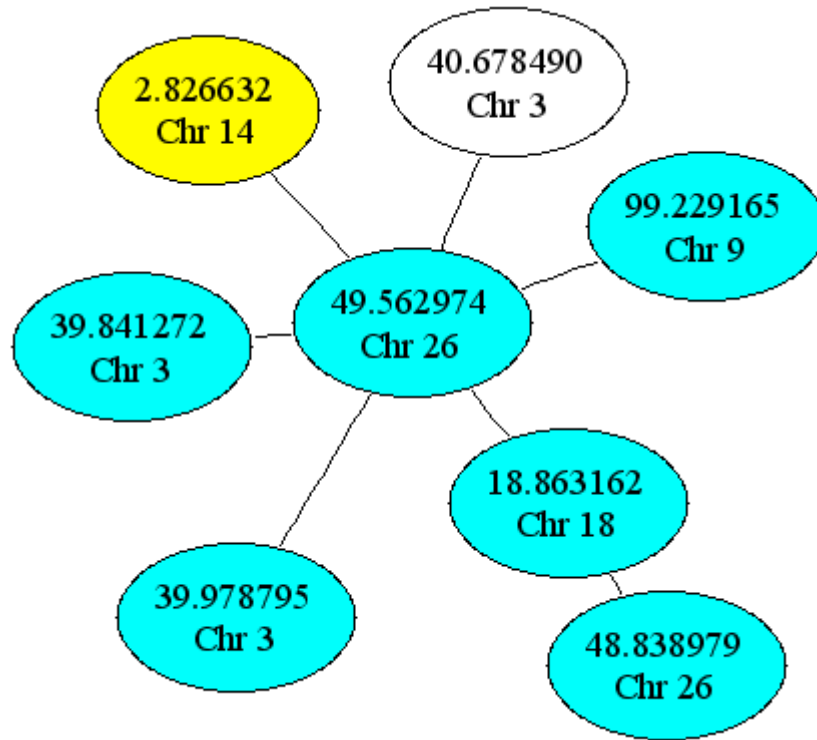
**Figure 3.S3 Top epistatic network for BW**

The largest epistatic network generated from the top 100 pairwise interactions for birth weight. Connecting line colours signify the type of epistatic interaction: black = overall interaction and red = additive x additive. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$ , green =  $p < 10^{-10}$  and white =  $p < 10^{-9}$ .



**Figure 3.S4 Second-ranked epistatic network for FAT**

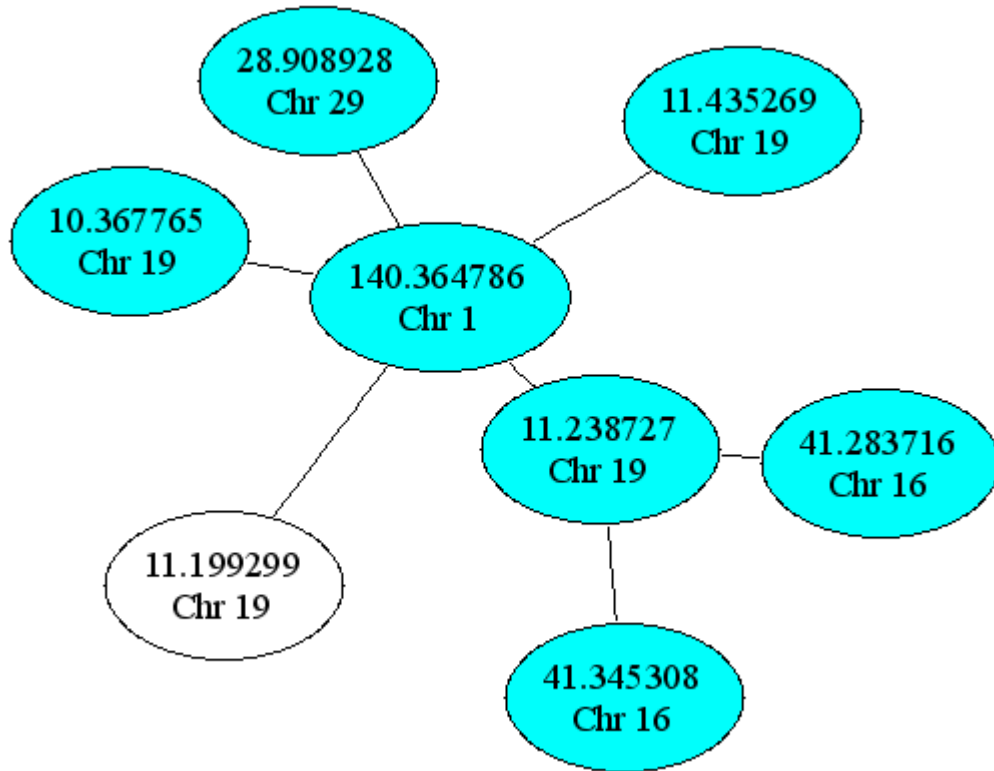
The second-largest epistatic network generated from the top 100 pairwise interactions for fat thickness. Connecting line colours signify the type of epistatic interaction: black = overall interaction. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$ , white =  $p < 10^{-9}$ , and yellow =  $p < 10^{-8}$ .





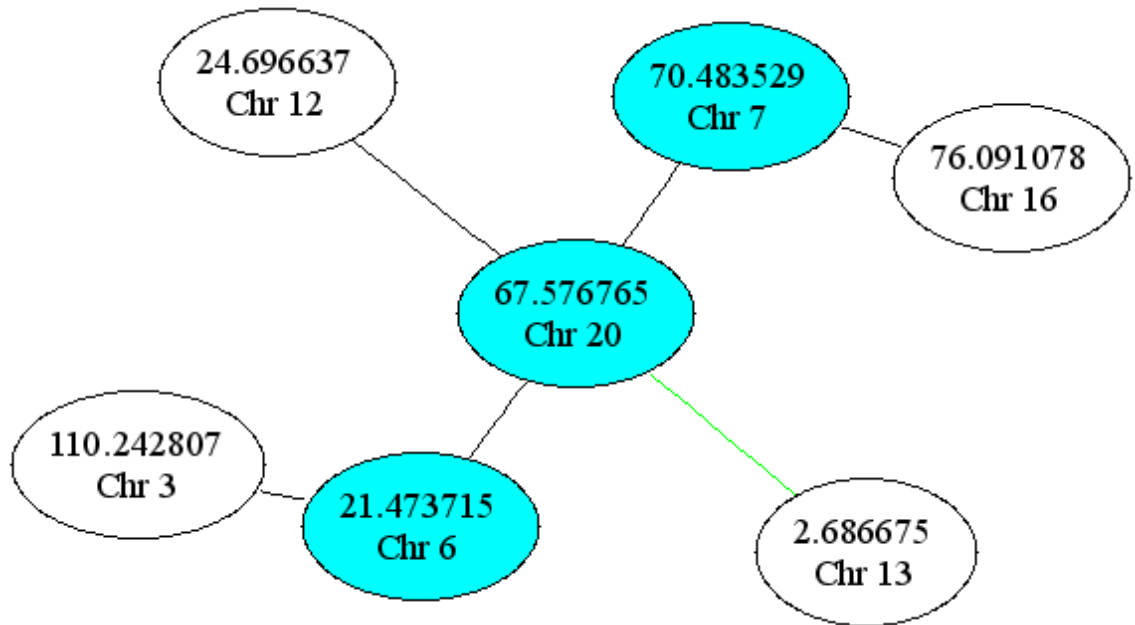
**Figure 3.S5 Second-ranked epistatic network for USFT**

The second-largest epistatic network generated from the top 100 pairwise interactions for ultrasound fat thickness. Connecting line colours signify the type of epistatic interaction: black = overall interaction. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$  and white =  $p < 10^{-9}$ .



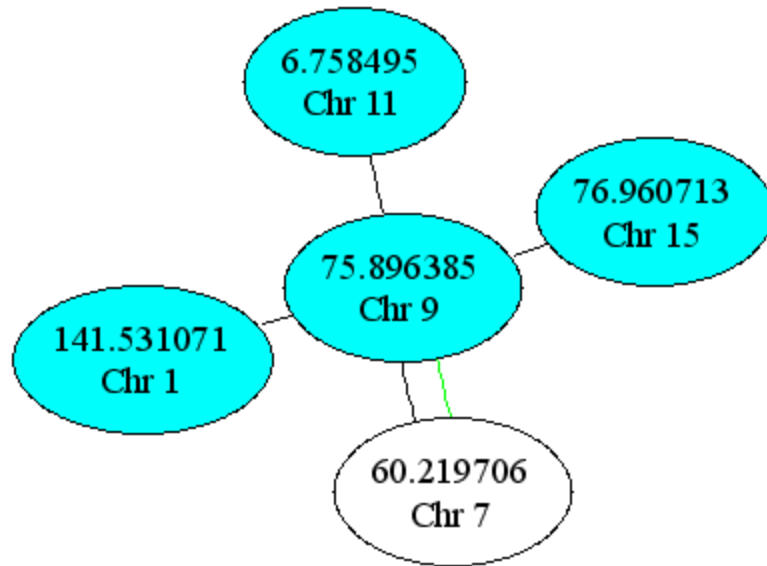
**Figure 3.S6 Third-ranked epistatic network for USFT**

The third-largest epistatic network generated from the top 100 pairwise interactions for ultrasound fat thickness. Connecting line colours signify the type of epistatic interaction: black = overall interaction and green = dominance x dominance. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$  and white =  $p < 10^{-9}$ .



