# DECIPHERING THE GENETIC BASIS FOR COMPLEX TRAIT VARIATION: UTILIZING ALTERNATIVE GENOME-WIDE ASSOCIATION METRICS AND MOLECULAR PHENOTYPES

By

Scott A Funkhouser

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Genetics – Doctor of Philosophy

2019

## ABSTRACT

DECIPHERING THE GENETIC BASIS FOR COMPLEX TRAIT VARIATION: UTILIZING ALTERNATIVE GENOME-WIDE ASSOCIATION METRICS AND MOLECULAR PHENOTYPES

By

Scott A Funkhouser

Within any population, complex trait variation can be attributed to an impressive number of genetic factors. Identification of such factors has been made possible, in part, by large biomedical datasets comprised of genotypes and phenotypes for hundreds of thousands of individuals. Furthermore, understanding the biological mechanisms through which genetic variation creates complex trait variation has been facilitated by high-throughput sequencing technology, used to quantify molecular, intermediate phenotypes. Despite such datasets being widely available, we lack understanding of the full spectrum of genetic effects, including gene-by-sex (G×S) interactions. We also have yet to uncover various molecular phenotypes that may "link" genetic variation to complex trait variation. To address these gaps in knowledge, the following chapters will 1) develop and utilize statistical methodology for mapping G×S interactions among human traits, and 2) utilize a pig model to characterize RNA editing—a relatively understudied form of transcriptional regulation—and evaluate its potential to link genetic variation with complex trait variation.

Growing evidence from genome-wide parameter estimates suggest males and females from human populations possess differing genetic architectures. Despite this, mapping G×S interactions remains challenging, suggesting that the magnitude of a typical G×S interaction is exceedingly small. We have developed a local Bayesian regression (LBR) approach to estimate sex-specific single nucleotide polymorphism (SNP) marker effects after fully accounting for local linkage-disequilibrium (LD) patterns. This provided means to infer G×S interactions either at the SNP level, or by aggregating multiple sex-specific SNP effects to make inferences at the level of small, LD-based regions. In simulations, LBR provided greater power and resolution to detect G×S interactions than the traditional approach to genome-wide association (GWA), single-marker regression (SMR).

When using LBR to analyze human traits from the UK Biobank (N ~ 250,000) including height, BMI, bone-mineral density, and waist-to-hip ratio, we find evidence of novel G×S interactions where sex-specific effects explain a very small proportion of phenotypic variance ($R^2 < 1x10^{-4}$) but are enriched in expression quantitative trait loci (eQTL). By leveraging large datasets and powerful metrics, we are providing evidence that G×S interactions may influence phenotypic variance for a variety of human complex traits.

Adenosine to inosine (A-to-I) RNA editing impacts gene function by converting adenosine to inosine molecules within specific regions of the transcriptome and is catalyzed by adenosine deaminase acting on RNA (ADAR). High-throughput sequencing studies, most of which utilizing human models, have found thousands of A-to-I edited loci commonly located within repetitive elements such as the primate-specific Alu element. Here, we utilized matched whole-genome sequencing and RNA sequencing from the same animal to demonstrate that widespread RNA editing occurs within pig transcriptomes, largely within pig-specific repetitive elements known as PRE-1.

The degree that sites in the transcriptome are edited by ADAR—the "editing level"—has been observed to vary within populations but it is largely unknown how genetic variation as whole influences editing level variation. Using 168 $F_2$ pigs with SNP genotyping data and RNA sequencing from skeletal muscle, we identified five RNA editing sites across four genes whose editing level variation was significantly attributed to the additive effects of all observed SNP markers (estimated genomic heritability $\hat{h}^2_g$ = 0.31–0.56; $p$-value = $8.2x10^{-5}$–$8.8x10^{-4}$). We then used bivariate models to estimate how genetics influences covariance between site-specific RNA editing levels and complex traits in pigs. We found modest evidence that SNPs near *ADAR* contribute to covariance in RNA editing activity and numerous growth traits such as average daily gain (local genetic correlation $\hat{\rho}_{g_{local}}$ [SE] = -0.87 [0.16]; $p$-value = 0.029). These results suggest potential pleiotropic effects between RNA editing activity and complex traits and encourages further use of multi-variate mixed models determine if RNA editing can "link" genetic variation with complex trait variation.

This work is dedicated to Kelsey Funkhouser and our cat, Duchess

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Cathy Ernst for giving me a chance to succeed as a graduate student. During my first year, I struggled to identify the area of genetics that both excited me and gave me confidence as a scientist. Cathy provided me an opportunity to work with the National Swine Registry (thanks to Doug Newcom and NSR as well) to develop methods to estimate breed composition from genetic data (I wouldn't have succeeded in that project without vital help from Juan Steibel and Ron Bates). This work exposed me to the greater world of statistics and quantitative genetics; today, these are two subjects that I *love* and feel compelled to continually acquire greater skills in. I am very thankful for a supportive committee who have allowed me to grow as a young quantitative geneticist (truly, I am thankful because prior to coming to MSU, I had never taken a statistics, probability, or linear algebra course).

I also cannot thank Cathy enough for her guidance. Without Cathy's help I could not have obtained a USDA NIFA pre-doctoral fellowship, which helped me obtain confidence as a scientist and a taste of successful grant writing experience.

I would like to thank Gustavo de los Campos who taught me many fundamentals of quantitative genetics, linear models, Bayesian estimation techniques, (the list goes on). In working with Gustavo, I began appreciating many open questions in the field of statistical and quantitative genetics. Together, we have spent a considerable amount of time brainstorming how we can improve upon existing GWAS methods. These experiences have been incredibly rewarding and will continue to inspire me well into my future scientific career.

Lastly, I have to thank my close friends, who have helped these six years become the most formative years of my life thus far. They include (but are not limited to) Ryan Corbett, Amanda Koenig, Agu González Reymúndez, and Kaitlyn Daza. Together, we've leaned on each other during the hardest times of graduate school. I also have to thank my best friend and partner, Kelsey Funkhouser, for being the strongest and most inspiring scientist I know.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 Genetic factors that influence variation in complex traits

For more than a decade, genome-wide association studies (GWAS) have identified single-nucleotide polymorphisms (SNPs) that associate with complex traits and diseases [1, 2]. Using human height as a model complex trait, it became clear that the proportion of phenotypic variance explained by GWAS-identified loci was much lower than the narrow-sense heritability [3, 4], where narrow-sense heritability (or simply, heritability) is the proportion of phenotypic variance explained by the additive effects at all quantitative trait loci (QTL). This "missing heritability" problem was partially solved; by estimating the proportion of variance explained by all SNPs genome-wide (estimating so-called "genomic" or "SNP" heritability), it became evident that many loci possess exceedingly small additive effects that go undetected by GWAS mainly due to the burden of multiple test correction [5]. It is generally accepted that any "still missing heritability" can be attributed to imperfect linkage-disequilibrium (LD) between observed SNPs and underlying QTL [6].

Although the nature of missing heritability is largely understood (in the sense that we understand the large disparity between GWAS-identified loci and trait heritability), work is ongoing to map the locations of both large- and small-effect QTL for a wide variety of complex, polygenic traits [7, 8, 9, 10]. This has been made possible, in part, by large biomedical datasets that contain genotypic and phenotypic records for hundreds of thousands of individuals (examples include the UK Biobank, China Kadoorie Biobank, FinnGen, 23andMe®, etc). When using ~700,000 individuals of European ancestry, GWAS-identified loci at a stringent $p$-value threshold of $1x10^{-8}$ are able to explain a much larger proportion of variance in height and BMI than before [11]. This suggests locations of small additive effects can be revealed with increasing sample size.

Still, many additional genetic factors contribute to the variance of complex traits (by contributing to broad-sense heritability). These factors cannot be discovered simply by increasing sample size

but rely more on appropriate modeling and experimental design. They include a multitude of complex interactions such as dominance effects (interactions between alleles at the same locus), epistasis (interactions between alleles at different loci), and various gene-by-environment (G×E) interactions. Conceptually, one may interpret G×E to be a locus with different effects depending on the environment it's placed in. Common approaches to GWAS or genomic heritability estimation assume that additive effects at SNPs are homogenous (constant for every member in the population). However, members within any population may undergo different environmental exposures (either endogenous or exogenous) creating the possibility of heterogenous genetic effects, should G×E exist.

Gene-by-sex (G×S) interactions are a form of G×E; sex directly influences both the endogenous (for instance, sex hormones influence transcriptional mechanisms) and exogenous (for instance, contraceptive use) environment. The existence of G×S interactions is one of several theories used to explain sex differences for numerous complex traits (for quantitative traits such as height, these include differences in both mean and variance). Evidence for G×S interactions largely stems from "between-sex" or "cross-sex" genetic correlation estimates [12, 13, 14]. When evidently less than one, these estimates indicate that genetic effects are disproportional between sexes [15]. Even with relatively large sample sizes (> 100,000), mapping G×S interactions remains challenging [16, 17], suggesting that for many traits the magnitude of any G×S is small and is likely escaping GWAS detection due to the burden of multiple test correction. Just as numerous small, homogenous, additive effects accumulate to influence narrow-sense heritability, numerous small G×S interactions may influence broad-sense heritability by inducing mean and variance differences between sexes.

In chapter 2, individuals of European ancestry from the UK Biobank (N $\sim$ 250,000) are used to map G×S interactions. In this chapter we discuss the difficulty of mapping G×S interactions using traditional GWAS methods and develop an alternative strategy for mapping such events. We find evidence of small-magnitude G×S interactions impacting such traits as height, BMI, and bone-mineral density.

## 1.2 Limitations to common GWAS techniques and alternative solutions

Despite sample sizes continuing to grow in recent years, the methods used for GWAS have remained largely the same since the first reported GWAS in 2006 [18]. These methods, commonly referred to as "single marker analysis" or "single marker regression" (SMR), test for an association between a complex trait and a SNP, doing so one SNP at a time. To control the family-wise error rate (probability of making at least one type I error), a *p*-value threshold of $5x10^{-8}$ is routinely adopted.

Although this straightforward approach is useful, it has a few drawbacks. As mentioned previously, the burden of multiple testing can severely hinder statistical power. Additionally, patterns of LD are not fully accounted for; as sample size continuous to grow, mapping resolution will worsen for moderate- to large-effect QTL. This is because many SNPs can be associated with one or more QTL due to LD, creating spurious signals that amplify as sample size increases.

Numerous methods have been developed that treat GWAS like a variable selection problem, one in which multiple covariates (SNPs, in this case) are considered simultaneously and only the most relevant ones are selected as non-null, useful predictors. Such methods have shown improved mapping resolution when compared to SMR [19, 20, 21, 22]. Under this variable selection paradigm, one can estimate the additive effects of multiple SNPs simultaneously using Bayesian multiple regression mixture models [23, 24]. Bayesian mixture models have several nice properties: i) the proportion of SNPs with non-null genetic effects (the polygenicity) can be treated as random and inferred from the data and ii) inferring whether each SNP has a non-null effect can be done formally by estimating the corresponding posterior probability (the probability of a non-null SNP effect, given the data). The first point is crucial for achieving appropriate error control [25]. One important criticism to Bayesian mixture models for GWAS (that applies to multiple regression models in general) is that in regions of high LD, the association of any given SNP with a trait may be exceedingly small when conditioning on all other SNPs in the region [26]. Fernando et al. demonstrated that in instances when individual SNP-trait associations are small due to LD, the aggregate effect of multiple SNPs in a window can more greatly associate with traits and be used

to locate QTL.

In chapter 2, we adapt Bayesian mixture models to infer sex-specific genetic effects and G×S interactions. Using simulations, we show that aggregating sex-specific SNP effects within small LD-based windows can enhance the power and precision to detect G×S interactions upon traditional SMR techniques.

## 1.3 Functional implications of noncoding variants on gene expression

In 2009, Visscher et al. [27] popularized the term "causal variant" to mean the genetic variant that causes an observed GWAS association signal. This term is somewhat unfortunate in that GWAS is limited to finding loci with allele content that associates with one or more phenotypes in a population, regardless if a causal relationship (such as a biological mechanism) can explain the association or not. One major observation from GWAS however is that common variants that explain some proportion of complex trait variation are typically within non-coding portions of genomes and enriched in expression QTL [28] (eQTL; QTL that explain variation in transcript abundance for one or more genes). This suggests variation in heritable complex traits is at least partially driven by variation in transcript abundance. More recently, it has been shown that for some complex traits, GWAS hits are equally enriched in splicing QTL (QTL that explain variation in alternatively-spliced isoforms), many of which did not influence transcript abundance [29].

In chapter 2, we show that G×S interactions identified using Bayesian mixture models generally show greater eQTL enrichment than G×S interactions identified from single marker regression. This may indicate our approach for finding G×S interactions is working well toward identifying functional regions that may contribute to sex differences and phenotypic variance. In chapter 3, we characterize a relatively understudied form of gene expression known as RNA editing using a pig model, and in chapter 4 we evaluate the potential that heritable RNA editing variation contributes to complex trait variation in pigs.

## 1.4 RNA editing and its role in gene expression

RNA editing comprises a wide set of modifications to RNA transcripts including deletion, insertion, and substitution of ribonucleotides [30, 31]. In mammals, RNA editing predominantly involves an adenosine to inosine transition within double-stranded pre-mRNA transcripts, catalyzed by adenosine deaminase acting on RNA (ADAR) [32]. At many edited genes, ADAR catalyzes this reaction without perfect efficacy, resulting in only a proportion of transcripts (named the editing level) containing the inosine variant. This means that like alternative splicing, RNA editing enables variation in transcript content from individual to individual, without necessarily affecting transcript abundance.

RNA editing by ADAR has been shown to be essential for the function of some genes and essential for life. For instance, the GluR-B receptor is highly edited at key positions (nearly all transcripts contain the inosine variant) and reduction in GluR-B editing in mice leads to Ca2+ permeability in neural cells and death from seizures [33]. Other editing events are shown to influence complex traits; editing of serotonin receptor 2C is known to influence energy dissipation and fat mass [34]. Perhaps most RNA editing sites show highly variable editing levels in a population [35]. Like SNPs, editing levels that show little variation in the population likely have large effects or essential functions (such as GluR-B editing), while those that exhibit large variation are expected to possess smaller effects but may still explain some proportion of complex trait variation.

Most RNA editing events in human transcriptomes are within Alu elements [32], a type of short interspersed nuclear element (SINE) unique to primates. In chapter 3, we find evidence that RNA activity among pigs largely occurs within PRE-1 elements, a type of SINE element unique to pigs, hogs, and peccaries. This finding has been replicated among several subsequent pig RNA editing studies [36, 37]. In chapter 4, we further evaluate the possibility that highly variable and heritable editing levels may explain variation in complex traits.

# CHAPTER 2

# DECIPHERING SEX-SPECIFIC GENETIC ARCHITECTURES USING LOCAL BAYESIAN REGRESSIONS

This chapter is available on the bioRxiv (doi: 10.1101/653386). It was prepared alongside co-authors Ana I Vazquez, Juan P Steibel, Catherine W Ernst, and Gustavo de los Campos.

## 2.1 Abstract

Many complex human traits exhibit differences between sexes. While numerous factors likely contribute to this phenomenon, growing evidence from genome-wide studies suggest a partial explanation: that males and females from the same population possess differing genetic architectures. Despite this, mapping gene-by-sex (G×S) interactions remains a challenge likely because the magnitude of such an interaction is typically and exceedingly small; traditional genome-wide association techniques may be underpowered to detect such events partly due to the burden of multiple test correction. Here, we developed a local Bayesian regression (LBR) method to estimate sex-specific SNP marker effects after fully accounting for local linkage-disequilibrium (LD) patterns. This enabled us to infer sex-specific effects and G×S interactions either at the single SNP level, or by aggregating the effects of multiple SNPs to make inferences at the level of small LD-based regions. Using simulations in which there was imperfect LD between SNPs and causal variants, we showed that aggregating sex-specific marker effects with LBR provides improved power and resolution to detect G×S interactions over traditional single-SNP-based tests. When using LBR to analyze traits from the UK Biobank, we detected a relatively large G×S interaction impacting bone-mineral density within *ABO* and replicated many previously detected large-magnitude G×S interactions impacting waist-to-hip ratio. We also discovered many new G×S interactions impacting such traits as height and BMI within regions of the genome where both male- and female-specific effects explain a small proportion of phenotypic variance ($R^2 < 1x10^{-4}$), but are enriched in known expression quantitative trait loci. By combining biobank-level data and techniques to estimate

sex-specific SNP effects after accounting for local-LD patterns, we are providing evidence that numerous small-magnitude G×S interactions exist to influence sex differences in a variety of complex traits.