**Figure 3.S7 Fourth-ranked epistatic network for USFT**

The third-largest epistatic network generated from the top 100 pairwise interactions for ultrasound fat thickness. Connecting line colours signify the type of epistatic interaction: black = overall interaction and green = dominance x dominance. Node colour indicates p-value significance with cyan =  $p < 10^{-11}$  and white =  $p < 10^{-9}$ .



## **4. *BMP3* does not cause the brachycephaly phenotype within**

### **Pug dogs**

**Holly R. Ramey<sup>1</sup>, Gary S. Johnson<sup>2</sup>, Robert D. Schnabel<sup>1</sup>, and Jeremy F. Taylor<sup>1§</sup>**

<sup>1</sup>Division of Animal Sciences, University of Missouri, Columbia MO 65211 USA

<sup>2</sup>Veterinary Pathobiology, School of Veterinary Medicine, University of Missouri,  
Columbia MO 65211 USA

<sup>§</sup>Corresponding author

Email addresses:

Holly R. Ramey ([hrrxb2@missouri.edu](mailto:hrrxb2@missouri.edu))

Gary S. Johnson ([johnsongs@missouri.edu](mailto:johnsongs@missouri.edu))

Robert D. Schnabel ([schnabelr@missouri.edu](mailto:schnabelr@missouri.edu))

Jeremy F. Taylor ([taylorjerr@missouri.edu](mailto:taylorjerr@missouri.edu))

#### **Corresponding Author:**

Dr. Jeremy F. Taylor

S135B Animal Sciences

920 East Campus Drive

University of Missouri

Columbia, MO 65211-5300

E-mail: [taylorjerr@missouri.edu](mailto:taylorjerr@missouri.edu)

Tel.: +1-573-884-4946

Fax: +1-573-882-6827

## Abstract

The dog genome represents a valuable source of breed-specific variation for the genomics research community. No other domesticated species exhibits such an extreme amount of variation via the creation of morphologically dissimilar breeds. Advancements in sequencing technologies and rapidly decreasing costs of whole genome sequencing allow the ability, for the first time, to study the variation within the whole-genome sequences of individuals from specific breeds for genotype-phenotype mapping studies. The brachycephaly phenotype is an example of a breed hallmark presumably caused by a mutation at a single major locus. Previous research has reported a causal mutation within *BMP3* at 5,231,894 bp on CFA32, which appears to be almost fixed within extremely brachycephalic breeds. In the whole genome sequences of four brachycephalic Pug dogs, we found all four dogs to be homozygous for the alternative wild type allele at this locus indicating that brachycephaly in Pugs is not caused by the *BMP3* mutation. To identify the mutation responsible for brachycephaly in Pugs, we first scanned the genomes of the four sequenced dogs to identify regions of homozygosity that were common to all four dogs. This revealed no regions of common homozygosity on CFA32, but eight large regions of common homozygosity were identified on CFA1, 10, 18, 21 and 25. The largest region extended for over 2.7 Mb and was located between 55.8-58.6 Mb on BTA1 and overlaps a previously identified selective sweep region in brachycephalic dogs. Within the region are three genes including *SMOC2* and *THBS2*, which are associated with cell-cell and cell-matrix interactions, and *PLN* that is important in cardiac diastolic function. Variants in the *SMOC2* gene have been implicated in orodental abnormalities. We sought to identify variants within this region for which alleles found in the Pug dogs

were not found in the whole genome sequences of 113 dogs from non-brachycephalic breeds. We identified only a single candidate for the causal mutation underlying the brachycephaly phenotype at 55,902,909 bp on CFA1 where the Pugs were homozygous for the non-reference allele, G, and all 113 controls were homozygous for the reference allele, C. This variant falls within the *SMOC2* gene where variants have previously been reported to have an effect on odontogenesis. Additional analysis is required to identify whether large insertion or deletion events exist within this region and to identify whether copy number variants may exist within the Pug dogs in this region before the causal mutation is identified.

### **Keywords**

brachycephaly, Pugs, *BMP3*, *SMOC2*, *THBS2*, candidate mutation, runs of homozygosity

### **Background**

Genome-wide association studies (GWAS) can be applied to individuals from different breeds to identify the variants responsible for the large phenotypic differences between breeds of dogs [138]. Understanding the genetic basis of breed hallmarks or phenotypes shared by subsets of breeds is of considerable interest to both dog breeders and researchers interested in the processes of breed formation following domestication. Considerable research has focussed on disease-related phenotypes and potential therapies [139-141] and characteristics such as size variation [20], furnishings [52], and skull shape morphology [142], however, opportunities still exist to identify other variants of large effect that underlie quantitative trait variation in dogs. As dogs are increasingly being sequenced to enable the identification of disease-causing variants and to study the origins of their domestication, these data can also be used to perform GWAS to identify large

effect variants. Performing GWAS on whole genome sequence has the advantage that many more variants are used to detect trait associations, resulting in many more loci being in strong linkage disequilibrium with the underlying causal variants and provided these have not been missed in the variant calling process, the causal variants will be included in the GWAS analysis. Few studies analysing deeply sequenced individuals have been performed in domesticated species. However, analysis pipelines and methodologies are now beginning to be developed to handle such large datasets [143, 144].

The Pug breed of dogs is brachycephalic. Literally meaning “shortened head,” the brachycephalic phenotype is associated with craniofacial development and morphology and is characterized by a shortened, up-turned snout or muzzle in many dog breeds. Brachycephalic breeds are of all sizes ranging from the toy-sized Pug and Cavalier King Charles to the medium-sized Boxer on up to the giant Dogue de Bordeaux breed. Brachycephalic breeds have many health-related issues, in particular breathing abnormalities which may require intervention [145-147]. As some groups of brachycephalic breeds, such as the Boxer, Bulldog, Boston Terrier, and Pekingese, have little phenotypic variation in their craniofacial features, some believe that there is a single allele responsible for brachycephaly that is fixed within these breeds [148].

Several studies have been performed to map the brachycephaly locus and quantitative trait loci (QTL) have been identified on *Canis familiaris* chromosomes (CFAs) 1, 5, 24, 30, 32, and on the X chromosome [51, 148, 149] which strongly suggests that the phenotype is not inherited as a simple Mendelian. In the Boxer, which is the breed of the reference genome sequence, two studies have linked brachycephaly to

CFA1 via across breed mapping [148] and selective sweep mapping [21]. One study presented data supporting a causal mutation where all extremely brachycephalic breeds share a missense mutation at position at 8,196,098 bp on CFA32 in the CanFam2 reference build, which is at 5,231,894 bp on CFA32 on the CanFam3.1 assembly [149]. This mutation lies within the bone morphogenetic protein 3 gene, *BMP3*, and the authors suggest the causality of the mutation via fine mapping and development and functional assessment in zebrafish. This was the first study of the phenotype that was able to suggest a causal mutation – the other studies concluded that none of the identified mutations could be convincingly linked to brachycephaly.

Using available whole-genome sequence data, we found no evidence for an association between this locus in 4 dogs sequenced from brachycephalic and 113 dogs sequenced from non-brachycephalic breeds. We found that dogs from non-brachycephalic breeds are frequently homozygous for the allele claimed to be responsible for the phenotype in severely brachycephalic breeds. Therefore, the goal of this study was to identify further candidates for the brachycephaly locus (or loci) within the extremely brachycephalic Pug breed.

## **Materials and Methods**

### *Sample selection, sequencing, and variant calling*

The samples selected for this analysis were determined by sequence availability from a genome resequencing project conducted at the University of Missouri – Columbia to identify disease-causing mutations in breed dogs and wild canids sequenced by collaborators to establish the origins of domesticated dogs. Overall, 102 individuals representing 48 domesticated breeds as well as one crossbred individual, a diverse group



of wolves, coyotes, a dingo and a jackal were available for this analysis. Most animals were sequenced due to their affected status for an inherited disease. Sequences were obtained using Illumina (San Diego, CA) and Applied Biosystems' SOLiD (Carlsbad, CA) sequencing chemistry. Sequences underwent filtering for quality control including duplicate removal, adapter trimming, and error correction before alignment to the CanFam3.1 reference genome.

Variant calling was performed on the alignments using NextGENe (Soft Genetics, State College, PA). Variants are defined as any position within the genome where a called allele differed from the reference, the non-reference allele. Instead of specifying hard quality control filters, an empirically derived overall mutation score based on five separate sub-scores, including read depth, and number of forward and reverse reads harbouring the variants, was used to identify variants that were likely to have been correctly identified. All variants were then uploaded to a custom designed database within mutually exclusive priority classes based upon the type of mutation (single base change, insertion, deletion, etc.) and annotation (untranslated region, CpG site, etc.) as well as any frameshift information, amino acid substitution or premature stop codon information. These classes allow for downstream filtering of the data based on classification or functional information. With this method all variants are called without excessive regard for accuracy resulting in a reasonably high Type I error rate, but with the maximum possible sensitivity for capturing rare variants.

### *Dataset filtering*

Once the variant calls were made for each of the 102 sequenced genomes, the data were empirically analysed to assess which of the samples should be used for the analysis. We removed any sequenced individual with less than 90% of the entire reference genome

**Table 4.1 Sample demography and classification**

<b>Breed</b>	<b>Average Coverage</b>	<b>% Genome Covered</b>	<b>Brachycephaly Status*</b>	<b>Brachy Genotype**</b>
Airedale Terrier	24.57	95.65	N	CC
Airedale Terrier	35.71	99.45	N	CC
Airedale Terrier	20.99	99.70	N	CC
Australian Shepherd	19.76	99.74	N	CC
Basenji	19.24	95.31	N	CC
Basenji	24.63	91.98	N	CC
Basenji	30.70	99.27	N	CC
Beagle	30.92	98.68	N	CC
Berger Picard	23.10	99.57	N	CC
Berger Picard	21.02	99.65	N	CC
Black Russian Terrier	35.80	99.49	N	CC
Border Collie	26.61	98.87	N	CC
Border Collie	11.02	91.96	N	CC
Border Collie	31.75	99.39	N	CC
Border Terrier	25.61	99.39	N	CC
Boxer	30.66	98.67	B	CC
Brittany Spaniel	12.60	90.49	N	CC
Cavalier King Charles Spaniel	12.79	94.06	B	CC
Chinese Crested	27.59	98.57	N	AA
Chinese Crested	33.23	99.60	N	CA
Chinook	26.59	98.27	N	CC
Clumber Spaniel	19.01	96.40	N	CC
Cocker Spaniel - American	24.25	96.55	B	CC
Coyote <sup>1</sup>	27.04	99.36	N	CC
Doberman	32.75	99.33	N	CC
Doberman	42.12	94.39	N	CC
Doberman	30.14	98.26	N	CC
English Cocker Spaniel	26.98	98.57	N	CC
English Pointer	24.55	96.13	N	CC
English Springer Spaniel	33.33	99.34	N	CC
German Shepherd	32.02	99.36	N	CC
Golden Retriever	21.62	99.41	N	CC
Golden Retriever	41.39	97.73	N	CC
Great Pyrenees	25.34	98.33	N	CC
Italian Greyhound	31.10	99.26	N	CC
Italian Greyhound	33.53	98.36	N	CC
Jack Russell Terrier	26.53	99.50	N	CA
Jack Russell Terrier	33.90	99.41	N	AC

Jack Russell Terrier	27.78	99.56	N	CC
Jackal <sup>1</sup>	14.91	97.07	N	CC
Kerry Blue Terrier	29.97	99.39	N	CC
Labrador Retriever	16.32	96.34	N	CC
Labrador Retriever	35.85	99.49	N	CC
Labrador Retriever	37.75	97.12	N	CC
Mastiff	20.74	98.98	B	CC
Newfoundland	42.52	91.62	B	CC
Norwegian Lundehund	31.28	97.82	N	CC
Pembroke Welsh Corgi	25.61	99.43	N	AA
Pembroke Welsh Corgi	32.08	99.38	N	AA
Pembroke Welsh Corgi	35.43	97.72	N	AA
Pointer	32.84	99.00	N	CC
Portuguese Podengo	31.33	99.40	N	AA
Portuguese Pointer	12.54	91.74	N	CC
Pug	16.21	97.83	B	AA
Pug	12.57	94.91	B	AA
Pug	13.61	93.23	B	AA
Pug	38.53	95.92	B	AA
Racing Greyhound	21.01	93.03	N	CC
Rhodesian Ridgeback	27.68	99.11	N	CC
Rottweiler	46.31	98.24	N	CC
Saint Bernard	31.96	99.22	N	CC
Saluki	21.54	95.18	N	CC
Scottish Deerhound	27.06	99.43	N	CC
Scottish Terrier	35.45	99.41	N	AA
Scottish Terrier	35.15	99.44	N	AA
Shetland Sheepdog	28.52	99.60	N	CC
Soft Coated Wheaten Terrier	18.38	99.53	N	CC
Soft Coated Wheaten Terrier	20.39	99.35	N	CC
Soft Coated Wheaten Terrier	29.36	95.31	N	CC
Soft Coated Wheaten Terrier	20.12	99.24	N	CC
Standard Poodle	23.57	99.36	N	CC
Standard Poodle	30.40	99.15	N	CC
Standard Poodle	15.41	94.63	N	CC
Standard Schnauzer	30.59	99.49	N	CA
Tibetan Terrier	21.89	99.69	N	CC
Tibetan Terrier	29.28	99.48	N	CC
West Highland White Terrier	14.60	93.12	N	CC
West Highland White Terrier	34.58	99.46	N	CC
West Highland White Terrier	31.64	99.31	N	CC
West Highland White Terrier	33.62	99.37	N	CC
West Highland White Terrier	34.41	99.33	N	CC

Wolf <sup>1</sup>	21.59	98.82	N	CC
Wolf (Chinese) <sup>1</sup>	30.85	99.41	N	CC
Wolf (Great Lakes) <sup>1</sup>	28.94	99.41	N	CC
Wolf (Iberian) <sup>1</sup>	24.26	98.45	N	CC
Wolf (Iranian) <sup>1</sup>	25.42	98.72	N	CC
Wolf (Israeli) <sup>1</sup>	23.28	97.72	N	CC
Wolf (Mexican) <sup>1</sup>	8.13	68.50	N	CC
Wolf (Portuguese) <sup>1</sup>	24.58	98.45	N	CC
Wolf (Red) <sup>1</sup>	24.98	98.69	N	CC
Wolf (Yellowstone) <sup>1</sup>	25.16	99.09	N	CC
Wolf (Yellowstone) <sup>1</sup>	23.62	99.01	N	CC
Wolf (Croatian) <sup>1</sup>	26.90	93.53	N	CC

---

\*Brachy status defined as N = normal and B = brachycephalic.

\*\*Brachy genotype here references genotype call at 5,231,894 bp on CFA32.

<sup>1</sup>These individuals are not representative of domestic dog breeds recognized by the American Kennel Club. They are wild ancestors or outgroups

covered with aligned sequence reads. This resulted in the loss of 8 individuals with total genome coverages ranging from 68.49% to 89.99%. Within the 8 removed individuals were two brachycephaly cases – a Pug and a Dogue de Bordeaux. Since there was only one sequenced Dogue de Bordeaux, this breed was eliminated from the analysis. One crossbred individual (a Kerry Blue Terrier x Beagle) was also eliminated from the dataset. The dataset was next filtered to contain only biallelic autosomal variants because the X chromosome assembly is not as reliable as the autosomal assemblies. The total number of individuals available for analysis was ultimately 93 representing 47 domestic breeds as well as a variety of wolves, a jackal and a coyote (Table 4.1). A total of 75,661,818 variants were called and used for the analysis.

#### *Runs of homozygosity analysis*

Runs of homozygosity (ROH) were determined using minor allele frequency (MAF) calculated for each variant among the 4 Pugs. A sliding window approach was used to interrogate 3,333 markers at a time. The intermarker interval is approximately 30 bp therefore every window should cover, on average, ~100,000 bp of sequence. Minor allele frequencies were averaged over all loci within each window to identify regions of reduced heterozygosity which were defined as regions with an average bin  $MAF \leq 0.02$  and that were greater than 1 Mb in length. Custom Perl scripts were used for this analysis. Within each ROH region, variants were scanned in the non-brachycephalic dogs to identify if any of these dogs were homozygous for the same allele that was found in the Pugs. Any variants that were unique to Pugs were considered to be candidates for the causal mutation underlying brachycephaly in Pugs assuming the phenotype to be determined by a single autosomal recessive mutation. Annotation of the ROH regions and

variants found in these regions within Pugs was performed using the UCSC Genome Browser [35] and an in-house database.

## **Results and Discussion**

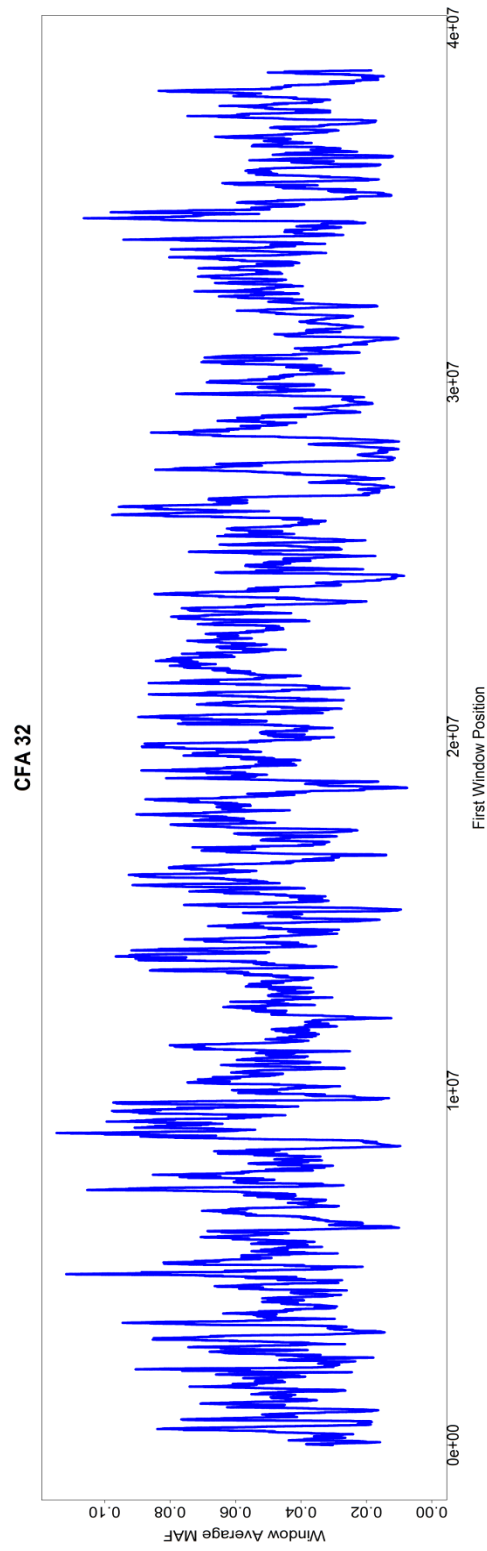
On examining a variant at 5,231,894 bp on CFA32 within *BMP3*, we found that the genotypes of dogs from some dolichocephalic, long-snouted, and mesocephalic, medium-snouted, breeds matched those of dogs from the brachycephalic breeds (Table 4.1). Following the discovery of these discordant genotypes, we re-examined the literature for instances of the non-reference, brachycephalic allele (A) in dolichocephalic and mesocephalic breeds where the reference allele (C) was expected. Interestingly, Scottish Terriers had previously been observed as being potentially fixed for the brachycephaly genotype (AA; n=3) [142, 149]. We also observed both of our sequenced Scottish Terriers to be homozygous AA. This breed is the only instance pointed out explicitly within the literature as being likely fixed for the allele causing brachycephaly, but when comparing previously reported genotypes to those in our dataset, there were several other instances in which the A allele was found in medium and long snouted breeds; some of these at high frequency. These breeds include Chinese Cresteds, Pembroke Welsh Corgis, Russell Terriers and Standard Schnauzers. Furthermore, the brachycephaly allele is not fixed within Pugs and some Pugs were either heterozygous (AC) or homozygous for the reference allele (CC). With this discordancy in Pug genotypes and such a high prevalence of the brachycephaly allele in non-brachycephalic breeds, we conclude that this mutation either does not cause brachycephaly or is epistatic to another locus such that brachycephaly is actually polygenic in dogs.