## 2.2 Author Summary

Many complex human traits are known to be influenced by an impressive number of causal variants each with very small effects, posing great challenges for genome-wide association studies (GWAS). To add to this challenge, many causal variants may possess context-dependent effects such as effects that are dependent on biological sex. While GWAS are commonly performed using specific methods in which one single nucleotide polymorphism (SNP) at a time is tested for association with a trait, alternatively we utilize methods more commonly observed in the genomic prediction literature. Such methods are advantageous in that they are not burdened by multiple test correction in the same way as traditional GWAS techniques are, and can fully account for linkage-disequilibrium patterns to accurately estimate the true effects of SNP markers. Here we adapt such methods to estimate genetic effects within sexes and provide a powerful means to compare sex-specific genetic effects.

## 2.3 Introduction

Sex differences are widespread in nature, observed readily among many human traits and diseases. For quantitative traits, sex may affect the distribution of phenotypes at various levels, including mean-differences between genetic males and genetic females (hereafter referred to as males and females, respectively) as well as differences in variance. Sex differences are likely due to a myriad of factors including differential environmental exposures, unequal gene dosages for sex-linked genes as well as sex-heterogeneity in the architecture of genetic effects at one or more autosomal loci (i.e. gene-by-sex (G×S) interactions). In this way, sex is considered an environmental variable, providing two well-defined conditions in which allele frequencies and linkage disequilibrium (LD) patterns are equivalent but nevertheless genetic effects of one or many

autosomal loci may differ.

Evidence for different genetic architectures between sexes among human populations is largely supported by genome-wide parameters [38, 12, 13, 14] including unequal within-sex heritabilities $\left(h^2_{\text{male}} \neq h^2_{\text{female}}\right)$ and between-sex genetic correlations less than one ($r_g < 1$); the former suggests that the proportion of phenotypic variance explained by genetic factors varies between sexes, while the latter suggests genetic effects are disproportional between sexes [15]. Although many traits seem to have between-sex genetic correlation that is evidentially less than one, genome-wide association (GWA) studies intended to map G×S interactions have struggled to pinpoint such loci [17, 39]. Based on this dichotomy, G×S interactions presumably exist for many traits, but the magnitude of a typical G×S interaction is suspected to be exceedingly small, explaining why such events commonly elude detection, particularly after multiple test correction. However, just as numerous small effect causal loci accumulate to affect phenotypic variance, small G×S interactions may accumulate to influence both sex differences and phenotypic variance.

Most GWA studies utilize single-marker regression (SMR), in which the phenotype is regressed upon allele content one SNP at a time, thereby obtaining marginal SNP effect size estimates that do not fully account for LD patterns. In contrast, whole-genome regression methods, in which the phenotype is regressed upon all SNPs across the genome concurrently, fully account for multi-locus LD. These methods are increasingly being used as a one-stop solution to estimate true (conditional) effect sizes of SNP markers and to provide genome-wide estimates including genomic heritability [23, 6, 5] and between-sex genetic correlations [12, 13, 14]. By estimating true SNP effect sizes, the goal across many studies is to select SNPs with non-zero effects and to build a model for predicting polygenic scores [40, 41, 42]. Other works have directly illustrated the use of whole-genome regression methods for GWAS [26, 19, 43, 44]. Whole-genome regressions are computationally challenging to use with biobank-level data; however, recent work suggests relatively accurate genomic prediction and SNP effect estimation can be achieved by simply accounting for local LD patterns (as opposed to global LD patterns) [45].

Building on the idea of utilizing true SNP marker effects, here we developed local Bayesian

regressions (LBR) in which the phenotype is regressed upon multiple SNPs spanning multiple LD blocks (thereby accounting for local LD patterns) to study sex differences in complex traits from the UK Biobank. The LBR model uses random-effect SNP-by-sex interactions [46, 47] that decompose conditional SNP effects into three components: i) one shared across sexes, ii) a male-specific deviation from the shared component, and iii) a female-specific deviation from the shared component. Using samples from the posterior distribution of conditional SNP effects, we developed methods to infer sex-specific effects and G×S interactions at the single SNP level and by aggregating SNP effects within small LD-based regions, offering multiple perspectives to study sex-specific genetic architectures.

In this study, we have utilized genotypes for 607,497 autosomal SNPs from ~259,000 distantly related Caucasians from the UK Biobank for assessing LBR's performance in analyzing simulated and real complex traits including height, BMI, waist-to-hip ratio (WHR), and heel bone-mineral density (BMD). Simulations showed that (i) for inferences of G×S interactions, LBR offers higher power with lower FDR than methods based on marginal effects (aka single-marker regression) and (ii) we show that under imperfect LD between SNPs and causal variants (i.e., when causal variants are not genotyped), aggregating SNP effects within small LD-based regions offers higher power than methods based on testing individual SNPs.

The traits analyzed in this study span a range of genome-wide metrics and G×S suggestibility; from height and BMI for which previous studies indicate males and females possess very similar genetic architectures [13], to WHR, a trait with well-documented G×S interactions [48, 16, 49, 50], and BMD, for which G×S interactions are thought to exist but have eluded detection [51]. LBR provided evidence of G×S interactions impacting height, BMI, and BMD at regions of the genome where sex-specific genetic effects are relatively small, however such regions are enriched in known eQTL. For WHR, LBR replicated many large-magnitude G×S interactions previously discovered using single-marker regression, but also located novel G×S interactions near such genes as the estrogen receptor *ESR1*.

## 2.4 Results

### 2.4.1 Overview of the LBR model, inference methods, and implementation

To study sex differences we regressed male and female phenotypes ($\boldsymbol{y}_{\mathrm{m}}$ and $\boldsymbol{y}_{\mathrm{f}}$) on male and female genotypes ($\boldsymbol{X}_{\mathrm{m}}$ and $\boldsymbol{X}_{\mathrm{f}}$) using a SNP-by-sex interaction model of the form

$$\begin{bmatrix} \boldsymbol{y}_{\mathrm{m}} \\ \boldsymbol{y}_{\mathrm{f}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}\mu_{\mathrm{m}} \\ \mathbf{1}\mu_{\mathrm{f}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{X}_{\mathrm{m}} \\ \boldsymbol{X}_{\mathrm{f}} \end{bmatrix} \boldsymbol{b}_0 + \begin{bmatrix} \boldsymbol{X}_{\mathrm{m}} \\ \mathbf{0} \end{bmatrix} \boldsymbol{b}_{\mathrm{m}} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{X}_{\mathrm{f}} \end{bmatrix} \boldsymbol{b}_{\mathrm{f}} + \begin{bmatrix} \boldsymbol{\varepsilon}_{\mathrm{m}} \\ \boldsymbol{\varepsilon}_{\mathrm{f}} \end{bmatrix}. \tag{2.1}$$

Above, $\mu_{\mathrm{m}}$ and $\mu_{\mathrm{f}}$ are male and female intercepts, $\boldsymbol{b}_0 = b_{0_j} (j = 1, \ldots, p)$ is a vector of main effects, $\boldsymbol{b}_{\mathrm{m}} = b_{\mathrm{m}_j}$ and $\boldsymbol{b}_{\mathrm{f}} = b_{\mathrm{f}_j}$ are male and female interactions, respectively and $\boldsymbol{\varepsilon}_{\mathrm{m}} = \varepsilon_{\mathrm{m}_i}$ and $\boldsymbol{\varepsilon}_{\mathrm{f}} = \varepsilon_{\mathrm{f}_i}$ are male and female errors which were assumed to follow normal distributions with zero mean and sex-specific variances. Female-specific and male-specific SNP effects are defined as $\beta_{\mathrm{f}_j} = b_{0_j} + b_{\mathrm{f}_j}$ and $\beta_{\mathrm{m}_j} = b_{0_j} + b_{\mathrm{m}_j}$, respectively.

#### 2.4.1.1 Prior assumptions

For SNP effects we adopted priors from the spike-slab family with a point of mass at zero and a Gaussian slab [24] specifically, $p\left(b_{k_j}\right) = \pi_k N\left(0, \sigma^2_{b_k}\right) + (1 - \pi_k)\, 1\left(b_{k_j} = 0\right)$ (where $k = 0$, f or m). Here, $\pi_k$ and $\sigma^2_{b_k}$ are hyper-parameters representing the proportion of nonzero effects and the variance of the slab; these hyper-parameters were treated as unknown and given their own hyperpriors (see Methods).

#### 2.4.1.2 Local-regression

Implementing the above model with whole-genome SNPs ($p \sim 600$K) and very large sample size ($n \sim 300$K) is computationally extremely challenging. However, LD in homogeneous un-structured human populations spans over relatively short regions ($R^2$ between allele dosages typically vanishes within 1-2 Mb; Fig A.1). Therefore, we applied LBR to long, overlapping chromosome segments (Fig 2.1). Specifically, we divided the genome into "core" segments containing 1,500 contiguous

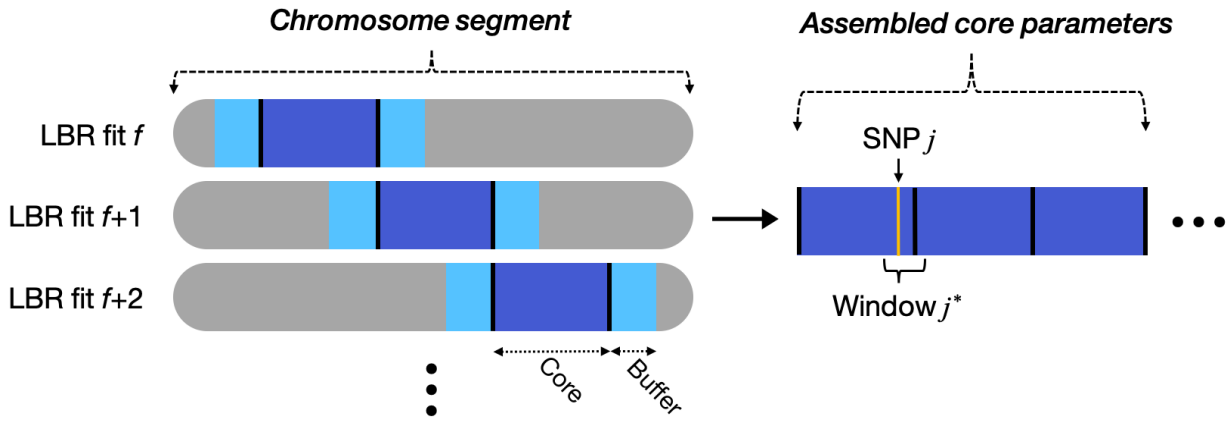


Figure 2.1: Strategy for implementing local Bayesian regressions genome-wide
The phenotype is regressed upon multiple sequential SNPs using a sliding window approach. The core region contained 1500 SNPs (roughly 8Mb, on average) and each buffer region contained 250 SNPs (roughly 1Mb, on average). Core parameters (posterior samples) are stitched together, then sex-specific effects and G×S interactions are inferred at the level of SNP $j$ and window $j^*$.

SNPs (roughly 8Mb, on average), then applied the regression in equation 2.1 to SNPs in the core segment plus 250 SNPs (i.e., roughly 1Mb) in each flanking region, which were added to account for LD between SNPs at the edge of each core segment with SNPs in neighboring segments.

#### 2.4.1.3 Inferences

We used the BGLR [52] software to draw samples from the posterior distribution of the model parameters and used these samples to make inference about individual SNP effects including: (i) the posterior probability that the $j^{\text{th}}$ SNP has a nonzero effect in males $\left(\text{PPM}_{\text{SNP}_j}\right)$ and females $\left(\text{PPF}_{\text{SNP}_j}\right)$ and (ii) the posterior probability that the female and male effects are different $\left(\text{PPDiff}_{\text{SNP}_j}\right)$.

In regions involving multiple SNPs in strong LD, inferences at the individual-SNP level may be questionable. Therefore we borrowed upon previous work by Fernando et al. [26], enabling us to aggregate multiple sex-specific SNP effects within relatively small regions using "window variances". For each SNP $j$ we defined a window $j^*$ around the SNP based on local LD patterns (see Methods). We then defined the male-specific and female-specific window variances as $\sigma^2_{g\text{m}_{j^*}} =$

$var\left(\boldsymbol{X}_{j^*}\boldsymbol{\beta}_{\mathrm{m}_{j^*}}\right)$ and $\sigma^2_{g_{\mathrm{f}_{j^*}}} = var\left(\boldsymbol{X}_{j^*}\boldsymbol{\beta}_{\mathrm{f}_{j^*}}\right)$, respectively. Here, $\boldsymbol{X}_{j^*}$ represent genotypes at SNPs within the $j^*$ window and $var()$ is the sample variance operator. Prior to model fitting, the phenotype is scaled across sexes; thus, sex-specific window variances may be interpreted as the proportion of total phenotypic variance explained by sex-specific SNP effects. From samples of sex-specific window variances, we computed the posterior probability of (i) nonzero male-specific window variance $\left(\mathrm{PPM}_{\sigma^2_{g_{j^*}}}\right)$, (ii) nonzero female-specific window variance $\left(\mathrm{PPF}_{\sigma^2_{g_{j^*}}}\right)$, and (iii) sex difference in window variances $\left(\mathrm{PPDiff}_{\sigma^2_{g_{j^*}}}\right)$.

### 2.4.2 LBR offers improved power with lower false-discovery rates

We used simulations to assess the power and false discovery rate (FDR) of LBR and to compare it with that of standard single-marker-regression (SMR). Traits were simulated using SNP genotypes from the Axiom UK-Biobank (119,190 males and 139,738 females, all distantly related Caucasians). We simulated a highly complex trait with one causal variant (CV) per ~ 2Mb which on average explained a proportion of the phenotypic variance equal to $3.3\text{x}10^{-4}$. Our simulation used a total of 60,000 SNPs (consisting of 6,000 consecutive SNPs taken from 10 different chromosomes) and 150 CVs; on the complete human genome "scale" this corresponds to a trait with 1,500 CVs and a heritability of 0.5 (see Methods for further details). 40% of the CVs (a total of 60 SNPs in our simulation) had differing sex-specific effects and the remaining 60% (90 SNPs) had effects that were the same in males and females.

#### 2.4.2.1 Power and FDR when causal variants are genotyped

First, we analyzed the simulated phenotypes using all SNPs (including the 150 causal ones). Initially interested in inferring G×S interactions, we ranked SNPs based on LBR's $\mathrm{PPDiff}_{\mathrm{SNP}_j}$ metric and based on SMR's *p*-value for sex difference (*pvalue*-diff , see Methods) and used the two ranks to estimate power and FDR as a function of the number of SNPs selected (Fig 2.2). LBR showed consistently higher power (achieving a power of 80% when selecting the top-50 SNPs with

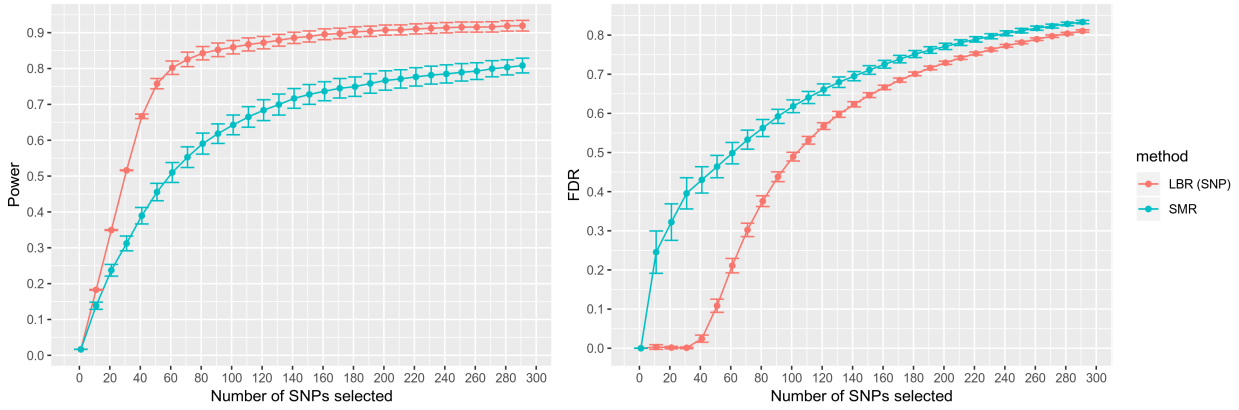


Figure 2.2: Estimated power and false-discovery rate for discovering observed SNPs with G×S interactions

Shown as a function of the number of SNPs selected. Each point represents a sample average and error bars represent 95% confidence intervals, each derived using 30 Monte Carlo replicates. LBR (SNP): local Bayesian regression, utilizing $\text{PPDiff}_{\text{SNP}_j}$. SMR: single-marker regression, utilizing *pvalue*-diff.

highest $\text{PPDiff}_{\text{SNP}_j}$ and lower FDR than SMR. The false discovery rate of LBR was very low when selecting the top-50 SNPs with highest $\text{PPDiff}_{\text{SNP}_j}$ and exhibited a very sharp phase-transition with fast increase in FDR thereafter.