To further assess the possibility that other genomic regions are associated with brachycephaly in Pugs, whole-genome derived sequence variants were analysed to identify ROH within the autosomal genomes of the four Pugs. Overall, eight regions were identified on five chromosomes including CFA 1, 10, 18, 21, and 25 (Table 4.2). No region on CFA32 near *BMP3* showed any evidence that all four dogs were homozygous for the same haplotype (Figure 4.1), until the most relaxed filter ( $MAF \leq 0.02$  and length  $\geq 100$  kb) was used, when a region 19 Mb from *BMP3* was identified. If brachycephaly is caused by an autosomal recessive which has been selected to fixation in the development of the breed, we would expect to see a selective sweep within the genome in the region harbouring the causal variant. Since we see no evidence of a ROH in the vicinity of *BMP3* on CFA32, we can conclude that either brachycephaly is polygenic or that mutations in *BMP3* are not responsible for brachycephaly in Pugs.

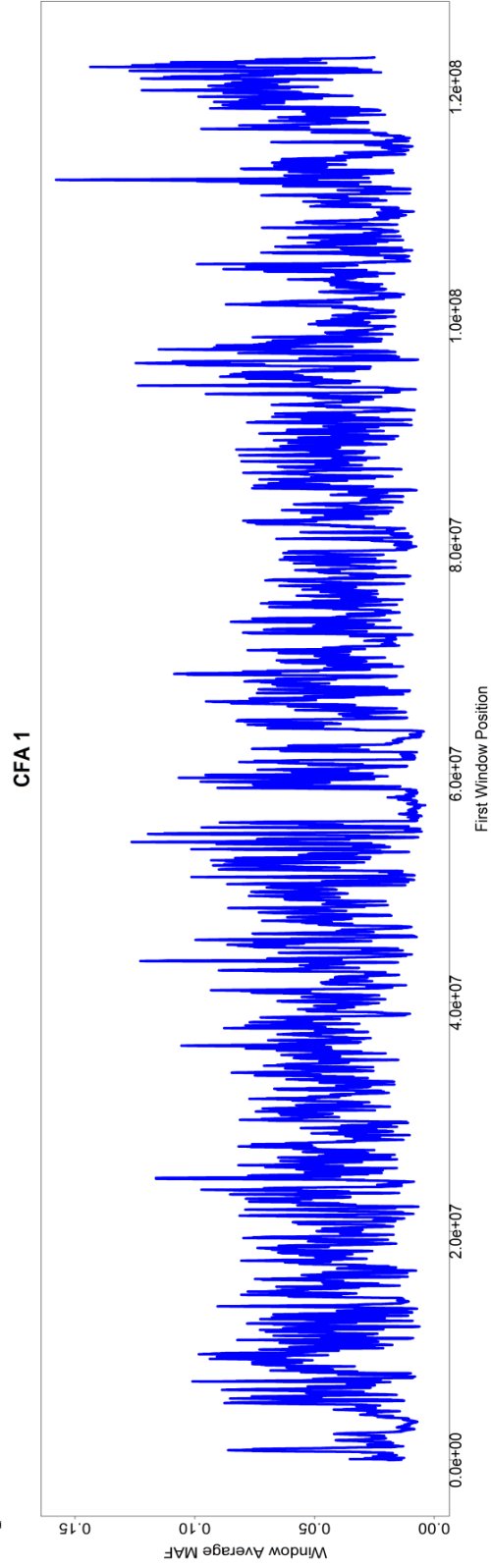
The largest of the detected ROH regions was on CFA1 at 55.8-58.6 Mb (reported as CanFam3.1) spanning 2,798,890 bp and containing 89,855 variants (Table 4.2; Figure 4.2). The region includes two genes annotated in the canine (*SMOC2* and *THBS2*) and approximately 27 orthologous regions transcribed in a variety of other species including, but not limited to, humans, mouse, rat, horse, chicken, pig, *Xenopus*, Rhesus macaque, and cattle. This region overlaps a previously reported region harbouring a brachycephaly locus detected as a 296 kb selective sweep (CFA1:59,536,208-59,832,965, CanFam2; CFA1:56,486,417-56,783,203, CanFam3.1) that was subsequently fine mapped down to a 31 kb homozygous haplotype upstream of the associated sweep region. No candidate variants for brachycephaly were predicted as being causal in that analysis [149]. Coordinates between CanFam2 and CanFam3.1 reference assemblies



**Figure 4.1 Selective sweep scan results on CFA 32 in Pugs**  
Average MAF of the sliding window on canine chromosome 32 in the pug breed. The previously published causal mutation of extreme brachycephaly is located at 5,231,894 bp. No selective sweep regions were identified as overlapping the causal mutation or across the entire chromosome.



**Figure 4.2 Selective sweep scan results on CFA 1 in Pugs**  
Windowed average minor allele frequencies across canine chromosome 1 in the Pug breed. Three selective sweep regions were identified at 2,434,072-3,720,260 bp (1.28 Mb), 55,850,267-58,649,157 bp (2.79 Mb), and 62,633,253-63,786,840 bp (1.15 Mb).



were validated using UCSC Genome Browser’s liftOver utility (Batch Coordinate Conversion) [35] to ensure that the published region was contained within the largest ROH found in this study. The genes, SPARC related modular calcium binding 2 (*SMOC2*) and thrombospondin 2 (*THBS2*), annotated and used for fine mapping in the previously published study were both contained in the identified 2.79 Mb region. This study possesses a greater power for determining candidate mutations within this region as the sequence level data reveal both a selective sweep and identifies the variation that exists within the sweep region. Candidates within this region were considered a priority and assessed for alleles that were not found in the non-brachycephalic dogs.

**Table 4.2 Detected runs of homozygosity**

<b>CFA</b>	<b>Start (bp)</b>	<b>Stop (bp)</b>	<b>Region Size (bp)</b>	<b>Number Variants</b>	<b>Lowest Binned MAF in Region</b>	<b>Highest Binned MAF in Region</b>
<b>1</b>	2,434,072	3,720,260	1,286,188	32,986	0.007063	0.019989
<b>1</b>	55,850,267	58,649,157	2,798,890	89,855	0.003800	0.019901
<b>10</b>	62,633,253	63,786,840	1,153,587	32,844	0.004437	0.019989
<b>18</b>	550,974	1,606,758	1,055,784	32,091	0.009563	0.0199394
<b>18</b>	2,861,231	4,428,927	1,567,696	51,903	0.008600	0.019901
<b>21</b>	797,423	2,114,984	1,317,561	48,518	0.008688	0.0199644
<b>25</b>	360,687	1,500,027	1,139,340	39,081	0.009400	0.0199769

The scan for alternate homozygotes was performed in the sequences of the four Pug cases and the 113 controls. The expanded number of controls came from samples that became available from active, on-going sequencing projects. This expanded sample identified 42,721 variants for interrogation. Of these, only a single positional candidate was identified in which alternate homozygotes were found in all cases *versus* all controls. This locus is located at 55,902,909 bp on CFA1. Genotypes for this region were

homozygous non-reference, GG, in all four cases and homozygous reference, CC, for all controls. This position falls within an intron of *SMOC2*. The *SMOC2* gene is involved in matrix assembly, the stimulation of endothelial cell proliferation and migration, and angiogenic activity. Online Mendelian Inheritance in Man indicates that variation in this gene is associated with orodental development [150]. Knockdown experiments in zebrafish resulted in tooth abnormalities and depletion of *SMOC2* altered expression in 3 genes that affect odontogenesis – *DLX2*, *BMP2*, and *PITX2* [151]. Interestingly, the mutation affecting odontogenesis was found to be an intronic variant causing a splice site mutation [151].

This sweep region was next assessed for the presence of small and large deletions and for regions of no sequence coverage in the Pugs as well as the presence of non-coding RNA elements annotated in other species in order to elucidate other potential candidates. No concordant variants were identified. The current sole candidate should be further assessed for concordance with phenotype by genotyping additional brachycephalic and normal dogs. We examined the genotype of the Dogue de Bordeaux that had been eliminated from the analysis for insufficient genome coverage and found that its genotype could be ascertained to be homozygous reference, CC, for this mutation. Functional analysis is also necessary to confirm that the mutation produces an altered transcript in dogs and perhaps also the altered phenotype in a laboratory species.

## **Conclusion**

We dispute the causality of a previously published mutation identified as causal for brachycephaly in extremely brachycephalic dog breeds, including Pugs. Discordant genotypes within medium to long-snouted dogs led us to pursue other analyses to identify

regions of the genome concordant with the autosomal recessive inheritance of the phenotype in Pugs. Homozygosity mapping was conducted to identify regions of homozygosity in Pugs likely to harbour the brachycephaly locus. The lack of a ROH on CFA32 led to our exploration of a large 2.79 Mb ROH region on CFA1 that overlaps a smaller region suggested prior to the publication of the *BMP3* mutation to be the major brachycephaly locus. Identifying alternate homozygotes within this larger sweep region led to the identification of a single candidate mutation at position 55,902,909 bp within *SMOC2*. Further analysis is required to provide support for this mutation as being the causal mutation for brachycephaly in Pug dogs.

### **List of abbreviations**

Abbreviations are defined in the text.

### **Competing interests**

The authors declare that they have no competing interests.

### **Author's contributions**

HRR and JFT designed the experiment. HRR, RDS, and JFT analysed the data. GSJ contributed sequence data. RDS managed the sequence database. HRR and JFT wrote the manuscript.

### **Acknowledgements**

We would like to thank the dog owners, organizations, and institutions that contributed DNA samples and sequences. We would also like to thank Zhiyuan Chen for his help debugging Perl scripts.

## References

1. Smith JM, Haigh J: **The hitch-hiking effect of a favourable gene.** *Genet Res* 1974, **23**(1):23-25.
2. Rasmuson J: **A contribution to the theory of quantitative character inheritance.** *Hereditas* 1933, **18**(1-2):245-261.
3. Thoday JM: **Location of polygenes.** *Nature* 1961, **191**(4786):368-370.
4. Zuk O, Hechter E, Sunyaev SR, Lander ES: **The mystery of missing heritability: Genetic interactions create phantom heritability.** *Proceedings of the National Academy of Sciences* 2012, **109**(4):1193-1198.
5. Ross-Ibarra J, Morrell PL, Gaut BS: **Plant domestication, a unique opportunity to identify the genetic basis of adaptation.** *PNAS* 2007, **104**:8641-8648.
6. Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P: **Evidence for two independent domestications of cattle.** *PNAS* 1994, **91**(7):2757-2761.
7. Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP, Chen K, Cooper A, Vilkki J, Seabury CM, Caetano AR *et al*: **Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics.** *PNAS* 2009, **106**(44):18644-18649.
8. Brenneman RA, Davis SK, Sanders JO, Burns BM, Wheeler TC, Turner JW, Taylor JF: **The Polled locus maps to BTA1 in a *Bos indicus* x *Bos taurus* cross.** *J Hered* 1996, **87**(2):156-161.
9. Drögemüller C, Wöhlke A, Mömke S, Distl O: **Fine mapping of the polled locus to a 1-Mb region on bovine chromosome 1q12.** *Mamm Genome* 2005, **16**(8):613-620.
10. Medugorac I, Seichter D, Graf A, Russ I, Blum H, Göpel KH, Rothhammer S, Förster M, Krebs S: **Bovine polledness – an autosomal dominant trait with allelic heterogeneity.** *PLoS One* 2012, **7**(6):e39477.
11. Klunghand H, D I Vage LG-R, S Adalsteinsson and S Lien: **The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination.** *Mamm Genome* 1995, **6**(9):636-639.
12. Grosz MD, MacNeil MD: **The "spotted" locus maps to bovine chromosome 6 in a Hereford-cross population.** *J Hered* 1999, **90**(1):233-236.
13. Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG: **Recent and ongoing selection in the human genome.** *Nat Rev Genet* 2007, **8**(11):857-868.

14. Nielsen R: **Molecular signatures of natural selection.** *Annu Rev Genet* 2005, **39**:197-218.
15. The Bovine HapMap Consortium: **Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds.** *Science* 2009, **324**(5926):528-532.
16. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C: **Genomic scans for selective sweeps using SNP data.** *Genome Res* 2005, **15**:1566-1575.
17. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES: **Positive natural selection in the human lineage.** *Science* 2006, **312**:1614-1620.
18. Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S *et al*: **Whole-genome resequencing reveals loci under selection during chicken domestication.** *Nature* 2010, **464**(7288):587-591.
19. Pollinger JP, Bustamante CD, Fledel-Alon A, Schmutz S, Gray MM, Wayne RK: **Selective sweep mapping of genes with large phenotypic effects.** *Genome Res* 2005, **15**(12):1809-1819.
20. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG *et al*: **A single IGF1 allele is a major determinant of small size in dogs.** *Science* 2007, **316**(5821):112-115.
21. Quilez J, Short AD, Martinez V, Kennedy LJ, Ollier W, Sanchez A, Altet L, Francino O: **A selective sweep of >8 Mb on chromosome 26 in the Boxer genome.** *BMC Genomics* 2011, **12**:12.
22. Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, Loredó AI, Bellone RR, Mezey JG, Brooks SA *et al*: **Four loci explain 83% of size variation in the horse.** *PLoS One* 2012, **7**(7):e39929.
23. Van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M *et al*: **A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig.** *Nature* 2003, **425**(6960):832-836.
24. Rubin C-J, Megens H-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg Ö, Jern P, Jørgensen CB *et al*: **Strong signatures of selection in the domestic pig genome.** *PNAS* 2012, **109**(48):19529-19536.
25. Andersson L: **How selective sweeps in domestic animals provide new insight into biological mechanisms.** *J Intern Med* 2012, **271**(1):1-14.

26. Barendse W, Harrison B, Bunch R, Thomas M, Turner L: **Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits.** *BMC Genomics* 2009, **10**(1):178.
27. Hayes BJ, Lien S, Nilsen H, Olsen HG, Berg P, Maceachern S, Potter S, Meuwissen THE: **The origin of selection signatures on bovine chromosome 6.** *Anim Genet* 2008, **39**(2):105-111.
28. Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H: **A genome-wide scan for signatures of recent selection in Holstein cattle.** *Animal Genetics* 2010, **41**(4):377-389.
29. Sodeland M, Kent M, Hayes B, Grove H, Lien S: **Recent and historical recombination in the admixed Norwegian Red cattle breed.** *BMC Genomics* 2011, **12**(1):33.
30. Gautier M, Flori L, Riebler A, Jaffrezic F, Laloe D, Gut I, Moazami-Goudarzi K, Foulley J-L: **A whole genome Bayesian scan for adaptive genetic divergence in West African cattle.** *BMC Genomics* 2009, **10**(1):550.
31. Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, Foulley J-L, Gautier M: **The genome response to artificial selection: A case study in dairy cattle.** *PLoS One* 2009, **4**(8):e6595.
32. Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS: **SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries.** *Nat Methods* 2008, **5**(3):247-252.
33. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TPL, Sonstegard TS *et al*: **Development and characterization of a high density SNP genotyping assay for cattle.** *PLoS One* 2009, **4**(4):e5350.
34. Rincon G, Weber KL, Van Eenennaam AL, Golden BL, Medrano JF: **Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys.** *J Dairy Sci* 2011, **94**(12):6116-6121.
35. Kent W, Sugnet C, Furey T, Roskin K, Pringle T, Zahler A, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12**(6):996-1006.
36. Dennis Jr. G, Sherman BT, Hosack DA, Yang J, Baseler MW, Lane HC, Lempicki RA: **DAVID: Database for annotation, visualization, and integrated discovery.** *Genome Biol* 2003, **4**(5):P3.
37. Decker JE, Vasco DA, McKay SD, McClure MC, Rolf MM, Kim JW, Northcutt SL, Bauck S, Woodward BW, Schnabel RD *et al*: **A novel analytical method,**



- birth date selection mapping, detects response of the Angus (*Bos taurus*) genome to artificial selection on complex traits.** *BMC Genomics*, under review.
38. The Bovine Genome Sequencing and Analysis Consortium, Elsik CG, Tellam RL, Worley KC: **The genome sequence of taurine cattle: a window to ruminant biology and evolution.** *Science* 2009, **324**(5926):522-528.
  39. Margolis RL, McInnis MG, Rosenblatt A, Ross CA: **Trinucleotide repeat expansion and neuropsychiatric disease.** *Arch Gen Psychiatry* 1999, **56**(11):1019-1031.
  40. Castermans D, Wilquet V, Parthoens E, Huysmans C, Steyaert J, Swinnen L, Fryns JP, Van de Ven W, Devriendt K: **The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism.** *J Med Genet* 2003, **40**:352-356.
  41. Castermans D, Volders K, Crepel A, Backx L, De Vos R, Freson K, Meulemans S, Vermeesch JR, Schrandt-Stumpel CT, De Rijk P, Del-Favero J, Van Geet C, Van De Ven WJ, Steyaert JG, Devriendt K, Creemers JW: **SCAMP5, NBEA and AMISYN: three candidate genes for autism involved in secretion of large dense-core vesicles.** *Hum Mol Gen* 2010, **19**(7):1368-1378.
  42. Liu Z, Chang GQ, Leibowitz SF: **Diacylglycerol kinase Z in hypothalamus interacts with long form leptin receptor. Relation to dietary fat and body weight regulation.** *J Biol Chem* 2001, **276**(8):5900-5907.
  43. Albrecht E, Gotoh T, Ebara F, Xu JX, Viergutz T, Nürnberg G, Maak S, Wegner J: **Cellular conditions for intramuscular fat deposition in Japanese Black and Holstein steers.** *Meat Sci* 2011, **89**(1):13-20.
  44. Candille SI, Kaelin CB, Cattanaach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS: **A  $\beta$ -defensin mutation causes black coat color in domestic dogs.** *Science* 2007, **318**(5855):1418-1423.
  45. Georges M, Drinkwater R, King T, Mishra A, Moore SS, Nielsen D, Sargeant LS, Sorensen A, Steele MR, Zhao X *et al*: **Microsatellite mapping of a gene affecting horn development in *Bos taurus*.** *Nat Genet* 1993, **4**:206-210.
  46. Qu Z, Adelson DL: **Bovine ncRNAs are abundant, primarily intergenic, conserved and associated with regulatory genes.** *PLoS One* 2012, **7**(8):e42638.
  47. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C *et al*: **Identification of ten loci associated with height highlights new biological pathways in human growth.** *Nat Genet* 2008, **40**(5):584-591.
  48. Karim L, Takeda H, Lin L, Druet T, Arias JAC, Baurain D, Cambisano N, Davis SR, Farnir F, Grisart B *et al*: **Variants modulating the expression of a**