We also compared the two methods based on arbitrary, albeit commonly used, mapping thresholds for SMR and LBR. At $\text{PPDiff}_{\text{SNP}_j} \geq 0.95$, LBR selected an average (across simulation replicates) of 38.33 SNPs with an estimated power of 0.634 and estimated FDR of 0.007. Conversely, at *pvalue*-diff $\leq 5\text{x}10^{-8}$, SMR selected an average of 50.7 SNPs with an estimated power of 0.436 and estimated FDR of 0.451. Altogether, these results suggest that for G×S discovery, LBR offers higher power and lower FDR than SMR—the method most widely used in GWA studies—at least when G×S interactions are observed.

When trying to map SNPs that had effect in at least one sex, we used

$$\text{PP}_{\text{SNP}_j} = \max\left[\text{PPM}_{\text{SNP}_j}, \text{PPF}_{\text{SNP}_j}\right]$$

and $p$-values from an F-test (see Methods) as metrics for LBR and SMR methods, respectively. Again, LBR showed higher power with lower FDR than a standard SMR $p$-value (Fig A.2). At traditional mapping thresholds, LBR and SMR had similar power but LBR achieved that power

with much lower FDR; at $PP_{SNP_j} \geq 0.95$, the average number of SNPs selected was 120.83 with an estimated power of 0.799 and estimated FDR of 0.009 while at *p*-value $\leq 5x10^{-8}$, the number of SNPs selected was 374.56 with an estimated power of 0.794 and FDR of 0.66.

#### 2.4.2.2 Power and FDR under imperfect LD

In a second round of analyses, we removed all CVs from the panel of SNPs used in the analysis to represent a situation where CVs are not observed, and genotyped SNPs are tagging CVs at varying degrees. As before, we initially assessed the relative performance of LBR to infer segments harboring G×S interactions. Power and FDR were assessed at several resolutions: 1Mb, 500Kb and 250Kb regions around each CV. At each resolution, a discovery was considered true if the finding laid within a segment harboring a G×S CV. Power and FDR were computed at different thresholds ($PPDiff_{SNP_j}$ and $PPDiff_{\sigma^2_{g_{j*}}}$ for LBR and *pvalue*-diff for SMR; Fig 2.3). When using a 1Mb target area—such that correct G×S discoveries must be within 500Kb on either side of a true G×S event—$PPDiff_{\sigma^2_{g_{j*}}}$ thresholds (LBR's window-based metrics) provided highest power within an FDR range of 0-0.3, thereafter SMR provided slightly higher power. As expected, when removing CVs, power was estimated to be much lower than when CVs were observed; at $PPDiff_{\sigma^2_{g_{j*}}} \geq 0.95$, the estimated power and FDR were 0.454 and 0.004, respectively, while at *pvalue*-diff $\leq 5x10^{-8}$, estimated power and FDR were 0.22 and 0.006. As seen in Fig 2.3, when considering a finer resolution (500Kb and 250Kb) the performance of both LBR-based approaches was more robust than SMR. Altogether this indicates that for the discovery and mapping of unobserved G×S interactions, LBR's window-based metric provides higher power with equivalent FDR and finer resolution than single-marker regression methods.

To infer segments containing CVs that affect at least one sex, we again used LBR to decide whether either sex-specific effect was nonzero at the level of individual SNPs or windows. Using a 1MB target area, LBR's window-based metrics provided the highest power within an FDR range of 0-0.025. When decreasing the target area, LBR provided the highest power over larger FDR ranges (Fig A.3).

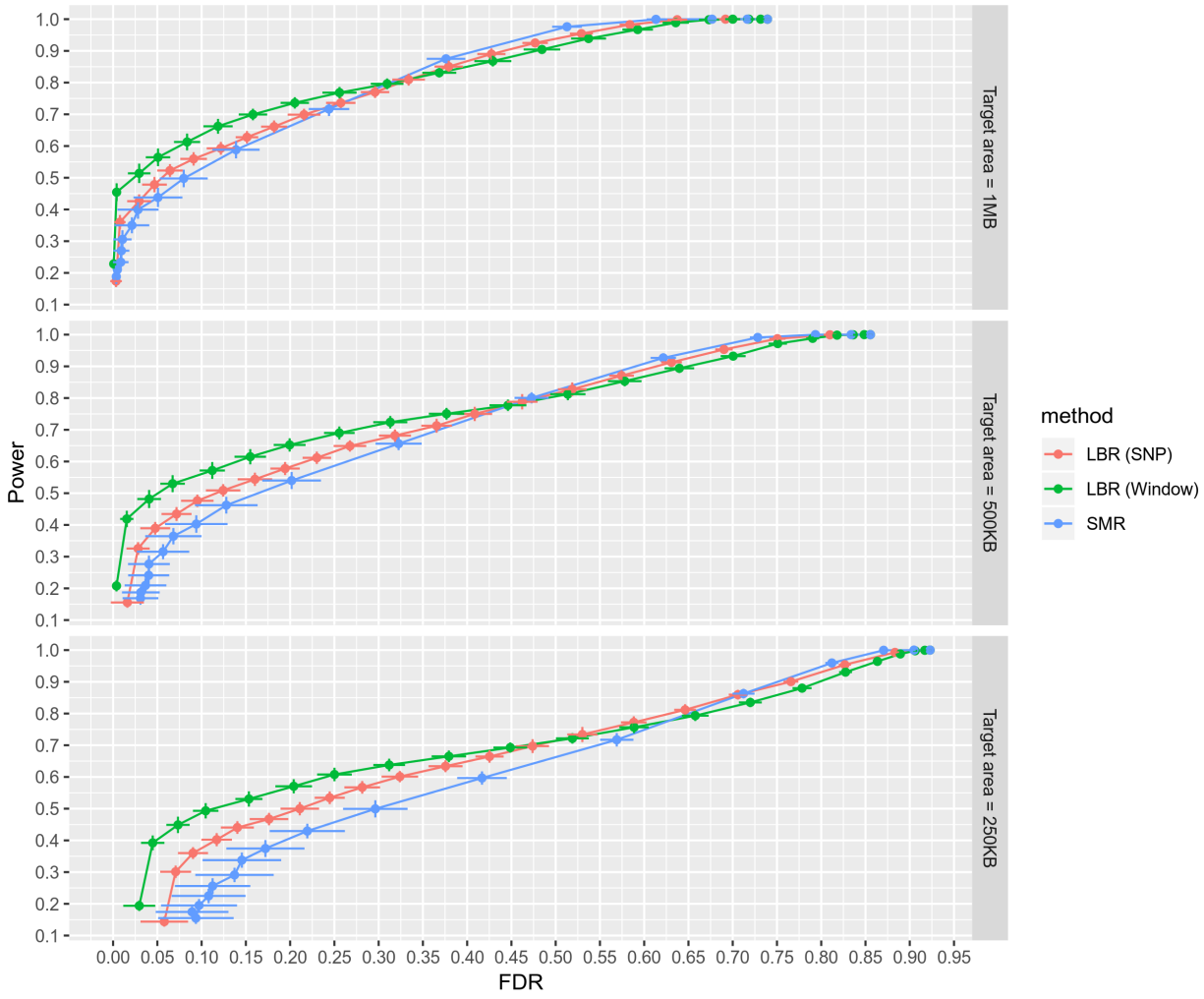


Figure 2.3: Power vs false-discovery rate for discovering genomic regions containing masked G×S interactions

Here power is defined as the expected proportion of G×S interactions that are being tagged by at least one selected SNP $j$ or window $j^*$. False discovery rate is defined as the expected proportion of selected SNPs or windows that are not tagging any G×S interactions. Each point is an estimate and error bars for both axes represent 95% confidence intervals. Point estimates and intervals were derived using 30 Monte Carlo replicates. Each facet corresponds to a different "target area", a fixed width around each G×S interaction that defines the set of SNPs effectively tagging it. LBR (SNP): uses the $\text{PPDiff}_{\text{SNP}_j}$ metric spanning 1-0. LBR (Window): uses the $\text{PPDiff}_{\sigma^2_{g_{j^*}}}$ metric spanning 1-0. SMR: uses the *pvalue*-diff metric spanning (on the -$\log_{10}$ scale) 8-0.

### 2.4.3 For real human traits, many newly discovered G×S interactions show relatively small sex-specific effects

We analyzed four complex human traits (height, BMI, BMD, and WHR) measured among ~259,000 distantly related Caucasians from the UK Biobank (~119,000 males and ~140,000 females). For each trait, we fit the LBR model (equation 2.1) across the entire autosome consisting of 607,497 genotyped SNPs using 417 overlapping segments (Fig 2.1) and obtained evidence of G×S interactions at the level of SNP $j$ and window $j^*$.

To compare both the magnitude and sign of sex-specific SNP effects, we plotted each $\hat{\beta}_{\mathrm{f}_j}$ against $\hat{\beta}_{\mathrm{m}_j}$ (Fig 2.4A). The trait was scaled across sexes prior to model fitting; thus, male- and female-specific effects were not constrained to the same scale. In this way, one might expect male-specific SNP effects to uniformly differ from female-specific SNP effects by a multiplicative factor if the variance of the phenotype is different between sexes (sample statistics within each sex are provided within Table A.1). Surprisingly, we did not observe evidence of sex-specific SNP effects uniformly differing due to differences in phenotypic scale; for height, BMD, and BMI, as seen in Fig 2.4A, most large effect SNPs lie near the blue diagonal line. For WHR, we observed largely consistent results from prior studies [48, 16, 49]: namely the prevalence of numerous SNPs with relatively large effects in females but little to no effect in males. No traits exhibited evidence of any SNPs with (i) high confidence male- and female- specific effects (no SNPs with $\mathrm{PPM}_{\mathrm{SNP}_j} \geq 0.9$ and $\mathrm{PPF}_{\mathrm{SNP}_j} \geq 0.9$) and (ii) differing signs between sexes.

We then aggregated sex-specific SNP effects within small LD-based regions to estimate sex-specific window variances $\sigma^2_{g\mathrm{m}_{j^*}}$ and $\sigma^2_{g\mathrm{f}_{j^*}}$ and compared the magnitude of each (Fig 2.4B). Interestingly for traits such as height, many large effect regions bear slightly larger window variances for males than for females. This was not observed at the single SNP level, suggesting that many regions bearing numerous small effect SNPs produce aggregate effects that are potentially larger (although not reaching a $\mathrm{PPDiff}_{\sigma^2_{g_{j^*}}} \geq 0.9$ threshold) in males than in females. One example is the *GDF5* locus, previously known to strongly associate with adult height [53], where a peak $\mathrm{PPDiff}_{\sigma^2_{g_{j^*}}}$ signal centered on rs143384 had slightly different estimated sex-specific window

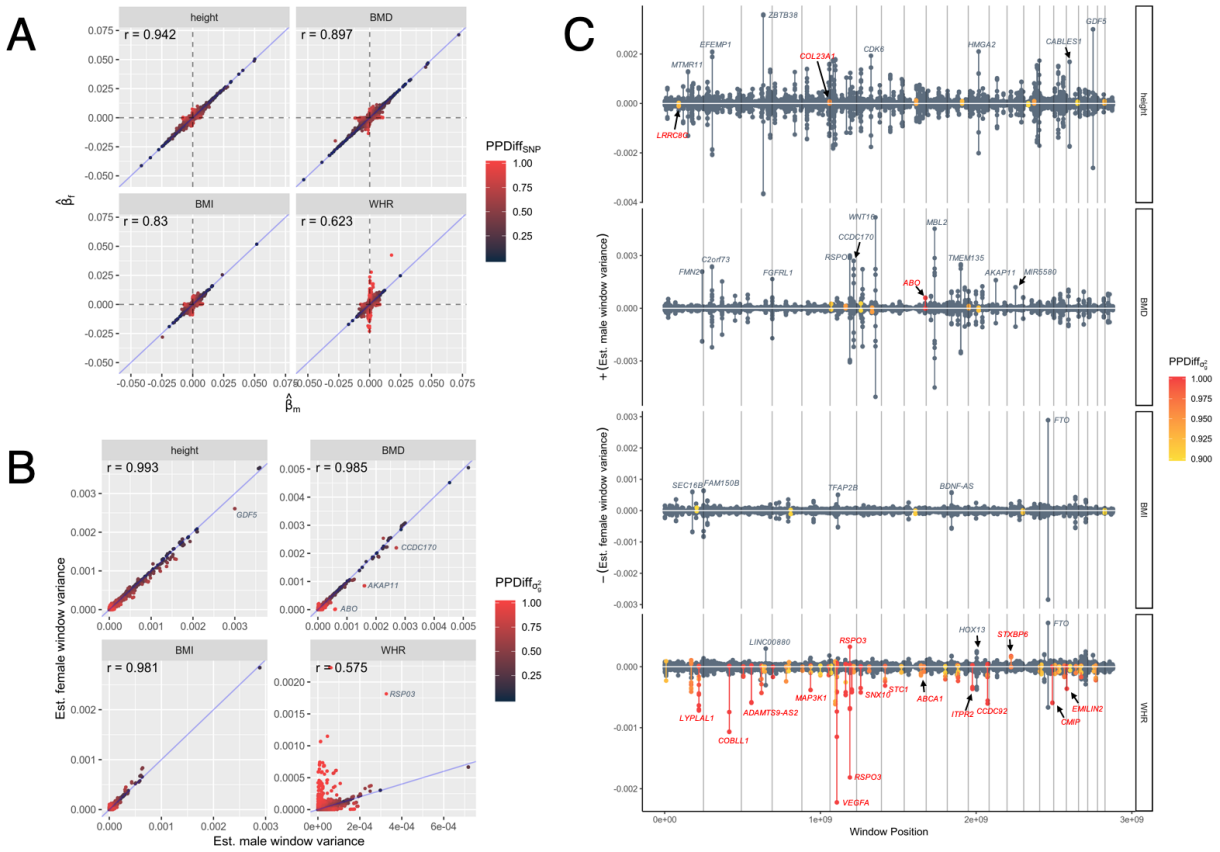


Figure 2.4: Comparing sex-specific genetic effects

**(A)** Plot of estimated female SNP effects against estimated male SNP effects for all 607,497 genotyped autosomal SNPs. Points are colored by their posterior probability of sex difference at the level of individual SNPs. **(B)** Plot of estimated female window variances against estimated male window variances for all 607,497 LD-based windows, with each window $j^*$ centered on a different focal SNP $j$. Points are colored by their posterior probability of sex difference at the level of window variances. **(C)** Miami-like plot depicting location and magnitude of G×S interactions identified through sex-specific window variances. For each trait, showing estimated male window variance above the x-axis and estimated female window variance below the x-axis. Vertical lines denote changing chromosomes. A sample of windows is labeled with nearest gene annotation, obtained from Axiom UKB WCSG annotations, release 34. Gray labels indicate nearest genes with relatively large window variances evidently shared across sexes, while red labels indicate nearest genes with detected G×S interactions.

variances ($\hat{\sigma}^2_{g\mathrm{m}_{j^*}}$ = 3.0x10$^{-3}$ and $\hat{\sigma}^2_{g\mathrm{f}_{j^*}}$ = 2.6x10$^{-3}$) but weak evidence of a G×S interaction PPDiff$_{\sigma^2_{g_{j^*}}}$ = 0.544). For BMD, several large effect regions show suggestive evidence of G×S interactions including the *AKAP11* locus and the *CCDC170* locus (PPDiff$_{\sigma^2_{g_{j^*}}}$ = 0.856 and 0.745, respectively), both previously associated with bone mineral density [54, 55, 56, 57].

To make G×S inferences at the level of window variances irrespective of the magnitude of sex-specific effects, we adopted a PPDiff$_{\sigma^2_{g_{j^*}}}$ threshold of 0.9, which in simulations (Fig 2.3) provided optimal power at an estimated FDR of 0.029 when using a 1MB target area. For height, a total of eight distinct regions possessed a PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$, two of which possessed a PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.95$. For BMI, 5 distinct regions possessed a PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$ with none reaching a more stringent PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.95$ threshold, and none overlapping with two previously suggested BMI G×S SNPs [58]. As seen in Fig 2.4C, inferred G×S interactions for height and BMI possess relatively small sex-specific window variances; as an example, for height, the window centered on SNP rs1535515 (near *LRRC8C*) had a PPDiff$_{\sigma^2_{g_{j^*}}}$ = 0.96, while $\hat{\sigma}^2_{g\mathrm{m}_{j^*}}$ = 2.1x10$^{-5}$ and $\hat{\sigma}^2_{g\mathrm{m}_{j^*}}$ = 1.1x10$^{-4}$. For BMD, seven regions reached a 0.9 PPDiff$_{\sigma^2_{g_{j^*}}}$ threshold while one higher-confidence G×S interaction (PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.95$) was detected within *ABO*, the gene controlling blood type.