- chromosome domain encompassing PLAG1 influence bovine stature.** *Nat Genet* 2011, **43**(5):405-413.
49. Dib MG, Van Vleck LD, Spangler ML: **Genetic analysis of mature size in American Angus cattle.** *Nebraska Beef Cattle Report* 2010:29-30.
50. Andersson EA, Harder MN, Pilgaard K, Pisinger C, Stančáková A, Kuusisto J, Grarup N, Færch K, Poulsen P, Witte DR *et al*: **The birth weight lowering C-allele of rs900400 near *LEKRI* and *CCNLI* associates with elevated insulin release following an oral glucose challenge.** *PLoS One* 2011, **6**(11):e27096.
51. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K, Brisbin A, Parker HG, vonHoldt BM *et al*: **A simple genetic architecture underlies morphological variation in dogs.** *PLoS Biol* 2010, **8**(8):e1000451.
52. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Pielberg GR, Sigurdsson S, Fall T, Seppälä EH, Hansen MST, Lawley CT *et al*: **Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping.** *PLoS Genet* 2011, **7**(10):e1002316.
53. Sanders JO: **History and development of zebu cattle in the United States.** *J Anim Sci* 1980, **50**:1188-1200.
54. Obbard DJ, Welch JJ, Kim K-W, Jiggins FM: **Quantifying adaptive evolution in the *Drosophila* immune system.** *PLoS Genet* 2009, **5**(10):e1000698.
55. Obbard DJ, Jiggins FM, Bradshaw NJ, Little TJ: **Recent and recurrent selective sweeps of the antiviral RNAi gene argonaute-2 in three species of *Drosophila*.** *Mol Biol Evol* 2011, **28**(2):1043-1056.
56. Tiffin P, Moeller DA: **Molecular evolution of plant immune system genes.** *Trends Genet* 2006, **22**(12):662-670.
57. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J *et al*: **Analyses of pig genomes provide insight into porcine demography and evolution.** *Nature* 2012, **491**(7424):393-398.
58. Zhang S, Xu L, Lee J, Xu T: ***Drosophila* atrophin homolog functions as a transcriptional corepressor in multiple developmental processes.** *Cell* 2002, **108**(1):45-56.
59. Zoltewicz JS, Stewart NJ, Leung R, Peterson AS: **Atrophin 2 recruits histone deacetylase and is required for the function of multiple signaling centers during mouse embryogenesis.** *Development* 2004, **131**:3-14.

60. Plaster N, Sonntag C, Schilling TF, Hammerschmidt M: **REReA/Atrophin-2 interacts with histone deacetylase and Fgf8 signaling to regulate multiple processes of zebrafish development.** *Dev Dynam* 2007, **236**(7):1891-1904.
61. Bauersachs S, Mitko K, Ulbrich SE, Blum H, Wolf E: **Transcriptome studies of bovine endometrium reveal molecular profiles characteristic for specific stages of estrous cycle and early pregnancy.** *Exp Clin Endocrinol Diabetes* 2008, **116**(7):371-384.
62. Kageyama R, Ohtsuka T, Kobayashi T: **The Hes gene family: repressors and oscillators that orchestrate embryogenesis.** *Development* 2007, **134**(7):1243-1251.
63. Hansen TR, Austin KJ, Johnson GA: **Transient ubiquitin cross-reactive protein gene expression in the bovine endometrium.** *Endocrinology* 1997, **138**(11):5079-5082.
64. Bebington C, Bell SC, Doherty FJ, Fazleabas AT, Fleming SD: **Localization of ubiquitin and ubiquitin cross-reactive protein in human and baboon endometrium and decidua during the menstrual cycle and early pregnancy.** *Biol Reprod* 1999, **60**(4):920-928.
65. Popova NK, Voitenko NN, Kulikov AV, Avgustinovich DF: **Evidence for the involvement of central serotonin in mechanism of domestication of silver foxes.** *Pharmacol Biochem Be* 1991, **40**(4):751-756.
66. Popova NK KA, Avgustinovich DF, Voitenko NN, Trut LN: **Effect of domestication of the silver fox on the main enzymes of serotonin metabolism and serotonin receptors.** *Genetika* 1997, **33**:708-716.
67. Kukekova A, Temnykh S, Johnson J, Trut L, Acland G: **Genetics of behavior in the silver fox.** *Mamm Genome* 2012, **23**(1-2):164-177.
68. Li Y, vonHoldt BM, Reynolds A, Boyko AR, Wayne RK, Wu D-D, Zhang Y-P: **Artificial selection on brain expressed genes during the domestication of dog.** *Mol Bio Evol* 2013.
69. Birling M-C, Tait S, Hardy RJ, Brophy PJ: **A novel rat tetraspan protein in cells of the oligodendrocyte lineage.** *J Neurochem* 1999, **73**(6):2600-2608.
70. Terada N, BaracsKay K, Kinter M, Melrose S, Brophy PJ, Boucheix C, Bjartmar C, Kidd G, Trapp BD: **The tetraspanin protein, CD9, is expressed by progenitor cells committed to oligodendrogenesis and is linked to  $\beta$ 1 integrin, CD81, and Tspan-2.** *Glia* 2002, **40**(3):350-359.
71. Campbell EMG, Nonneman D, Rohrer GA: **Fine mapping a quantitative trait locus affecting ovulation rate in swine on chromosome 8.** *J Anim Sci* 2003, **81**(7):1706-1714.

72. Sudhof TC: **Neuroligins and neurexins link synaptic function to cognitive disease.** *Nature* 2008, **455**(7215):903-911.
73. Kim H-G, Kishikawa S, Higgins AW, Seong I-S, Donovan DJ, Shen Y, Lally E, Weiss LA, Najm J, Kutsche K *et al*: **Disruption of neurexin 1 associated with autism spectrum disorder.** *Am J Hum Genet* 2008, **82**(1):199-207.
74. Rice GI, Bond J, Asipu A, Brunette RL, Manfield IW, Carr IM, Fuller JC, Jackson RM, Lamb T, Briggs TA *et al*: **Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response.** *Nat Genet* 2009, **41**(7):829-832.
75. Lewis RS, Kolesnik TB, Kuang Z, D'Cruz AA, Blewitt ME, Masters SL, Low A, Willson T, Norton RS, Nicholson SE: **TLR regulation of SPSB1 controls inducible nitric oxide synthase induction.** *J Immunol* 2011, **187**(7):3798-3805.
76. Ghasemvand F, Salehi Z, Fakour Y, Omidinia E: **Association of 757 C/T polymorphism in PRODH gene with Schizophrenia in Iranian population.** *J Paramed Sci* 2013, **4**(1).
77. Marg A, Haase H, Neumann T, Kouno M, Morano I: **AHNAK1 and AHNAK2 are costameric proteins: AHNAK1 affects transverse skeletal muscle fiber stiffness.** *Biochem Biophys Res Commun* 2010, **401**(1):143-148.
78. Fernandes PR, Samuelson DA, Clark WR, Cousins RJ: **Immunohistochemical localization of cysteine-rich intestinal protein in rat small intestine.** *Am J Physiol Gastrointest Liver Physiol* 1997, **272**(4):G751-G759.
79. Schneider B, Hanke P, Jagla W, Wattler S, Nehls M, Grosse J, Schroder A, Laufs J: **Synergistic interaction of two independent genetic loci causes extreme elevation of serum IgA in mice.** *Genes Immun* 2004, **5**(5):375-380.
80. Chen DT, Jiang X, Akula N, Shugart YY, Wendland JR, Steele CJM, Kassem L, Park JH, Chatterjee N, Jamain S *et al*: **Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder.** *Mol Psychiatry* 2011.
81. Burgess HA, Martinez S, Reiner O: **KIAA0369, doublecortin-like kinase, is expressed during brain development.** *J Neurosci Res* 1999, **58**(4):567-575.
82. Short B, Haas A, Barr FA: **Golgins and GTPases, giving identity and structure to the Golgi apparatus.** *Biochim Biophys Acta* 2005, **1744**(3):383-395.
83. Vjugina U, Zhu X, Oh E, Bracero NJ, Evans JP: **Reduction of mouse egg surface integrin alpha9 subunit (ITGA9) reduces the egg's ability to support sperm-egg binding and fusion.** *Biol Reprod* 2009, **80**(4):833-841.

84. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MAR, Green T *et al*: **Association between microdeletion and microduplication at 16p11.2 and Autism.** *N Engl J Med* 2008, **358**(7):667-675.
85. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y *et al*: **Structural variation of chromosomes in autism spectrum disorder.** *Am J Hum Gen* 2008, **82**(2):477-488.
86. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, Gilliam TC, Nowak NJ, Cook EH, Dobyns WB *et al*: **Recurrent 16p11.2 microdeletions in autism.** *Hum Mol Genet* 2008, **17**(4):628-638.
87. Strotmann R, Schröck K, Bösel I, Stäubert C, Russ A, Schöneberg T: **Evolution of GPCR: Change and continuity.** *Mol Cell Endocrinol* 2011, **331**(2):170-178.
88. Cheutin T, McNairn AJ, Jenuwein T, Gilbert DM, Singh PB, Misteli T: **Maintenance of stable heterochromatin domains by dynamic HP1 binding.** *Science* 2003, **299**(5607):721-725.
89. Martin C, Beaujean N, Brochard V, Audouard C, Zink D, Debey P: **Genome restructuring in mouse embryos during reprogramming and early development.** *Dev Biol* 2006, **292**(2):317-332.
90. Ruddock-D'Cruz NT, Prashadkumar S, Wilson KJ, Heffernan C, Cooney MA, French AJ, Jans DA, Verma PJ, Holland MK: **Dynamic changes in localization of chromobox (CBX) family members during the maternal to embryonic transition.** *Mol Reprod Dev* 2008, **75**(3):477-488.
91. Arch JRS, Newsholme EA: **Activities and some properties of 5'-nucleotidase, adenosine kinase and adenosine deaminase in tissues from vertebrates and invertebrates in relation to the control of the concentration and the physiological role of adenosine.** *Biochem J* 1978, **174**:965-977.
92. Nei M, Niimura Y, Nozawa M: **The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity.** *Nat Rev Genet* 2008, **9**(12):951-963.
93. Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S: **Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates.** *PLoS Biol* 2004, **2**(1):e5.
94. Zhang X, De la Cruz O, Pinto JM, Nicolae D, Firestein S, Gilad Y: **Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray.** *Genome Biol* 2007, **8**:R86.

95. Glick G, Shirak A, Uliel S, Zeron Y, Ezra E, Seroussi E, Ron M, Weller JI: **Signatures of contemporary selection in the Israeli Holstein dairy cattle.** *Animal Genetics* 2012, **43**:45-55.
96. Mason IL: **A World Dictionary of Livestock Breeds, Types and Varieties**, 4 edn. Oxford, UK: CAB International, Wallingford; 1996.
97. Schwarzenbacher H, Dolezal M, Flisikowski K, Seefried F, Wurmser C, Schlotterer C, Fries R: **Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle.** *BMC Genomics* 2012, **13**(1):48.
98. Purfield DC, Berry DP, McParland S, Bradley DG: **Runs of homozygosity and population history in cattle.** *BMC Genetics* 2012, **13**:70.
99. Hu Z-L, Park CA, Wu X-L, Reecy JM: **Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era.** *Nucleic Acids Research* 2013, **41**(Database issue):D871-D879.
100. Dillard EU, Rodriguez O, Robison OW: **Estimation of additive and nonadditive direct and maternal genetic effects from crossbreeding beef cattle.** *Journal of Animal Science* 1980, **50**(4):653-663.
101. Große-Brinkhaus C, Jonas E, Buschbell H, Phatsara C, Tesfaye D, Jüngst H, Looft C, Schellander K, Tholen E: **Epistatic QTL pairs associated with meat quality and carcass composition traits in a porcine Duroc× Pietrain population.** *Genet Sel Evol* 2010, **42**(1):39.
102. Tortereau F, Sanchez M-P, Feve K, Gilbert H, Iannuccelli N, Billon Y, Milan D, Bidanel J-P, Riquet J: **Progeny-testing of full-sibs IBD in a SSC2 QTL region highlights epistatic interactions for fatness traits in pigs.** *BMC Genetics* 2011, **12**(1):92.
103. Stratz P, Baes C, Rückert C, Preuss S, Bennewitz J: **A two-step approach to map quantitative trait loci for meat quality in connected porcine F2 crosses considering main and epistatic effects.** *Animal Genetics* 2013, **44**(1):14-23.
104. García-Gómez E, Reverter A, Whan V, McWilliam SM, Arranz JJ, Kijas J, International Sheep Genomics C: **Using regulatory and epistatic networks to extend the findings of a genome scan: Identifying the gene drivers of pigmentation in Merino sheep.** *PLoS ONE* 2011, **6**(6):e21158.
105. Raadsma HW, Jonas E, Fleet MR, Fullard K, Gongora J, Cavanagh CR, Tammen I, Thomson PC: **QTL and association analysis for skin and fibre pigmentation in sheep provides evidence of a major causative mutation and epistatic effects.** *Animal Genetics* 2013, **44**(5):547-559.

106. Carvalho TDd, Siqueira F, Torres Júnior RAdA, Medeiros SRd, Feijó GLD, Junior S, Blecha IMZ, Soares CO: **Association of polymorphisms in the leptin and thyroglobulin genes with meat quality and carcass traits in beef cattle.** *Revista Brasileira de Zootecnia* 2012, **41**(10):2162-2168.
107. Bennett G, Shackelford S, Wheeler T, King D, Casas E, Smith T: **Selection for genetic markers in beef cattle reveals complex associations of thyroglobulin and casein1-s1 with carcass and meat traits.** *Journal of Animal Science* 2013, **91**(2):565-571.
108. Ma L, Runesha HB, Dvorkin D, Garbe J, Da Y: **Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies.** *BMC Bioinformatics* 2008, **9**(1):315.
109. Hu T, Chen Y, Kiralis JW, Moore JH: **ViSEN: Methodology and Software for Visualization of Statistical Epistasis Networks.** *Genetic Epidemiology* 2013, **37**(3):283-285.
110. Lishout FV, Mahachie John J, Gusareva E, Urrea V, Cleynen I, Theatre E, Charloteaux B, Calle ML, Wehenkel L, Steen KV: **An efficient algorithm to perform multiple testing in epistasis screening.** *BMC Bioinformatics* 2013, **14**(1):138.
111. Kogelman L, Kadarmideen H: **Weighted Interaction SNP Hub (WISH) network method for building genetic networks for complex diseases and traits using whole genome genotype data.** *BMC Syst Biol* 2014, **8**(2):1-12.
112. Neibergs H, Seabury C, Wojtowicz A, Wang Z, Scraggs E, Kiser J, Neupane M, Womack J, Eenennaam A, Hagevoort G *et al*: **Susceptibility loci revealed for bovine respiratory disease complex in pre-weaned holstein calves.** *BMC Genomics* 2014, **15**(1):1164.
113. Saatchi M, Beever J, Decker J, Faulkner D, Freetly H, Hansen S, Yampara-Iquise H, Johnson K, Kachman S, Kerley M *et al*: **QTLs associated with dry matter intake, metabolic mid-test weight, growth and feed efficiency have little overlap across 4 beef cattle studies.** *BMC Genomics* 2014, **15**(1):1004.
114. Saatchi M, Schnabel R, Taylor J, Garrick D: **Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds.** *BMC Genomics* 2014, **15**(1):442.
115. Beavis WD: **The power and deceit of QTL experiments: lessons from comparative QTL studies.** In: *Proceedings of the forty-ninth annual corn and sorghum industry research conference: 1994; Washington, DC: Chicago, IL; 1994: 250-266.*
116. McClure MC, Morsci NS, Schnabel RD, Kim JW, Yao P, Rolf MM, McKay SD, Gregg SJ, Chapple RH, Northcutt SL *et al*: **A genome scan for quantitative trait**

- loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Animal Genetics* 2010, **41**(6):597-607.**
117. Nandurkar HH, Huysmans R: **The myotubularin family: Novel phosphoinositide regulators.** *IUBMB Life* 2002, **53**(1):37-43.
118. Chandler RJ, Aswani V, Tsai MS, Falk M, Wehrli N, Stabler S, Allen R, Sedensky M, Kazazian HH, Venditti CP: **Propionyl-CoA and adenosylcobalamin metabolism in *Caenorhabditis elegans*: Evidence for a role of methylmalonyl-CoA epimerase in intermediary metabolism.** *Molecular Genetics and Metabolism* 2006, **89**(1-2):64-73.
119. Yang F, Wang Q, Wang M, He K, Pan Y: **Associations between gene polymorphisms in two crucial metabolic pathways and growth traits in pigs.** *Chin Sci Bull* 2012, **57**(21):2733-2740.
120. Ockerman HW, Jaworek D, VanStavern B, Parrett N, Pierson CJ: **Castration and sire effects on carcass traits, meat palatability and muscle fiber characteristics in Angus cattle.** *J Anim Sci* 1984, **59**(4):981-990.
121. Gregory KE, Ford JJ: **Effects of late castration, Zeranol and breed group on growth, feed efficiency and carcass characteristics of late maturing bovine males.** *J Anim Sci* 1983, **56**(4):771-780.
122. Kieslinger M, Folberth S, Dobreva G, Dorn T, Croci L, Erben R, Consalez GG, Grosschedl R: **EBF2 regulates osteoblast-dependent differentiation of osteoclasts.** *Developmental Cell* 2005, **9**(6):757-767.
123. Gimble JM, Robinson CE, Wu X, Kelly KA: **The function of adipocytes in the bone marrow stroma: an update.** *Bone* 1996, **19**(5):421-428.
124. Pittenger MF, Mackay AM, Beck SC, Jaiswal RK, Douglas R, Mosca JD, Moorman MA, Simonetti DW, Craig S, Marshak DR: **Multilineage potential of adult human mesenchymal stem cells.** *Science* 1999, **284**(5411):143-147.
125. Yokota T, Meka CSR, Medina KL, Igarashi H, Comp PC, Takahashi M, Nishida M, Oritani K, Miyagawa J-i, Funahashi T *et al*: **Paracrine regulation of fat cell formation in bone marrow cultures via adiponectin and prostaglandins.** *The Journal of Clinical Investigation* 2002, **109**(10):1303-1310.
126. Edwards DB, Ernst CW, Raney NE, Doumit ME, Hoge MD, Bates RO: **Quantitative trait locus mapping in an F2 Duroc × Pietrain resource population: II. Carcass and meat quality traits1.** *J Anim Sci* 2008, **86**(2):254-266.
127. Thaller G, Kühn C, Winter A, Ewald G, Bellmann O, Wegner J, Zühlke H, Fries R: **DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle.** *Animal Genetics* 2003, **34**(5):354-357.