For WHR, roughly 45 distinct genomic regions possessed a PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$, while 34 of these possessed a PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.95$. We found many previously detected G×S interactions known to associate with WHR or a related trait, WHR adjusted for BMI (WHRadjBMI) [48, 16, 49, 50]. These included interactions at *LYPLAL1*, *MAP3K1*, *COBLL1*, *RSPO3*, and *VEGFA* among others. We also detected numerous novel G×S interactions (Table 2.1) near physiologically intriguing genes such as the estrogen receptor gene *ESR1* and the ATP binding cassette transporter A1 gene *ABCA1* known to play a role in HDL metabolism (PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.95$). As seen in Table 2.1, both novel signals possessed a high-confidence female-specific effect with weak evidence for a male-specific effect (PPF$_{\sigma^2_{g_{j^*}}} \geq 0.95; \mathrm{PPM}_{\sigma^2_{g_{j^*}}} \leq 0.6$), however the magnitude of the female-specific effect was relatively small ($\hat{\sigma}^2_{g\mathrm{f}_{j^*}} \leq 1.4 \times 10^{-4}$). As evident from Table 2.1, most novel WHR G×S interactions detectable with LBR are those with relatively small sex-specific effects.

Additionally, we utilized a traditional SMR approach (see Methods) for the discovery of G×S

Table 2.1: G×S interactions inferred through sex-specific window variances[a]

| Focal SNP[b] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [c] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [c] | PPM$_{\sigma^2_{g_{j^*}}}$ | PPF$_{\sigma^2_{g_{j^*}}}$ | PPDiff$_{\sigma^2_{g_{j^*}}}$ | Nearest gene[d] | location | eQTL[e] |
|---|---|---|---|---|---|---|---|---|---|
| rs8176719 | BMD | 0.06000 | 0.00182 | 1.000 | 0.794 | 1.000 | *ABO* | exon/frameshift | yes |
| rs1535515 | height | 0.00211 | 0.01170 | 0.819 | 0.999 | 0.956 | *LRRC8C* | intron | yes |
| rs1544926 | height | 0.00763 | 0.00035 | 0.983 | 0.418 | 0.955 | *COL23A1* | UTR-3 | yes |
| rs6905288 | WHR | 0.00567 | 0.22200 | 0.920 | 1.000 | 1.000 | ***VEGFA*** | downstream | |
| rs72961013 | WHR | 0.03260 | 0.18100 | 1.000 | 1.000 | 1.000 | ***RSPO3*** | downstream | |
| rs1128249 | WHR | 0.00132 | 0.10700 | 0.614 | 1.000 | 1.000 | ***COBLL1*** | intron | yes |
| rs12022722 | WHR | 0.00080 | 0.07180 | 0.490 | 1.000 | 1.000 | ***LYPLAL1*** | downstream | yes |
| rs1776897 | WHR | 0.00870 | 0.06140 | 0.976 | 1.000 | 0.950 | ***HMGA1*** | upstream | yes |
| rs11057401 | WHR | 0.00438 | 0.06030 | 0.846 | 1.000 | 1.000 | ***CCDC92*** | exon/missense | yes |
| rs17777180 | WHR | 0.00031 | 0.05950 | 0.291 | 1.000 | 1.000 | ***CMIP*** | intron | yes |
| rs4607103 | WHR | 0.00195 | 0.05920 | 0.809 | 1.000 | 1.000 | ***ADAMTS9-AS2*** | intron | yes |
| rs6937293 | WHR | 0.00457 | 0.04660 | 0.839 | 1.000 | 1.000 | ***LOC728012*** | downstream | yes |
| rs16861373 | WHR | 0.00066 | 0.04300 | 0.389 | 1.000 | 0.995 | ***PLXND1*** | intron | |
| rs73068463 | WHR | 0.00068 | 0.04220 | 0.461 | 1.000 | 1.000 | ***SNX10*** | intron | yes |
| rs9376422 | WHR | 0.00107 | 0.04180 | 0.524 | 1.000 | 1.000 | *LOC645434* | upstream | |
| rs6867983 | WHR | 0.00192 | 0.03820 | 0.440 | 1.000 | 0.998 | ***MAP3K1*** | upstream | |
| rs2171522 | WHR | 0.00241 | 0.03650 | 0.561 | 1.000 | 0.998 | ***ITPR2*** | downstream | yes |
| rs3810068 | WHR | 0.00026 | 0.03590 | 0.174 | 1.000 | 1.000 | *EMILIN2* | upstream | yes |
| rs568890 | WHR | 0.00129 | 0.03110 | 0.809 | 1.000 | 1.000 | ***NKX2-6*** | upstream | yes |
| rs1332955 | WHR | 0.00647 | 0.02940 | 0.970 | 1.000 | 0.973 | ***LOC284688*** | downstream | yes |
| rs13133548 | WHR | 0.00019 | 0.02400 | 0.175 | 0.969 | 0.956 | ***FAM13A*** | intron | yes |
| rs11263641 | WHR | 0.00207 | 0.02340 | 0.723 | 1.000 | 0.991 | *MYEOV* | downstream | yes |
| rs2800999 | WHR | 0.00201 | 0.02220 | 0.691 | 1.000 | 0.979 | *TSHZ2* | intron | |
| rs2244506 | WHR | 0.00101 | 0.02070 | 0.453 | 0.998 | 0.985 | *MIR5694* | downstream | |
| rs7259285 | WHR | 0.00182 | 0.01710 | 0.767 | 1.000 | 0.989 | *HAUS8* | downstream | yes |
| rs4450871 | WHR | 0.00002 | 0.01680 | 0.027 | 1.000 | 1.000 | *CYTL1* | downstream | |
| rs4080890 | WHR | 0.00153 | 0.01630 | 0.594 | 0.999 | 0.975 | ***KCNJ2*** | downstream | |
| rs4684859 | WHR | 0.00039 | 0.01570 | 0.330 | 0.998 | 0.994 | ***PPARG*** | downstream | |
| rs7704120 | WHR | 0.00049 | 0.01370 | 0.476 | 0.998 | 0.991 | *STC2* | downstream | |
| rs10991417 | WHR | 0.00048 | 0.01230 | 0.339 | 0.986 | 0.966 | *ABCA1* | intron | yes |
| rs12454712 | WHR | 0.00087 | 0.01020 | 0.360 | 0.996 | 0.965 | ***BCL2*** | intron | yes |
| rs62070804 | WHR | 0.00004 | 0.00887 | 0.052 | 0.969 | 0.961 | *ABHD15* | exon/missense | yes |
| rs10760322 | WHR | 0.00027 | 0.00812 | 0.282 | 0.986 | 0.968 | *LHX2* | downstream | |
| rs1361024 | WHR | 0.00022 | 0.00760 | 0.203 | 0.982 | 0.962 | *ESR1* | intron | |
| rs1358503 | WHR | 0.00021 | 0.00716 | 0.309 | 0.989 | 0.966 | *SEMA3C* | upstream | yes |
| rs13156948 | WHR | 0.00016 | 0.00660 | 0.079 | 0.970 | 0.957 | *IRX1* | downstream | |
| rs12432376 | WHR | 0.01740 | 0.00074 | 1.000 | 0.552 | 0.994 | *STXBP6* | upstream | |

[a] Listed are loci with at least 0.95 posterior probability that sex-specific window variances differ. The table is sorted first by trait, then by magnitude of the female-specific window variance. Results are filtered such that each window listed consisted of a distinct set of SNPs. A full list of all G×S signals at a PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.90$ threshold is provided in Table A.2.

[b] Focal SNP is defined as the center SNP $j$ in window $j^*$.

[c] The proportion of variance explained by sex-specific SNP effects, expressed as a percentage.

[d] Nearest gene and location identified through Axiom UKB WCSG annotations, release 34. The gene/locus is bold if it has been previously detected as a G×S interaction for WHR or WHR adjusted for BMI [48, 16, 49, 50].

[e] If "yes", the focal SNP is significantly associated with gene expression in at least one tissue, according to GTEx V7.

interactions among traits to compare *pvalue*-diff signals to PPDiff$_{\sigma^2_{g_{j^*}}}$ signals (Fig A.4). At *pvalue*-diff $\leq 5\text{x}10^{-8}$, there were no genome-wide significant G×S-interacting SNPs for height, one significant SNP for BMI nearby a window with PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$, and one significant peak within *ABO* for BMD (the same signal detected using PPDiff$_{\sigma^2_{g_{j^*}}}$). Regions with a PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$ generally coincided with at least nominally-significant *pvalue*-diff signals; for height and BMD, regions with PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$ also possessed a peak SNP with *pvalue*-diff $\leq 0.01$. For BMI, PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$ signals possessed a peak SNP of *pvalue*-diff $\leq 0.1$. This, together with the fact that novel G×S interactions found using LBR possess relatively small sex-specific effects, suggests that LBR may be detecting G×S interactions that are otherwise missed due to low power. Lastly for WHR, most of the high-confidence PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$ signals coincided with clear and obvious *pvalue*-diff peaks.

### 2.4.4 Inferred G×S interactions are enriched in tissue-specific eQTL

As seen previously, many G×S interactions inferred using LBR have exceedingly small sex-specific effects. To further investigate whether G×S detections using the PPDiff$_{\sigma^2_{g_{j^*}}}$ metric may be functionally relevant, we inferred whether such signals are enriched in eQTL identified from GTEx. Specifically, using a hypergeometric test we asked whether PPDiff$_{\sigma^2_{g_{j^*}}}$-selected focal SNPs (SNP $j$ within window $j^*$) were enriched in eQTL, then compared to eQTL enrichment from *pvalue*-diff-selected SNPs as a function of the number of SNPs selected (Fig A.5). PPDiff$_{\sigma^2_{g_{j^*}}}$-selected focal SNPs showed consistently higher eQTL enrichment than *pvalue*-diff-selected SNPs for all traits except WHR. For instance, at PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$, the total number of windows (focal SNPs) selected was 36, 264, 34, and 13, for height, WHR, BMD, and BMI, respectively. With these selections, eQTL enrichment $p$-values were $2.39\text{x}10^{-4}$, $1.52\text{x}10^{-12}$, $2.01\text{x}10^{-12}$, and $8.33\text{x}10^{-4}$, for height, WHR, BMD, and BMI, respectively. When selecting the same number of SNPs using *pvalue*-diff, enrichment p-values were $2.25\text{x}10^{-2}$, $1.56\text{x}10^{-28}$, $5.54\text{x}10^{-8}$, $1.93^{-1}$, for height, WHR, BMD, and BMI, respectively.

To provide more information about how genetic regions bearing G×S interactions may impact

gene expression in specific tissues, we determined whether focal SNPs at PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$ are enriched in tissue-specific eQTL (Fig 2.5). For height, BMD, and WHR, such SNPs showed significant eQTL enrichment in at least one tissue, using a conservative bonferroni corrected enrichment p-value of $2.6\text{x}10^{-4}$ (correcting for 192 tests in total; 48 tissues and 4 traits). Interestingly, BMD's G×S signals are very strongly enriched in eQTL with associated eGenes (including *ABO* and *CYP3A5*) expressed in the adrenal gland, among other tissues. For height, we observed small enrichment *p*-values across many tissues since G×S focal SNPs are enriched in eQTL with associated eGenes (including *LOC101927975* and *CNDP2*) expressed across many tissues. Lastly for WHR, we observed G×S detections to be heavily enriched in eQTL with associated eGenes expressed in fibroblast, adipose, and skin tissues.

## 2.5 Discussion

We have investigated the degree to which sex-specific genetic architectures differ at local regions, using large biobank data (N ~ 119,000 males and ~140,000 females) and Bayesian multiple regression techniques that estimate sex-specific marker effects accounting for local LD patterns. The flexibility of the Bayesian approach enables multi-resolution inference of sex-specific effects: from individual SNP effects to window-variances that aggregate SNP effects within chromosome segments. These inferences can be drawn all using the results of the same model fit (equation 2.1) but different post-processing of samples of SNP effects from the posterior distribution.

The Bayesian multiple regression technique performed in this study, along with estimation of window variances, was largely inspired by Fernando et al. [26]. In that study, windows were defined using disjoint, fixed intervals. In contrast, for each SNP we define a window based on local LD patterns, resulting in heavily overlapping, dynamically sized windows. The methods presented here also bear resemblance to those of Vilhjálmsson et al. [45], which utilized point-normal priors to estimate human SNP effects after accounting for local LD patterns. In that study, posterior means of SNP effects were estimated for the purposes of prediction while in this study, we numerically derive the full posterior distribution, allowing for inference of non-null SNP effects and window

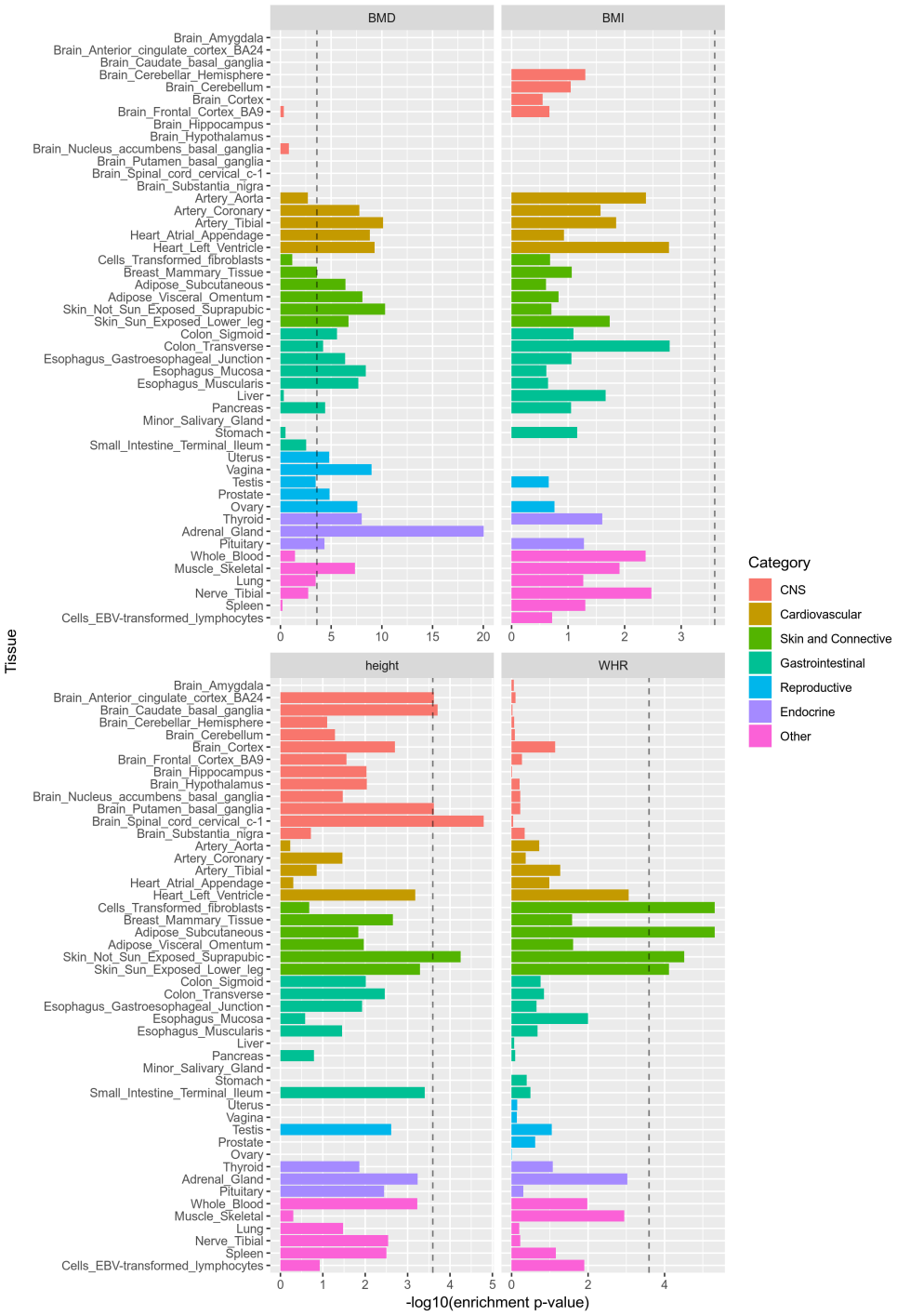


Figure 2.5: Evidence that LBR-identified G×S interactions are enriched in tissue-specific eQTL Plotted on the x-axis is the *p*-value obtained from a hypergeometic test providing evidence that focal SNPs selected using PPDiff $\sigma^2_{g_{j*}} \geq 0.9$ are enriched in tissue-specific eQTL. The dashed line represents a Bonferroni corrected significance threshold of $2.6 \times 10^{-4}$.

variances.

Through simulations, we showed that local Bayesian regressions (LBR) provide superior power and precision to detect causal variants and those specifically bearing G×S interactions. We rationalize improvements in power upon traditional SMR methods by noting that the magnitude of a typical causal variant or G×S interaction is exceedingly small and can elude hypothesis testing partly due to the burden of multiple test correction. We also note that the resolution (peak size) in SMR signals is relatively large when using large sample sizes (due to not fully accounting for local LD patterns). To overcome this problem, we provided evidence that LBR methods—either by estimating true marker effects or by aggregating true marker effects within relatively small regions—can achieve improved resolution when working with large sample sizes such as biobank-level data.

When using LBR to analyze real human traits, we have provided credence to our posterior probability-based discoveries by determining that LBR-detected G×S interactions are generally more enriched in eQTL than SMR-detected interactions. For BMD, we provided new evidence that sex-specific effects differ within *ABO* and that G×S interactions are highly enriched in adrenal gland-specific eQTL. This encourages the hypothesis that some G×S are eQTL that may modulate gene expression in the adrenal gland, with gene function dependent on the presence or absence of sex hormones. This was also an intriguing finding given that *ABO* blood groups have been known to associate with osteoporosis and osteoporosis severity [59, 60]. For WHR, we detected previously known, large-magnitude G×S interactions that were discovered using WHR or WHRadjBMI [48, 16, 49, 50], but additionally discovered novel, small magnitude G×S interactions near such genes as *ESR1* and *ABCA1*. In a previous work analyzing WHRadjBMI, *ABCA1* showed a significant female-specific genetic effect only, however the test for G×S interaction failed to reach significance [50].

For traits like height and BMI, large effect loci are estimated to have very similar effects between males and females and loci with evidence of G×S interactions were those possessing relatively small sex-specific effects. As seen in Fig 2.4B, many relatively large window variances for height are estimated to be slightly higher for males than for females albeit not reaching a $\text{PPDiff}_{\sigma^2_{g_{j*}}} \geq 0.9$

threshold. This is consistent with the fact that the global genomic variance for height was estimated to be higher in males than in females in a previous study using the interim release of the UK Biobank [14]. Similarly, the same prior study estimated the global genomic variance of BMI to be higher in females than in males and we observe, if anything, evidence of sex-specific window variances leading to the same conclusion. These observations may potentially indicate that relatively large causal variants have slightly different sex-specific effects for traits like height and BMI, however, if that is the case we are still underpowered to confidently detect such interactions.

It is important to acknowledge that while the methods presented here appear useful to decipher sex-specific genetic architectures from large human samples, additional work will be required to determine how these techniques may infer heterogeneous genetic effects in other contexts (other types of gene-by-covariate interactions), or when using different sample sizes or samples from different populations. With large sample sizes, the increased power and flexibility of LBR comes with the cost of a significantly larger computational burden than the one involved in the traditional SMR approach; however, working with large datasets can be made manageable by adjusting the size of each fitted segment (Fig 2.1) and parallel processing the fitting of each segment. Alternatively, LBR may be used as a follow up to traditional SMR tests, using pre-selected regions of interest. Another limitation inherent to aggregating SNP effects using window variances is that the sign of the effect is lost. In this way, when inferring G×S interactions through window variance differences, we cannot comment on whether sex-specific effects had the same sign or differing signs.

To conclude, we have demonstrated the powerful and flexible use of local Bayesian regressions for GWA to infer sex-specific genetic effects and G×S interactions using the UK Biobank. This was largely done by showing various means to utilize estimates of true (accounting for local LD), sex-specific SNP marker effects for GWA even when causal variants are not on the SNP panel for analysis. We anticipate that many more traits will be analyzed with this method to increasingly learn more about what is contributing to differences between males and females in human populations.

## 2.6 Methods

### 2.6.1 Genotype data

Individuals from the UK Biobank [61] were genotyped using the custom UK Biobank Axiom Array (http://www.ukbiobank.ac.uk/scientists-3/uk-biobank-axiom-array/) containing ~ 800,000 SNPs. SNP quality control proceeded with the Caucasian cohort (N = 409,700); SNPs with a minor allele frequency < 0.01 and missing call rate > 0.05 were removed. SNPs from sex chromosomes and the mitochondrial chromosome were not considered in this study, resulting in 607,497 autosomal SNPs. Individuals with coefficient of relatedness of 0.03 or greater were removed from analysis, resulting in 258,928 distantly related genotyped individuals for use in this study.

### 2.6.2 Phenotype data

All phenotypic data was collected using baseline measurements of UK Biobank participants. For height, the description "Standing height" from the UK Biobank was used. Individuals with heights (cm) less than 147 or more than 210 were removed from analysis. For BMD, the descriptions "Heel bone mineral density (BMD)", "Heel bone mineral density (BMD) (left)", and "Heel bone mineral density (BMD) (right)" were used in conjunction; for individuals with missing "Heel bone mineral density (BMD)" records, either the (left), the (right), or if available, the average between (left) and (right) was used. For BMI, the description "Body mass index (BMI)" was used and for WHR, the ratio of "Waist circumference" to "Hip circumference" was used. Prior to model fitting, all traits were pre-corrected for sex, age, batch, genotyping center, and the first 5 principle components derived from genomic data. The adjusted phenotypes consisted of least-squares residuals from a model that included the effects listed above. For each trait, sample sizes and within-sex summary statistics are provided in S1 Table.

### 2.6.3 LBR hyperparameters

Hyperparameters used in the LBR model (eq. 1) were error variances for each sex, the proportion of nonzero effects for each SNP effect component, and the variances of nonzero effects for each SNP effect component $\left\{\sigma^2_{\varepsilon_\mathrm{m}}, \sigma^2_{\varepsilon_\mathrm{f}}, \pi_0, \pi_\mathrm{m}, \pi_\mathrm{f}, \sigma^2_{b_0}, \sigma^2_{b_\mathrm{m}}, \sigma^2_{b_\mathrm{f}}\right\}$. Variances (of either SNP effect components or sex-specific errors) were given a scaled-inverse Chi-square prior, parameterized by a degree of freedom parameter $df$ (set to 5) and scaling parameter $S$. $S$ is set according to built-in rules of the BGLR package using a prior model R-squared of 0.03 for main effects and 0.01 for the sex-interaction terms. More detail on how the scale parameter $S$ is calculated can be found in Perez and de los Campos, 2014 [52]. $\pi_k$ was given a beta prior with shape parameters $\alpha = 2$ and $\beta = 2$. An example of how to implement LBR (eq. 2.1) using BGLR with the above hyperparameter specifications is provided at `https://github.com/funkhou9/LBR-sex-interactions`.

### 2.6.4 Inference using post-processing of posterior samples

BGLR uses Markov chain Monte Carlo (MCMC) to sample from the posterior distribution of sex-specific effects. For each MCMC sample we derived male and female effects using $\beta_{\mathrm{m}_{j(s)}} = b0_{j(s)} + b_{\mathrm{m}_{j(s)}}$ and $\beta_{\mathrm{f}_{j(s)}} = b0_{j(s)} + b_{\mathrm{f}_{j(s)}}$, where $s = 1, \ldots, 4,350$ indexes MCMC samples. Here, results were obtained using three separate MCMC chains. Each chain was obtained using 3,400 MCMC samples; the first 500 samples were discarded as burn-in and the remaining samples where thinned by an interval of 2, leading to 1,450 samples per chain.

Estimates of sex-specific SNP effects $\left(\hat{\beta}_{\mathrm{m}_j}$ and $\hat{\beta}_{\mathrm{f}_j}\right)$ were obtained from their posterior means. We estimated the posterior probability of a female-specific non-zero SNP effect using $\mathrm{PPF}_{\mathrm{SNP}_j} = \max\left[\Pr\left(\beta_{\mathrm{f}_j} > 0 \mid \mathcal{D}\right), \Pr\left(\beta_{\mathrm{f}_j} < 0 \mid \mathcal{D}\right)\right]$, where $\mathcal{D}$ represents the observed data. This was done by counting the proportion of $\beta_{\mathrm{f}_j}$ samples above zero and below zero. This was repeated for inferring the male-specific SNP effect. The posterior probability of sex-difference at individual SNP-effects was estimated using $\mathrm{PPDiff}_{\mathrm{SNP}_j} = \max\left[\Pr\left(\beta_{\mathrm{m}_j} > \beta_{\mathrm{f}_j} \mid \mathcal{D}\right), \Pr\left(\beta_{\mathrm{m}_j} < \beta_{\mathrm{f}_j} \mid \mathcal{D}\right)\right]$ where again these probabilities were estimated using the corresponding frequencies from the posterior distribution samples.

For each MCMC sample we also aggregated SNP effects within window $j^*$ using $\boldsymbol{u}_{\mathrm{m}_{j^*(s)}} = \boldsymbol{X}_{j^*}\boldsymbol{\beta}_{\mathrm{m}_{j^*(s)}}$ and $\boldsymbol{u}_{\mathrm{f}_{j^*(s)}} = \boldsymbol{X}_{j^*}\boldsymbol{\beta}_{\mathrm{f}_{j^*(s)}}$. For this calculation we used a common genotype matrix $\boldsymbol{X}_{j^*}$ consisting of all $N$ male and female genotypes to avoid differences in additive genetic values arising from allele frequency differences between males and females occurring by random sampling. Samples of sex-specific window variances were obtained using the sample variance: $\sigma^2_{\mathrm{m}_{j^*(s)}} = (N-1)^{-1}\sum_{i=1}^{N}\left(u_{\mathrm{m}_{i_{j^*(s)}}} - \bar{u}_{\mathrm{m}_{j^*(s)}}\right)^2$ and $\sigma^2_{\mathrm{f}_{j^*(s)}} = (N-1)^{-1}\sum_{i=1}^{N}\left(u_{\mathrm{f}_{i_{j^*(s)}}} - \bar{u}_{\mathrm{f}_{j^*(s)}}\right)^2$. Estimates of sex-specific window variances were obtained from their posterior means. Inferring sex-specific window variances was done by estimating $\mathrm{PPM}_{\sigma^2_{g_{j^*}}} = \Pr\left(\sigma^2_{g\mathrm{m}_{j^*}} > 0 \mid \mathcal{D}\right)$ and $\mathrm{PPF}_{\sigma^2_{g_{j^*}}} = \Pr\left(\sigma^2_{g\mathrm{f}_{j^*}} > 0 \mid \mathcal{D}\right)$ and inferring a G×S interaction at window j* was done by estimating:

$$\mathrm{PPDiff}_{\sigma^2_{g_{j^*}}} = \max\left[\Pr\left(\sigma^2_{g\mathrm{m}_{j^*}} - \sigma^2_{g\mathrm{f}_{j^*}} > t_{j^*} \mid \mathcal{D}\right), \Pr\left(\sigma^2_{g\mathrm{m}_{j^*}} - \sigma^2_{g\mathrm{f}_{j^*}} < t_{j^*} \mid \mathcal{D}\right)\right],$$

where $t_{j^*}$ was used to exert judgment about how different sex-specific window variances must be to declare a meaningful G×S interaction. Here, $t_{j^*}$ was one-tenth of the mean of all posterior samples of $\sigma^2_{g\mathrm{m}_{j^*}}$ and $\sigma^2_{g\mathrm{f}_{j^*}}$. Functions to process posterior samples to estimate and infer non-null sex-specific effects and G×S interactions is provided at `https://github.com/funkhou9/LBR-sex-interactions`.

### 2.6.5 Defining local, LD-based windows

To define SNPs contained within window $j^*$, a region of LD centered on SNP $j$, we collected all SNP $j'$ immediately surrounding SNP $j$ for which $cor(x_j, x_{j'})^2 \geq 0.1$. We allowed up to two consecutive SNPs in which $cor(x_j, x_{j'})^2 < 0.1$ to allow for potential mapping errors or other unexplained instances where LD with SNP $j$ dips only briefly. The function `getWindows()`, which provides windows given a genotype matrix $\boldsymbol{X}$, is provided in `https://github.com/funkhou9/LBR-sex-interactions`.

### 2.6.6 Single marker regression

We also performed single-marker regression analyses using following model:

$$\begin{bmatrix} \boldsymbol{y}_{\mathrm{m}} \\ \boldsymbol{y}_{\mathrm{f}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}\mu_{\mathrm{m}} \\ \mathbf{1}\mu_{\mathrm{f}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{x}_{\mathrm{m}_j} \\ \boldsymbol{x}_{\mathrm{f}_j} \end{bmatrix} \beta_j + \begin{bmatrix} \boldsymbol{x}_{\mathrm{m}_j} \\ \mathbf{0} \end{bmatrix} \beta_{\mathrm{G\times S}} + \begin{bmatrix} \boldsymbol{\varepsilon}_{\mathrm{m}} \\ \boldsymbol{\varepsilon}_{\mathrm{f}} \end{bmatrix}. \tag{2.2}$$

As with the LBR model (equation 2.1), we assume sex-specific errors are distributed normally with zero mean and sex-specific variances. SNP effects and interactions were estimated using weighted least squares. To test for a G×S interaction at SNP $j$, a t-test is used: $\hat{\beta}_{j_{\mathrm{G\times S}}} / SE\left(\hat{\beta}_{j_{\mathrm{G\times S}}}\right) \sim t_{N-3}$. The $p$-value from such a test is referred to as *pvalue*-diff. To test for any association (either among males, females, or both), we used an F-test, comparing a restricted model:

$$\begin{bmatrix} \boldsymbol{y}_{\mathrm{m}} \\ \boldsymbol{y}_{\mathrm{f}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}\mu_{\mathrm{m}} \\ \mathbf{1}\mu_{\mathrm{f}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{\mathrm{m}} \\ \boldsymbol{\varepsilon}_{\mathrm{f}} \end{bmatrix}$$

against the unrestricted model in equation 2.2.

### 2.6.7 Simulations

Simulated traits were developed using 60,000 genotyped SNPs (the first 6,000 SNPs from the first ten chromosomes) from 119,190 males and 139,738 females. Using these SNP genotypes, each trait was simulated as follows:

1. A total of 150 causal variants (CVs) were randomly sampled from 60,000 SNPs.

   Let $\boldsymbol{Z}_{\mathrm{m}} = \left\{z_{\mathrm{m}_{ik}}\right\}_{i=1,k=1}^{N_{\mathrm{m}}=119{,}190, q=150}$ and $\boldsymbol{Z}_{\mathrm{f}} = \left\{z_{\mathrm{f}_{ik}}\right\}_{i=1,k=1}^{N_{\mathrm{f}}=139{,}738, q=150}$ denote matrices of male and female genotypes at sampled CVs.

2. Additive CV effect sizes were randomly sampled from the gamma distribution. 90 CVs (those with homogenous effects) were sampled from $Gamma(k = 10, \theta = 1)$ and were made negative with a probability of 0.5. Of the 60 CVs with differing sex-specific effects, 30 had nonzero effects in both sexes but with deferring magnitudes: at random one sex's effects

were sampled from $Gamma(k = 5, \theta = 1)$ and the other from $Gamma(k = 20, \theta = 1)$. For the remaining 30 CVs, at random one sex's effects were exactly zero while the other sex's effects were sampled from $Gamma(k = 10, \theta = 1)$.

Let $\boldsymbol{\gamma}_{\mathrm{m}} = \left\{\gamma_{\mathrm{m}_k}\right\}_{k=1}^{q=150}$ and $\boldsymbol{\gamma}_{\mathrm{f}} = \left\{\gamma_{\mathrm{f}_k}\right\}_{k=1}^{q=150}$ denote vectors of male-specific and female-specific CV effects, respectively, for all 150 CVs.

3. Error variances for males $\sigma^2_{\delta_{\mathrm{m}}}$ and females $\sigma^2_{\delta_{\mathrm{f}}}$ were adjusted such that the proportion of phenotypic variance explained by all QTL is 0.05 for both males and females (on the complete genome scale this corresponds to a heritability of about 0.5).

   Let $\delta_{\mathrm{m}_i} \sim N\left(0, \sigma^2_{\delta_{\mathrm{m}}}\right)$ and $\delta_{\mathrm{f}_i} \sim N\left(0, \sigma^2_{\delta_{\mathrm{f}}}\right)$ denote residual error for the $i^{\text{th}}$ male and $i^{\text{th}}$ female.

4. Male traits $\boldsymbol{\phi}_{\mathrm{m}} = \left\{\phi_{\mathrm{m}_i}\right\}_{i=1}^{N_{\mathrm{m}}=119{,}190}$ and female traits $\boldsymbol{\phi}_{\mathrm{f}} = \left\{\phi_{\mathrm{f}_i}\right\}_{i=1}^{N_{\mathrm{f}}=139{,}738}$ were simulated from a linear combination of QTL genotypes plus a residual error: $\boldsymbol{\phi}_{\mathrm{m}} = \boldsymbol{Z}_{\mathrm{m}}\boldsymbol{\gamma}_{\mathrm{m}} + \boldsymbol{\delta}_{\mathrm{m}}$ and $\boldsymbol{\phi}_{\mathrm{f}} = \boldsymbol{Z}_{\mathrm{f}}\boldsymbol{\gamma}_{\mathrm{f}} + \boldsymbol{\delta}_{\mathrm{f}}$

5. Steps 1-4 are repeated for 30 Monte Carlo replicates.

## 2.7 Acknowledgments

Enrichment analysis performed in this manuscript was done using data from the Genotype-Tissue Expression (GTEx) Project. Single-tissue cis-eQTL data was downloaded from `https://gtexportal.org/home/datasetson02/01/19`.