128. Yuan Z, Li J, Li J, Gao X, Gao H, Xu S: **Effects of DGAT1 gene on meat and carcass fatness quality in Chinese commercial cattle.** *Mol Biol Rep* 2013, **40**(2):1947-1954.
129. Sonstegard T, Ma L, Van Tassell C, Kim E, Cole J, Wiggans G, Crooker B, Mariani B, Matukumalli L, Garbe J: **Forty years of artificial selection in US Holstein cattle had genome-wide signatures.** *Poster presentation at 9th World Congr Genet Appl Livest Prod, Leipzig, Germany* 2010.
130. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P: **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 research* 2002, **12**(2):222-231.
131. Ashwell MS, Heyen DW, Sonstegard TS, Van Tassell CP, Da Y, VanRaden PM, Ron M, Weller JI, Lewin HA: **Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle.** *Journal of Dairy Science* 2004, **87**(2):468-475.
132. Tantia M, Vijn R, Mishra B, Mishra B, Kumar SB, Sodhi M: **DGAT1 and ABCG2 polymorphism in Indian cattle (*Bos indicus*) and buffalo (*Bubalus bubalis*) breeds.** *BMC Veterinary Research* 2006, **2**(1):32.
133. Kaupé B, Winter A, Fries R, Erhardt G: **DGAT1 polymorphism in *Bos indicus* and *Bos taurus* cattle breeds.** *Journal of Dairy Research* 2004, **71**(02):182-187.
134. Cantley LC: **The phosphoinositide 3-kinase pathway.** *Science* 2002, **296**(5573):1655-1657.
135. Glass DJ: **Skeletal muscle hypertrophy and atrophy signaling pathways.** *The International Journal of Biochemistry & Cell Biology* 2005, **37**(10):1974-1984.
136. Latres E, Amini AR, Amini AA, Griffiths J, Martin FJ, Wei Y, Lin HC, Yancopoulos GD, Glass DJ: **Insulin-like growth factor-1 (IGF-1) inversely regulates atrophy-induced genes via the phosphatidylinositol 3-kinase/Akt/mammalian target of rapamycin (PI3K/Akt/mTOR) pathway.** *Journal of Biological Chemistry* 2005, **280**(4):2737-2744.
137. Léger B, Cartoni R, Praz M, Lamon S, Dériaz O, Crettenand A, Gobelet C, Rohmer P, Konzelmann M, Luthi F *et al*: **Akt signalling through GSK-3 $\beta$ , mTOR and Foxo1 is involved in human skeletal muscle hypertrophy and atrophy.** *The Journal of Physiology* 2006, **576**(3):923-933.
138. Raven L-A, Cocks B, Hayes B: **Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle.** *BMC Genomics* 2014, **15**(1):62.

139. Farias FHG, Zeng R, Johnson GS, Winger FA, Taylor JF, Schnabel RD, McKay SD, Sanders DN, Lohi H, Seppälä EH *et al*: **A truncating mutation in ATP13A2 is responsible for adult-onset neuronal ceroid lipofuscinosis in Tibetan terriers.** *Neurobiol Dis* 2011, **42**(3):468-474.
140. Beltran WA, Cideciyan AV, Lewin AS, Iwabe S, Khanna H, Sumaroka A, Chiodo VA, Fajardo DS, Román AJ, Deng W-T *et al*: **Gene therapy rescues photoreceptor blindness in dogs and paves the way for treating human X-linked retinitis pigmentosa.** *PNAS* 2012, **109**(6):2132-2137.
141. Ahonen SJ, Pietilä E, Mellersh CS, Tiira K, Hansen L, Johnson GS, Lohi H: **Genome-wide association study identifies a novel canine glaucoma locus.** *PLoS ONE* 2013, **8**(8):e70903.
142. Schoenebeck JJ, Ostrander EA: **The genetics of canine skull shape variation.** *Genetics* 2013, **193**(2):317-325.
143. You N, Murillo G, Su X, Zeng X, Xu J, Ning K, Zhang S, Zhu J, Cui X: **SNP calling using genotype model selection on high-throughput sequencing data.** *Bioinformatics* 2012, **28**(5):643-650.
144. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al*: **A framework for variation discovery and genotyping using next-generation DNA sequencing data.** *Nat Genet* 2011, **43**(5):491-498.
145. Amis TC, Kurpershoek C: **Pattern of breathing in brachycephalic dogs.** *Am J Vet Res* 1986, **47**(10):2200-2204.
146. Hendricks JC: **Brachycephalic airway syndrome.** *Veterinary Clinics of North America: Small Animal Practice* 1992, **22**(5):1145-1153.
147. Roedler FS, Pohl S, Oechtering GU: **How does severe brachycephaly affect dog's lives? Results of a structured preoperative owner questionnaire.** *The Veterinary Journal* 2013, **198**(3):606-610.
148. Bannasch D, Young A, Myers J, Truvé K, Dickinson P, Gregg J, Davis R, Bongcam-Rudloff E, Webster MT, Lindblad-Toh K *et al*: **Localization of canine brachycephaly using an across breed mapping approach.** *PLoS ONE* 2010, **5**(3):e9632.
149. Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, Faden DL, Rimbault M, Decker B, Kidd JM, Sood R *et al*: **Variation of BMP3 contributes to dog breed skull diversity.** *PLoS Genet* 2012, **8**(8):e1002849.
150. **Online Mendelian Inheritance in Man, OMIM®** [<http://omim.org/>]

151. Bloch-Zupan A, Jamet X, Etard C, Laugel V, Muller J, Geoffroy V, Strauss J-P, Pelletier V, Marion V, Poch O *et al*: **Homozygosity mapping and candidate prioritization identify mutations, missed by whole-exome sequencing, in SMOC2, causing major dental developmental defects.** *The American Journal of Human Genetics* 2011, **89**(6):773-781.

## VITA

Holly René Ramey was born in 1988 in Greensboro, North Carolina. Shortly after birth she moved following her father, David's, job to Acworth, Georgia, a small Atlanta suburb. There she grew up in a Southern home with her parents, David and Cindy, and her older sister Peggy.

Following her parents' divorce in 2001 she moved back to North Carolina the summer after her sixth grade year. There she began to understand and foster an appreciation and knowledge of the rural lifestyle. Once she entered Eastern Randolph High School in 2002, she began taking agriculture classes as electives out of curiosity. The introductory course led her to begin enrolling in Animal Science courses within the high school's vocational department. Shortly after finding a passion for agriculture she got involved in her FFA Chapter. Starting as the Historian on the local officer team she began developing her leadership and teamwork skills. Not only was she active in the leadership but also in Career Development Events such as tool identification, poultry judging, livestock judging and dairy judging. She competed on the state level in these events and was successful enough to attend the national competition in dairy judging. She also participated and competed extracurricularly in the ERHS Marching Wildcat Band for four years as a clarinet player and clarinet section leader during her senior year. Beyond her love of agriculture classes, youth group at Ramseur Wesleyan Church, and marching band she was involved with advance placement and honours courses in almost every topic offered. She attained a very high level of academic success in her rural school. The academic achievements and extracurricular activities helped her to apply for a Bachelor's degree in Animal Sciences at North Carolina State University.

Beginning her studies at NC State in the fall of 2006, she quickly became involved on campus attending a variety of Wolfpack sporting events, becoming a sister in Sigma Alpha and a College of Agriculture and Life Sciences Ambassador. In her academic life she became enamoured with the genetics section of all the Animal Science courses she took. This led her to pursue a Minor in Genetics alongside her Bachelor's. The academic success and interest in research she had resulted in an undergraduate research position in Dr. Joe Cassady's quantitative genetics lab. During this time she collected field data and scored films to collect data on suckling piglets. Dr. Cassady approached her about a summer research internship opportunity that he thought would be beneficial at the University of Missouri for the summer of 2009. Upon applying for this position she landed in the Taylor lab at MU. This began the path of her graduate journey as she was invited back at the end of the summer to do a Ph.D. with Dr. Taylor. Holly graduated from NC State in 2010.

She joined Dr. Jerry Taylor's lab at MU in July of 2010. Foregoing her Master's degree she launched directly into a doctoral program. Here she has learned how to become a true scientist and worked on many different genomic projects. This includes the completion of her Comprehensives Exam in March of 2014. The work described in this dissertation catalogues her work and successes throughout her career at MU. Holly successfully defended for her Ph.D. on April 29, 2015. Her graduation will be May 16, 2015.