# CHAPTER 3

# EVIDENCE FOR TRANSCRIPTOME-WIDE RNA EDITING AMONG SUS SCROFA PRE-1 SINE ELEMENTS

This chapter has been published previously [62]. The manuscript was prepared alongside co-authors Juan P Steibel, Ronald O Bates, Nancy E Raney, Darius Schenk, and Catherine W Ernst.

## 3.1 Abstract

RNA editing by ADAR (adenosine deaminase acting on RNA) proteins is a form of transcriptional regulation that is widespread among humans and other primates. Based on high-throughput scans used to identify putative RNA editing sites, ADAR appears to catalyze a substantial number of adenosine to inosine transitions within repetitive regions of the primate transcriptome, thereby dramatically enhancing genetic variation beyond what is encoded in the genome. Here, we demonstrate the editing potential of the pig transcriptome by utilizing DNA and RNA sequence data from the same pig. We identified a total of 8550 mismatches between DNA and RNA sequences across three tissues, with 75% of these exhibiting an A-to-G (DNA to RNA) discrepancy, indicative of a canonical ADAR-catalyzed RNA editing event. When we consider only mismatches within repetitive regions of the genome, the A-to-G percentage increases to 94%, with the majority of these located within the swine specific SINE retrotransposon PRE-1. We also observe evidence of A-to-G editing within coding regions that were previously verified in primates. Thus, our high-throughput evidence suggests that pervasive RNA editing by ADAR can exist outside of the primate lineage to dramatically enhance genetic variation in pigs.

## 3.2 Background

Eukaryotes are known for relatively complex mechanisms used to regulate gene expression. One such mechanism, RNA editing, enables the cell to alter sequences of RNA transcripts [30] such that they are no longer forced to match the "hard-wired" genome sequence. High throughput

methods for studying targets of this mechanism transcriptome-wide have been applied to primate studies, where evidence for massive amounts of ADAR (adenosine deaminase acting on RNA) catalyzed A-to-I RNA editing has been discovered, preferentially within SINE retrotransposons such as the primate Alu [32, 63, 64, 65, 66, 67, 68]. Such work has yet to be performed with pig transcriptomes using the latest sequencing technology. Although little is known about pig SINE elements compared to those in primates, key features of the pig-specific PRE-1 retrotransposon make pigs an intriguing model to further elucidate transcriptome-wide patterns of ADAR targets.

ADAR can only catalyze A-to-I editing within dsRNA. The high editibility of the primate specific Alu element is attributed to its capacity to induce dsRNA; these elements have a high copy number, are short, relatively undiverged from one another, and tend to cluster in gene rich regions of the genome [69]. When appearing as tandem and inverted pairs within the same transcribed region, these properties facilitate intra-molecular dsRNA formation that serve as ADAR targets [32, 70]. Comparatively, the pig PRE-1 element possesses many of these same properties that are believed to contribute to dsRNA formation within the transcriptome. Notably, PRE-1 has the 3rd highest copy number of any SINE cataloged on SINEBase [71].

Since Alu elements are generally found within and near genes, ADAR editing in humans preferentially targets non-coding regions of many genes such as introns, UTRs and upstream and downstream gene proximal regions. ADAR editing of these regions is thought to be a key component of RNA processing via mechanisms that include Alu exonization [72] and RNAi pathway alteration [73]. By demonstrating that RNA editing in pigs generally targets SINE elements within non-coding regions of genes, this would suggest that RNA processing by way of ADAR editing of SINE elements predated the emergence of primate and pig-specific retrotransposons. Rarely, ADAR editing occurs within coding regions to alter amino acid sequences [74]. This type of editing is particularly mysterious in that its pattern is less traceable than non-coding editing, but is nevertheless site-specific and required for the function of essential protein coding genes such as *GluR-B* in mice [33]. Therefore, in addition to the regulation of transcripts by way of editing non-coding SINE elements, editing of coding regions is an essential form of transcriptional regulation in mice, with

the extent of its conservation across Mammalia yet to be fully determined.

Here, we demonstrate the pig's capacity for RNA editing. By studying this process in a relatively distant species to human with a distinct repetitive element repertoire, we want to determine if RNA editing patterns seen in Alu bearing genomes can likewise be observed in pigs. RNA editing detection was done by analyzing a single pig using whole genome sequencing data and RNA sequencing data from liver, subcutaneous fat, and *longissimus dorsi* muscle. Based on previous studies done in primates, a bioinformatic strategy was used to find A-to-I (observed as A-to-G) DNA to RNA mismatches that give evidence of ADAR catalyzed RNA editing events.

## 3.3 Results and discussion

### 3.3.1 DNA and RNA sequencing

To provide the materials needed for a transcriptome-wide survey of RNA editing candidates, genomic DNA as well as total RNA from liver, subcutaneous fat, and *longissimus dorsi* (LD) muscle were purified from samples obtained from a single animal, similar to another single-animal editome study [68]. Sequencing was done using the Illumina HiSeq 2500 to generate 150x2 paired end reads from genomic DNA, with PolyA RNA sequencing used to generate cDNA reads in the same format. Roughly 250M pass-filter genomic DNA reads were generated with an average overall alignment rate of 89% to the *Sus scrofa* reference genome sequence (*Sus scrofa* 10.2.69). An average of 106M pass-filter strand specific cDNA reads were obtained from each tissue, with an average overall alignment rate of 76%.

### 3.3.2 Identification of candidate RNA editing events

To scan the transcriptome for possible RNA editing sites, we utilized a custom pipeline influenced by previous studies done in human cell lines and primates [75, 68]. Prior to alignment, in order to avoid utilizing bases with relatively poor base qualities at the ends of reads, raw genomic DNA and cDNA sequencing reads were trimmed for base quality at their 3' ends before aligning to the *Sus scrofa* 10.2.69 reference genome. Additional trimming 6bp from the 5' ends of cDNA reads

was done to prevent misidentification of DNA RNA mismatches due to artifacts associated with the use of random hexamers during cDNA library preparation [76]. When conducting a search for RNA editing candidates with RNA-seq, strand-specific RNA-seq libraries can be utilized to account for the strandedness of each transcript, thereby enabling A-to-G DNA-to-RNA mismatches to be distinguished from T-to-C DNA-to-RNA mismatches. In order to utilize our strand-specific cDNA alignments for variant calling while preserving the strandedness of each alignment to distinguish A-to-G from T-to-C mismatches, plus-strand alignments were separated from minus-strand alignments for each cDNA sample. From all genomic DNA and cDNA alignments, we extracted those reads that had only 1 recorded alignment in order to optimize our chances that genomic DNA and cDNA reads arising from the same locus map to the same location. Joint variant calling using SAMTools [77] was performed, combining genomic DNA alignments with cDNA plus-strand alignments from each tissue. This was repeated for all cDNA minus-strand alignments. Both resulting VCF files were analyzed using editTools, an in-house R package made to efficiently scan VCF files for DNA RNA mismatches using C++ source code. editTools was developed to implement RNA editing detection within the R framework and to provide visualization tools; editTools was used to generate all figures in this manuscript pertaining to sequencing data. Default editTools parameters were used, in which a mismatch was considered a candidate RNA editing site if at a particular locus 1) the genotype is homozygous according to 95% of the DNA reads, 2) at least 10 reads were used to determine the genotype, 3) neither genomic DNA nor cDNA samples are indels, 4) at least 5 cDNA reads from the same tissue differ from the genotype call, and 5) these cDNA reads must have a Phred-scaled strand-bias P-value of 20 or less. Specific thresholds for DNA and cDNA sequencing depths were determined according to a previous study that profiled the rhesus macaque editome from a single animal [68]. Using this approach, we identified a total of 6410 A-to-G mismatch events representing 75% of all mismatches found (8550 total mismatches; Fig 3.1). When we restrict our search to known swine repetitive sequences, 5993 out of 6410 A-to-G mismatches are retained, representing 93.8% of all mismatches in repetitive regions. Of the remaining mismatches in repetitive regions, 4.1% are T-to-C. It is not surprising that T-to-C mismatches are the second most

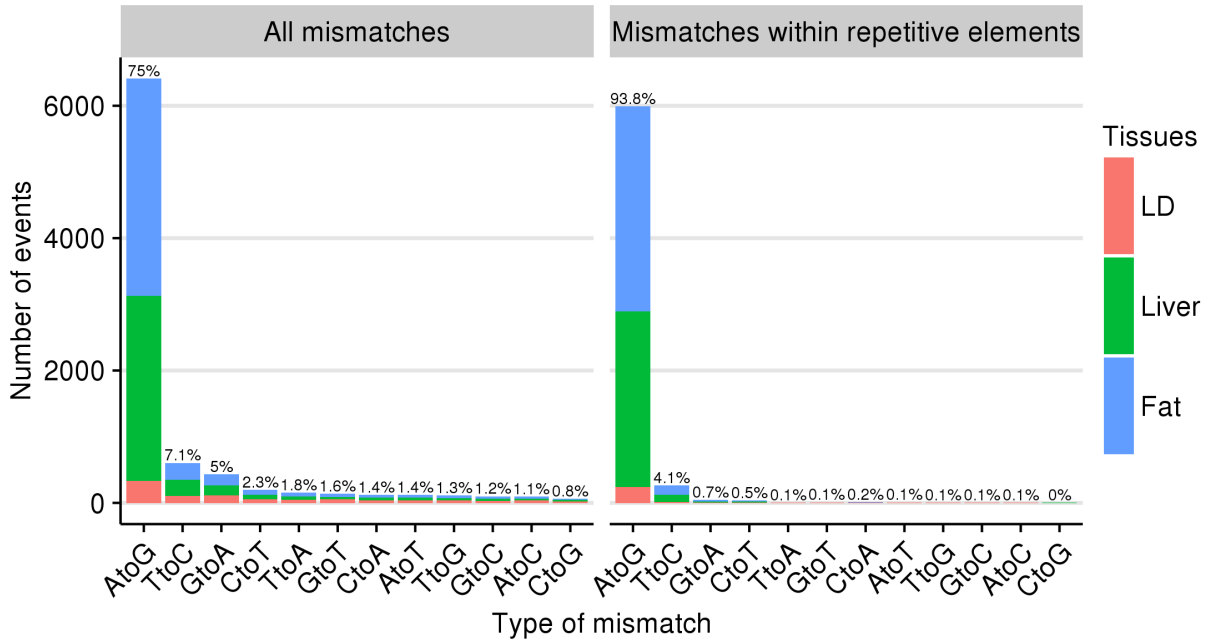


Figure 3.1: DNA to RNA mismatch counts
Comparing all mismatches found transcriptome wide (Left) to those within the body of a repetitive element (Right). Percentages shown are out of all mismatches found in each category.

common since T-to-C artifacts could arise if at a true A-to-G editing site, plus-strand alignments were incorrectly identified as minus-strand alignments or vice versa. Note that our observation of 8550 A-to-G mismatches is intended to be a conservative estimate of the total number of ADAR-catalyzed editing sites in these three tissues, primarily because we have restricted our search to homozygous sites; at heterozygous sites, it is not feasible to directly determine which allele is being edited, or if editing is truly occurring at either allele.

### 3.3.3 Tissue differences

To understand differences in candidate RNA editing sites between tissues, canonical A-to-G mismatches were aligned across tissues if they were detected at the same physical position and on the same strand. The number of candidate RNA editing events was fewer in LD compared to liver or fat (Fig 3.1), consistent with lower RNA editing activity in muscle compared to other tissues for rhesus macaque [68]. Despite candidate RNA editing sites showing strong tissue specificity, a total

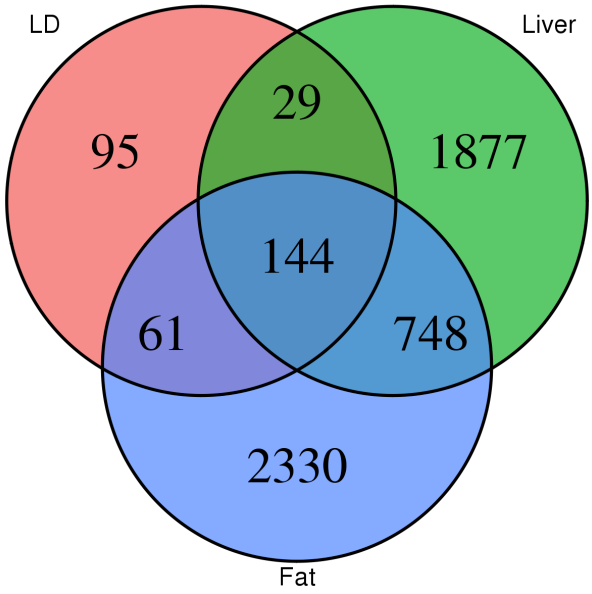


Figure 3.2: Shared A-to-G mismatches between tissues
A mismatch between two or more tissues was considered shared if it occurred at the same physical position and on the same strand.

of 144 A-to-G mismatches were found to be common among all three tissues, whereas 748 were found to be common between liver and fat (Fig 3.2).

One factor that may contribute to tissue specificity of RNA editing is differential expression of ADAR [78]. Using RNA samples from 33 additional pigs, a quantitative real-time PCR assay was used to infer ADAR transcript abundance differences between liver, subcutaneous fat, and LD muscle (Fig 3.3). Average ADAR expression was determined to be significantly lower in LD muscle tissue than in either fat ($p < 0.0003$) or liver ($p < 0.00001$) tissues, suggesting that differential ADAR expression may contribute to differences in candidate RNA editing sites between tissues.

### 3.3.4 Controlling for errors due to mapping quality

After imposing such strict restrictions as excluding genomic DNA and cDNA reads that had more than one recorded alignment and trimming the ends of reads pre-alignment, we wanted to assess how well such measures protect against mapping errors, which are among the leading causes of RNA editing misidentification when using short reads [76, 79]. Mapping quality is a measurement

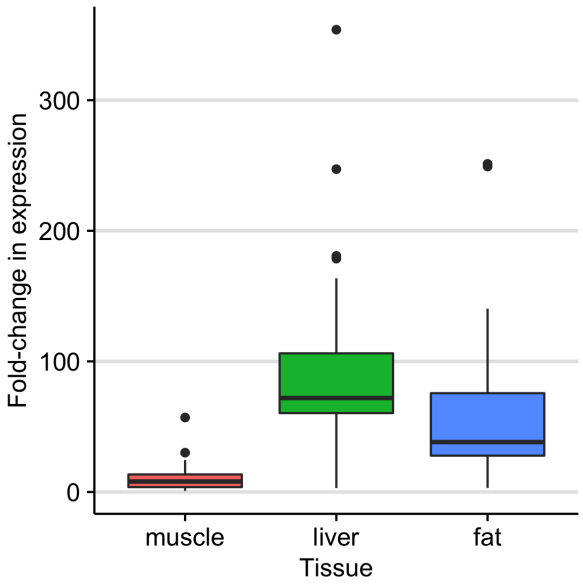


Figure 3.3: Relative ADAR transcript abundance between tissues
Expression was measured relative to the LD muscle sample used for sequencing. Using a one-way ANOVA, a significant effect of tissue on ADAR expression was detected ($p < 0.0001$). Pairwise comparisons of tissue means using Tukey HSD shows significant differences in ADAR expression between LD and liver ($p < 0.00001$) and between LD and fat ($p < 0.003$), but no significant difference between fat and liver ($p = 0.0505563$).

that provides a probability that a read is misaligned, given its number of possible alignments and sum of base qualities for each alignment [80]. Knowing this, and under the assumption of no RNA editing, for each mismatch locus $i$ we computed the probability of observing at least 5 "edited" reads given the cDNA sequencing depth $N_i$ and average sample mapping quality $MQ_i$. Among all 8550 repetitive and non-repetitive mismatch positions, the maximal probability of observing at least 5 "edited" reads was $\sim 6.772 \times 10^{-15}$ for a site with $N = 13$ and average $MQ = 29$. If Bonferroni correction is used then 0.05 / 189,638 = $6.23\text{x}10^{-7}$ can be used as a threshold for transcriptome-wide significance, where 189,638 was the total number of queried cDNA positions with a sequencing depth of at least 5 cDNA reads that were at the location of homozygous loci in the genomic sequence. From this evidence we conclude that our pipeline sufficiently minimizes artifacts associated with mapping quality when using the *Sus scrofa* 10.2.69 assembly.

### 3.3.5 Pig editome functional implications

Little is known about the average effect of RNA editing transcriptome wide. For humans, one prevailing hypothesis is that the exonization of Alu SINE elements is controlled in part by A-to-G editing. An instance of this mechanism has been demonstrated, where intronic A-to-G editing events contribute to alternative splicing of *nuclear prelamin A* so that an Alu element is included in an exon [72]. To explore the possibility that RNA editing in pigs targets introns to affect splicing, editTools was used to synthesize mismatch data with Variant Effect Predictor data to find the relative locations of each mismatch relative to annotated transcripts. Consistent with what has been found in humans [32], nearly half of all detected A-to-G mismatches are located in retained introns (Fig 3.4). The remaining sites are concentrated in other non-coding regions including 3' UTRs, intergenic, and gene proximal regions. While the majority of non-coding editing events in humans are attributed to the position and orientation of SINE elements within transcripts [70], coding RNA editing occurs rarely, usually outside repetitive elements but nevertheless site-specifically. It has been suggested that site-specificity of coding RNA editing events is facilitated by nearby SINE elements, which through their induction of long dsRNA regions, recruit ADAR in sufficient density to affect coding regions in close proximity [75]. From our data, only 49 pig A-to-G mismatches were found within coding regions and of those, 34 would result in a missense variant (Table 3.1). It can be noted that a number of amino acid changes resulting from verified macaque DNA RNA mismatches [68] can be found among our pig dataset – mismatches that control I/V in COPA, Y/C in BLCAP, I/V in COG3, K/R in NEIL1, and Q/R in GRIA2. Interestingly, Y/C recoding of BLCAP via RNA editing has been associated with hepatocellular carcinoma (HCC) in humans as HCC samples were shown to express edited BLCAP in significantly higher amounts than non-HCC samples [81]. Additionally, exon 6 K/R recoding of NEIL1 by RNA editing was previously thought to be primate specific and attributed to the K/R site's proximity to Alu dense regions [82], however we witness evidence of the same K/R recoding of exon 6 via an A-to-G editing event in pigs. If in fact SINE elements recruit ADAR to affect nearby coding regions, then our data suggest the remarkable conservation of NEIL1 K/R recoding across genomes with entirely different SINE

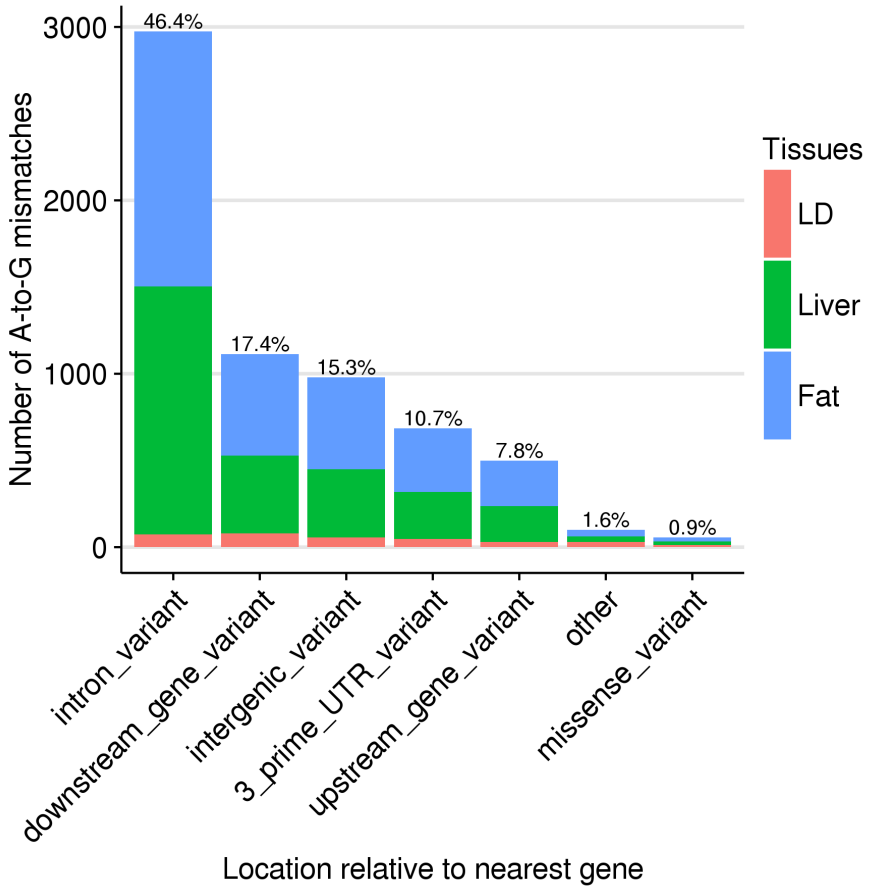


Figure 3.4: A-to-G mismatch locations relative to the nearest annotated genes
Percentages shown are out of all A-to-G mismatches.

elements.

### 3.3.6 Pig editome association with pig-specific SINE elements

Since properties of the primate Alu element are suggested to influence RNA editing in both coding and non-coding regions, one of our primary interests was to determine which SINE elements in pigs are capable of attracting the majority of ADAR activity. Again using the functionality of editTools, we merged our mismatch data with data from RepeatMasker to determine which repetitive regions contain putative RNA editing sites. As mentioned previously, 5993 out of 6410 A-to-G mismatches are located within the body of a repetitive element. Upon closer inspection, 5715 of the 5993 are within pig SINE elements as opposed to LINE elements and others (Fig 3.5A), although SINEs

Table 3.1: A-to-G mismatches resulting in amino acid changes

| Position | Gene symbol/ID | AA | SIFT | Tissues |
|---|---|---|---|---|
| 1:63408856 | ENSSSCG00000029003 | L/P | tolerated(1) | Fat LD Liver |
| 1:125424444 | ENSSSCG00000024660 | Q/R | tolerated(1) | Fat LD Liver |
| 2:12622576 | LDHB | I/M | tolerated(1) | Fat LD Liver |
| 2:49316285 | ARNTL | K/E | tolerated low confidence(1) | Liver |
| 4:98044799 | COPA | I/V | deleterious(0.02) | Fat |
| 5:42375023 | KRR1 | I/T | deleterious(0.01) | Liver |
| 6:92516721 | PTPRM | K/R | tolerated(1) | Fat |
| 6:146168578 | NDC1 | E/G | deleterious(0.01) | Liver |
| 7:62951442 | NEIL1 | K/R | deleterious(0.02) | Fat LD |
| 7:81602273 | ENSSSCG00000002045 | C/R | tolerated(1) | Fat LD Liver |
| 7:102789222 | ACOT4 | T/A | tolerated(0.61) | Fat |
| 7:129322238 | RPS21 | C/R | - | Fat LD Liver |
| 8:28015971 | ENSSSCG00000008767 | H/R | tolerated(1) | Fat LD Liver |
| 8:31629014 | TLR1 | I/V | tolerated(1) | Liver |
| 8:32309809 | RPL9 | I/V | tolerated(0.4) | Fat |
| 8:32309814 | RPL9 | E/G | deleterious(0.01) | Fat |
| 8:48244993 | GRIA2 | Q/R | tolerated(0.07) | Fat |
| 9:41146365 | ENSSSCG00000023913 | Q/R | deleterious(0.04) | Fat |
| 9:74510703 | ENSSSCG00000015294 | K/R | tolerated(0.13) | Liver |
| 9:83273454 | SLC25A13 | E/G | deleterious(0.02) | LD |
| 11:22178068 | COG3 | I/V | tolerated(1) | Fat LD Liver |
| 12:20231860 | AOC3 | Q/R | tolerated(1) | Liver |
| 13:131377159 | EIF2B5 | Q/R | tolerated(1) | Fat |
| 13:156760971 | UBE2B | D/G | tolerated(0.48) | Fat LD Liver |
| 13:206979572 | SON | R/G | - | Fat |
| 14:40832826 | PLBD2 | R/G | tolerated low confidence(0.12) | Fat |
| 14:52398588 | IGLV-3 | E/G | tolerated(0.05) | Fat |
| 14:59613334 | LYST | S/G | - | LD |
| 14:81796679 | OIT3 | S/G | tolerated(1) | Liver |
| 15:59811585 | HNRNPA2B1 | L/P | tolerated(0.35) | Fat LD Liver |
| 15:98217885 | ENSSSCG00000028949 | R/G | tolerated low confidence(1) | Fat LD Liver |
| 16:29335640 | ENSSSCG00000016869 | N/D | tolerated(1) | Fat LD |
| 16:42512978 | ELOVL7 | S/G | tolerated(1) | Fat |
| 17:46041505 | BLCAP | Y/C | deleterious(0) | Fat Liver |

occupy just 11.4% of the swine genome, while LINEs occupy 17.5% [83]. Of the 5993 repetitive A-to-G mismatches, 58.8% are found within the Pre0_SS element, a SINE element of the PRE1 family (Fig 3.2B). Little is known about Pre0_SS, but among all elements of the PRE1 family, Pre0_SS is most identical to the consensus PRE1 sequence. In many instances, Pre0_SS elements are > 99% identical to one another, indicating that it is currently actively transposing in pigs [84]. Additional members of the PRE1 family contain A-to-G mismatches, although at a much lower

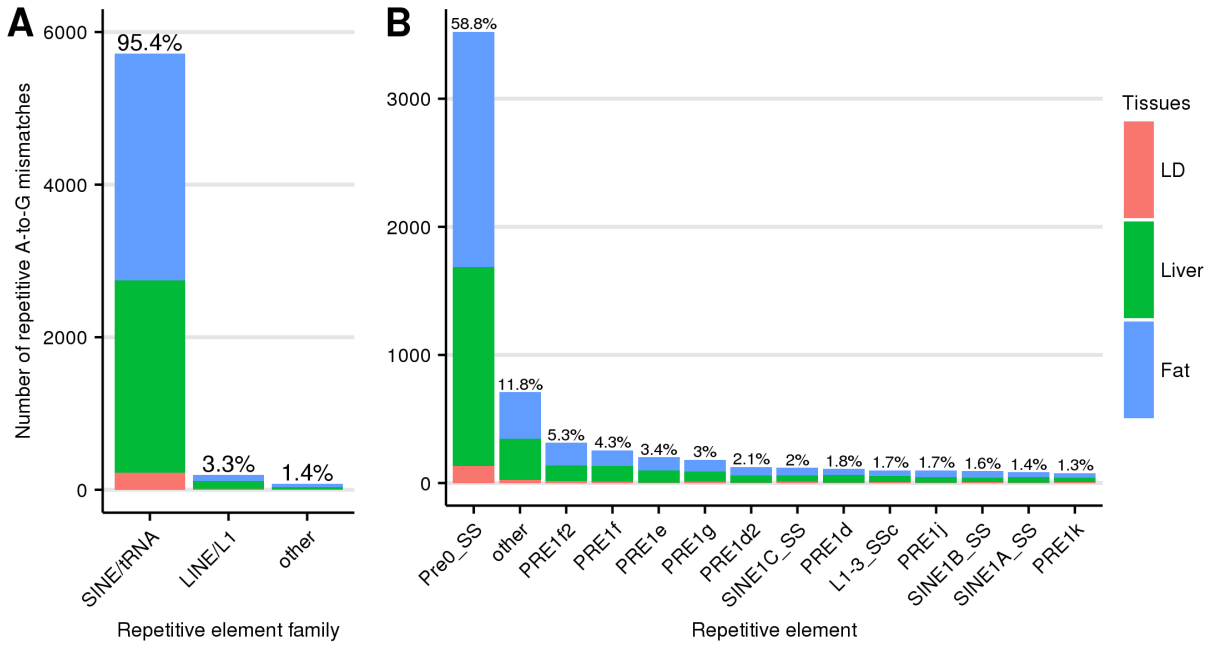


Figure 3.5: Distribution of repetitive A-to-G mismatches
The distribution is shown across major repetitive element families (A) and further broken down into specific repetitive element types (B). Percentages shown are out of all repetitive A-to-G mismatches.

frequency than Pre0_SS.

## 3.4 Conclusions

While Alu elements enable substantial RNA editing among primate genomes, we show that non-Alu bearing genomes can also utilize RNA editing as a means to achieve a similar result. Our high-throughput scan suggests that pig transcriptomes are highly editable among PRE-1 SINE retrotransposons. PRE-1, an element derived from an ancestral tRNA, has similar features to the primate Alu, derived from an ancestral 7SL RNA; a copy number of $1x10^6$, consensus length of 246bp, and very little diversity among such members as Pre0_SS. These features influence the secondary structure of the transcriptome, which in turn affect ADAR editable targets. Surprisingly, conservation of specific editing sites such as those in *NEIL1* and *BLCAP* appears evident between human and pigs. Therefore, we hypothesize that transcriptome secondary structure may be conserved among mammals enough to preserve particular RNA editing sites, and that SINE

elements, regardless of origin, may conform to certain positions and orientations in order to allow conservation to occur.

By demonstrating that pig transcriptomes have potential to be highly edited, we propose that pigs may be a valuable model to understand the patterns of ADAR controlled RNA editing. Additionally, by shedding light on the pig editome, we can begin to understand the extent to which this phenomenon enhances pig genetic variation. Such sources of variation may one day provide valuable explanatory power for a variety of traits of interest to both biomedical and agricultural communities.

## 3.5 Methods

### 3.5.1 Sequence data

From Michigan State University's pig resource population (MSUPRP), an $F_2$ population resulting from crosses between 4 $F_0$ Duroc sires and 15 $F_0$ Pietrain dams [85], a single female animal was chosen for whole genome and transcriptome sequencing. Total RNA was extracted from subcutaneous fat, liver, and LD skeletal muscle using TRIzol, and a RIN greater than 7 was determined with the Agilent 2100 Bioanalyzer. cDNA libraries were made using the Illumina TruSeq Stranded mRNA Library Preparation Kit. Sequencing was performed using the Illumina HiSeq 2500 in Rapid Run mode with 150x2 paired-end reads. Base calling was done by Illumina's Real Time Analysis v1.18.61 and the output was converted to FastQ format with Illumina's Bcl2fastq v1.8.4. Genomic DNA was purified from white blood cells using the Invitrogen Purelink Genomic DNA Mini Kit and libraries were made using the Illumina TruSeq Nano DNA Library Preparation Kit HT. Sequencing of genomic DNA was done using the Illumina HiSeq 2500 in Rapid Run mode with 150x2 paired-end reads. Real Time Analysis v.1.17.21.3 and Bcl2fastq v1.8.4 were used for base calling and FastQ conversion, respectively. Read quality of both whole genome and RNA data was assessed using the FastQC program [86].

### 3.5.2 Sequence preparation and mapping

DNA reads from whole genome sequencing were trimmed for quality at the 3' end using Condetri v2.2 [87] with parameters: `-sc=33 -minlen=75` and `b=fq`. Resulting mate 1, mate 2 and unpaired reads were mapped to *Sus Scrofa* 10.2.69 using Bowtie v2.2.1 [88] with parameters: `-p 7 -X 1000`. In order to filter out DNA reads that had more than one recorded alignment, alignments containing the `XS:i:<N>` tag, where `N` indicates the number of alternative alignments for a read, were removed. Strand specific cDNA sequencing reads from each tissue sample were trimmed with Condetri with parameters: `-sc=33 -minlen=75 -pb=fq -cutfirst=6 -pb=fq`. Resulting paired and unpaired cDNA reads were then mapped to *Sus Scrofa* 10.2.69 using TopHat v2.0.12 [89] with parameters: `-p 7 -mate-inner-dist 400 -mate-std-dev 100 -library-type "fr-firststrand"`. Filtering out cDNA reads that had more than one recorded alignment was done by selecting alignments with the `NH:i:1` tag, while separating plus strand transcript alignments from minus strand alignments was done by selecting alignments possessing the `XS:A:+` or `XS:A:-` tags, respectively. The resulting DNA and cDNA alignments are the "filtered" data used in downstream variant calling and mismatch detection.

### 3.5.3 Variant calling and mismatch detection

We utilized variant calling software Samtools v1.0 and Bcftools v1.2 to jointly call variants among DNA and cDNA reads from plus strand transcripts using:

```
samtools mpileup -f <reference_genome.fa> -C50 -E -Q25 -ug -t DP,DV,SP <
    ↪ DNA.bam> <liver_plusstrand.bam> <fat_plusstrand.bam> <LD_plusstrand.
    ↪ bam>
```

where `<DNA.bam>` includes all filtered DNA alignments, and

```
<liver_plusstrand.bam>, <fat_plusstrand.bam>, <LD_plusstrand.bam>
```

are filtered cDNA alignments from plus strand transcripts. Likewise, DNA and cDNA reads from minus strand transcripts were processed similarly with:

```
samtools mpileup -f <reference\_genome.fa> -C50 -E -Q25 -ug -t DP,DV,SP <
    ↪ DNA.bam> <liver_minusstrand.bam> <fat_minusstrand.bam> <
    ↪ LD_minusstrand.bam>.
```

Note that the parameter `-t DP,DV,SP` is required for downstream mismatch detection with editTools. Samtools output from each command was piped into bcftools with additional parameters: `-O v -m -v`. These steps produce two VCF files that are simultaneously processed with `find_edits()`, a function within editTools available at `https://github.com/funkhou9/editTools`. By default, `find_edits()` scans each variant site to search for candidate RNA editing sites according to the five criteria required for sufficient evidence (see Results and Discussion). Most figures in this report were generated using editTools plotting methods, which utilized the ggplot2 R package [90].

### 3.5.4 Quantitative real-time PCR

Total RNA was isolated from liver, LD skeletal muscle and subcutaneous fat tissues from 34 MSUPRP pigs, including the pig chosen for sequencing, using TRIzol reagent (Ambion) according to the manufacturer's instructions. Concentrations were measured using a NanoDrop spectrophotometer (Thermo Scientific), and quality and integrity were determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.). Total RNA was reverse transcribed using random primers with the High Capacity cDNA Reverse Transcription Kit with RNase Inhibiter (Applied Biosystems) according to the manufacturer's instructions. A pig ADAR Custom TaqMan Gene Expression assay was designed using the online Custom TaqMan Assay Design Tool (ThermoFisher Scientific). The assay was designed to span exons 2-3 of the pig ADAR gene (Accession No. NC_010446.4). Assays were performed in triplicate using 50 ng cDNA and the TaqMan Gene Expression Master Mix (20 $\mu$l final volume per reaction) in a StepOnePlus Real-Time PCR System (Applied Biosystems). Cycling conditions were 52°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Relative expression values were obtained using the

2-$^{\Delta\Delta}$CT method, with the muscle sample used for sequencing as a calibrator and Ubiquitin C as a reference gene (Applied Biosystems Assay No. Ss03374343_g1). Inference of differential ADAR expression was calculated by one-way ANOVA (main effect of tissue on ADAR expression), and Tukey HSD (pairwise comparisons of tissue means).

### 3.5.5 Calculating probability of mapping error

The average phred-scaled mapping quality $MQ$ across all samples at mismatch site $i$ is provided by SAMTools output. From $MQ$ we can compute the probability of mapping error $p$ according to:

$$p_i = 10^{\frac{-MQ_i}{10}}$$

It follows that the probability of observing 5 "edited" reads at a homozygous site with a cDNA sequencing depth of $N$ assuming no RNA editing can be modeled using the binomial distribution, where:

$$P(X \geq 5 \mid N, p) = 1 - P(X < 5) = 1 - \sum_{j=0}^{4} \binom{N}{j} p^j (1-p)^{N-j}$$

### 3.5.6 Incorporating RepeatMasker and Variant Effect Predictor data using editTools

The editTools function `add_repeatmask()` was used to merge a mismatch data object (generated with `find_edits()`) with susScr3, a Repeatmasker dataset available for download at: `http://www.repeatmasker.org/species/susScr.html`. This function utilizes a binary search algorithm implemented in C++ to process large RepeatMasker files efficiently. The function `write_vep()` was used to generate Variant Effect Predictor input from a mismatch data object. The output of Variant Effect Predictor was merged with the mismatch data object using `add_vep()`. Additional documentation for `find_edits()`, `write_vep()`, `add_vep()`, `add_repeatmask()` is available within editTools v2.1.

## 3.6 Authors' Contributions

Conceived and designed the study: CWE. Contributed samples from the MSU pig resource population: ROB, CWE, NER. Isolated RNA and DNA: NER. Developed software and analysis pipeline: SAF, JPS. Performed qPCR assays and analysis: DS, NER, SAF. Wrote the manuscript: SAF. All authors read and approved the final manuscript.

## 3.7 Acknowledgements

Computing resources were provided by Michigan State University's High Performance Computing Cluster. Sequencing was performed at the Michigan State University Research Technology Support Facility.

# CHAPTER 4

# ESTIMATING THE COHERITABILITY BETWEEN SITE-SPECIFIC RNA EDITING AND ECONOMICALLY IMPORTANT TRAITS IN PIGS

## 4.1 Abstract

The highly conserved post-transcriptional mechanism known as adenosine to inosine (A-to-I) RNA editing impacts gene function by converting adenosine to inosine molecules within specific regions of the transcriptome. The degree that specific sites are edited—the "editing level"—has been observed to vary within populations and can be considered a molecular quantitative trait hypothesized to influence higher-order phenotypes. Here we utilized 940 $F_2$ animals and a combination of univariate and bivariate mixed models to study the shared genetic contributions to RNA editing activity in *longissimus dorsi* muscle tissue and economically important pig traits. We identified five RNA editing sites across four genes whose editing level variation was significantly attributed to the additive effects of all observed SNP markers (estimated genomic heritability $\hat{h}^2_g$ = 0.31—0.56; $p$-value = $8.2\text{x}10^{-5}$—$8.8\text{x}10^{-4}$). Using a multi-polygenic model to localize genomic heritability estimates to a region of interest, across all five editing sites we found suggestive evidence that a portion of the genomic heritability can be attributed to SNPs flanking *ADAR*. When using bivariate models to estimate local genetic correlations between site-specific editing levels and 67 complex traits, we found nominally-significant evidence that the *ADAR* locus contributes to a negative relationship between editing activity and phenotypically related growth traits including average daily gain (local genetic correlation $\hat{\rho}_{g_{local}}[SE]$ = -0.87 [0.16]; $p$-value = 0.029). This work suggests potential pleiotropy between RNA editing activity and complex growth traits in pigs and encourages further use of mixed models to determine if RNA editing can link genetic variation to complex trait variation.

## 4.2 Introduction

According to popular and prevailing theory, quantitative trait loci (QTL) influence complex traits largely by influencing gene expression. Functional geneticists have investigated this theory primarily by performing genome-wide associations (GWA) to find QTL that associate with transcript abundance, otherwise known as expression QTL or eQTL [91, 92, 93]. Gene expression, however, involves a complex array of processes beyond upregulating and downregulating transcript abundance. For instance, RNA splicing has been shown to be influenced by genetic effects and can explain a substantial part of complex trait and disease risk variation [29, 94].

Still, additional forms of gene expression are continually being evaluated for their ability to link genetic variation to higher-order trait variation. A highly-conserved post-transcriptional mechanism known as adenosine to inosine (A-to-I) RNA editing regulates gene expression by converting adenosine to inosine molecules within pre-mRNA transcripts [31], a process catalyzed by adenosine deaminase acting on RNA (ADAR). At numerous edited sites in the transcriptome, the proportion of transcripts containing the edited inosine variant—the so-called "editing level"—has been shown to vary between individuals in a population [35]. In this way, site-specific RNA editing activity has been considered a heritable quantitative trait; numerous studies have performed GWA to identify editing QTL (edQTL) in such species as humans, mice, drosophila, cattle, and pigs [95, 96, 97, 98, 99, 36], with a general consensus across species and populations that genome-wide significant edQTL routinely co-localize with the RNA editing site they are associated with.

While the strongest edQTL signals mostly appear cis-acting, the degree that additive genetic effects (genome-wide) influence editing variation remains largely unknown. Similarly, we lack an understanding of how similarity (or covariance) in RNA editing activity and complex traits may be attributed to shared additive genetic sources. Here, we will refer to the proportion of variation attributable to the additive effects of all single-nucleotide polymorphisms (SNPs) as "genomic heritability" ($h^2_g$). Likewise, we will refer to the component of covariance related to additive effects of SNPs as "genomic covariance", or what is sometimes referred to as "coheritability" [100]. Indeed, if RNA editing activity and complex traits possess a genomic covariance, this could

indicate pleiotropic effects between the two and further the hypothesis that RNA editing can serve as a direct link from genetic variation to complex trait variation.

In this study, we have utilized animals from Michigan State University's Pig Resource Population (MSUPRP) [85] to quantify genetic contributions to RNA editing activity in *longissimus dorsi* muscle tissue. We utilize a combination of univariate and bivariate polygenic models to estimate the genomic heritability of site-specific RNA editing activity and estimate the genomic covariance between site-specific editing and economically important pig traits. We further decompose genomic heritability and genomic covariance estimates into local regions of interest—namely the ADAR locus—to infer how such regions may affect both RNA editing activity and higher-order traits. Using a sample of highly heritable RNA editing sites, we find suggestive evidence that SNPs near ADAR influence both RNA editing activity and complex growth traits, which encourages further study.

## 4.3 Results

### 4.3.1 Heritable RNA editing activity impacts pig *longissimus dorsi* muscle gene expression

To study RNA editing activity in pig *longissimus dorsi* (LD) muscle tissue, we utilized a "discovery cohort" consisting of three adult pigs, each with whole-genome sequencing (WGS) and LD RNA-sequencing (RNASeq) to identify high-confidence RNA editing sites that were detectable across multiple animals. We followed a standard procedure outlined previously [62] to identify DNA-to-RNA mismatches, resulting in 104 A-to-G mismatches (indicative of potential A-to-I RNA editing activity) detectable in at least two of the three discovery cohort animals.

We then utilized an "analysis cohort" consisting of a subset of animals from Michigan State University's Pig Resource Population (MSUPRP) each with LD RNAseq (N = 168); for each animal we estimated editing levels at each of the 104 sites, where we define the estimated editing level to be the ratio of the number of apparently edited reads (containing G) over the total number of reads. Unsurprisingly, only 47 / 104 sites showed evidence that the editing level could be normally distributed across pigs, using a low-threshold shapiro-wilk test ($p$-value $\geq 1\text{x}10^{-10}$). This may

Table 4.1: RNA editing sites exhibiting heritable variability in *longissimus dorsi* muscle tissue

| Site | Strand | Gene[a] | Location[a] | N[b] | $\hat{\sigma}^2_g$ (SE)[c] | $\hat{\sigma}^2_\varepsilon$ (SE)[d] | $\hat{h}^2_g$ | $p$-value[e] |
|---|---|---|---|---|---|---|---|---|
| Chr1:126,167,425 | - | *BLOC1S6* | 3'-UTR | 165 | 0.31 (0.16) | 0.69 (0.14) | 0.31 | $8.8\text{x}10^{-4}$ |
| Chr6:39,368,241 | - | *UQCRFS1* | intron | 166 | 0.41 (0.18) | 0.60 (0.14) | 0.41 | $1.6\text{x}10^{-4}$ |
| Chr15:110,910,484 | + | *CCNYL1* | 3'-UTR | 166 | 0.39 (0.17) | 0.61 (0.13) | 0.39 | $1.7\text{x}10^{-4}$ |
| Chr16:26,512,555 | - | *OXCT1* | intron/3'-UTR[f] | 139 | 0.58 (0.22) | 0.45 (0.15) | 0.56 | $8.2\text{x}10^{-5}$ |
| Chr16:26,512,645 | - | *OXCT1* | intron/3'-UTR[f] | 159 | 0.34 (0.17) | 0.65 (0.14) | 0.34 | $2.3\text{x}10^{-4}$ |

[a] Gene annotation and editing site location provided by ensembl 95 predictions.
[b] Sample size (the number of animals with a detectable editing level)
[c] Genomic variance component. REML estimate (Standard Error)
[d] Residual variance component. REML estimate (Standard Error)
[e] $p$-value from a likelihood ratio test, testing $H_0 : \sigma^2_g = 0$.
[f] Multiple predicted isoforms present at editing site

reflect that while RNA editing activity may be allowed to vary in the population at some sites, other sites show much more constraint, as shown previously [101].

For each of 47 RNA editing sites that showed variation in editing activity among the MSUPRP, we fit a restricted maximum likelihood (REML)-based univariate genomic best unbiased linear predictor (GBLUP) model (see Methods) to decompose editing level variance into genomic and residual components. Exactly five sites showed a significant genomic variance component using a likelihood ratio test (LRT) at a Bonferroni-corrected threshold of 0.05 / 47 = 0.001 (Table 4.1). Variance component estimates and sample sizes are shown for all 47 editing sites in Table B.1.

### 4.3.2 Genetic variants near ADAR are suspected to contribute to editing level variation across sites

For each heritable RNA editing site (Table 4.1), we sought to identify SNPs strongly associated with editing levels using mixed-model GWA methods that control for kinship among the $F_2$ animals (see section 4.5). We identified cis-acting genome-wide significant signals ($p$-value $\leq 1\text{x}10^{-5}$; 5% estimated FDR) for both editing sites within $OXCT1$ but surprisingly no cis-acting genome-wide significant signals were detected for the remaining three RNA editing sites (Fig 4.1A). Curiously, we observed suggestive GWA peaks (peak $p$-values: $3\text{x}10^{-4}$—$2.9\text{x}10^{-5}$) near *ADAR* (otherwise known as *ADAR1* in humans) for all five editing sites.

To investigate the suggestive *ADAR*-localized edQTL signals, pairwise linkage-disequilibrium

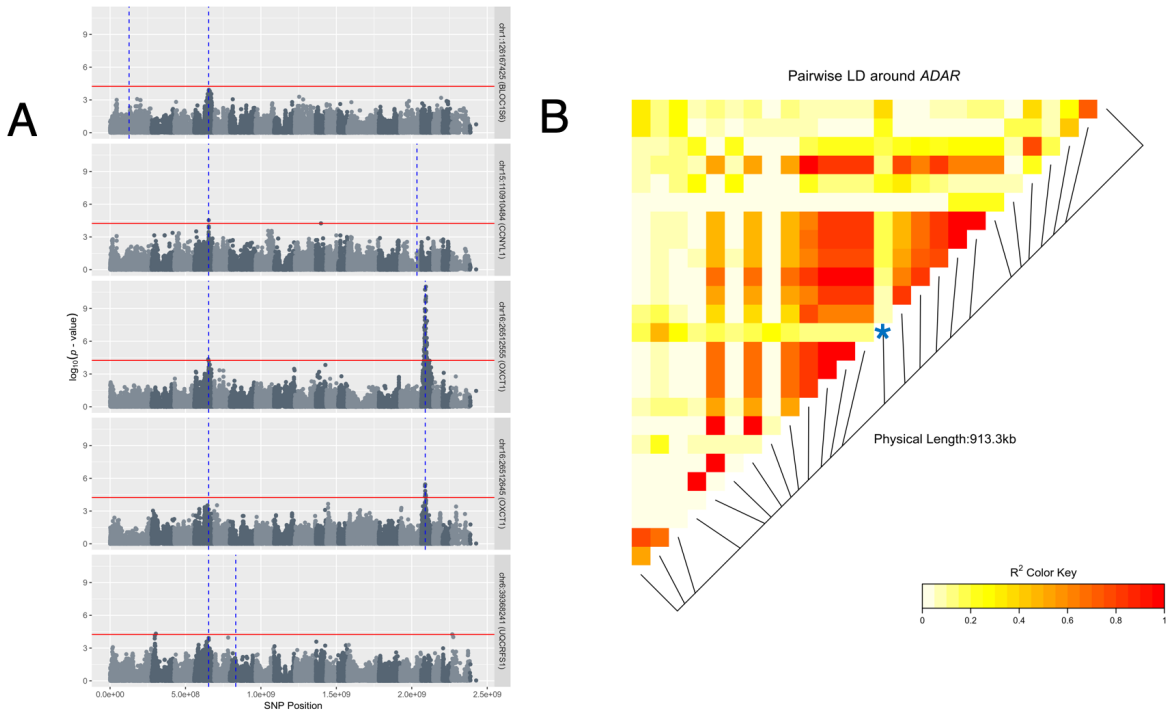


Figure 4.1: GWA for site-specific editing levels
**(A)** Manhattan plot for each RNA editing site. Red line indicates estimated FDR of 10%. Each facet corresponds to a different RNA editing site. For each facet, blue dashed lines indicate position of ADAR on chromosome 4 and position of the editing site. **(B)** Pairwise LD plot between 26 SNPs selected flanking ADAR. The SNP nearest ADAR is marked with a blue asterisk.

estimates ($R^2$) were obtained for a 1MB region surrounding ADAR (Fig 4.1B). This utilized genotypes at 26 SNPs for all genotyped MSUPRP animals (N = 1015), where pairwise two-SNP haplotype frequencies used in $R^2$ calculations were estimated using maximum likelihood [102]. Intriguingly, the SNP nearest *ADAR* (however not within *ADAR*), H3GA0013586, is in relatively poor linkage-disequilibrium with other SNPs in the ~1MB region (pairwise $R^2$ with H3GA0013586 < 0.6). Longer-range linkage disequilibrium with H3GA0013586 was observed to drop beyond this 1MB region (Fig B.1). Hypothetically, if causal variants within *ADAR* contribute to variance in RNA editing activity across sites, this suggests that a weak edQTL signal at *ADAR* could reflect poor linkage-disequilibrium between *ADAR* causal variants and SNP markers.

For each RNA editing site we sought to quantify the proportion of genomic variance in editing activity explained by SNPs flanking *ADAR*, as well as SNPs flanking *OXCT1*. We defined our

Table 4.2: Proportion of editing level genomic variance explained by SNPs flanking *ADAR* and *OXCT1*

| Local region | Editing Site (gene) | $\hat{\sigma}^2_{g_{local}}$ (SE)[a] | $\hat{\sigma}^2_{g_{BG}}$ (SE)[b] | $\hat{\sigma}^2_{g_{local}} / \hat{\sigma}^{2c}_{g}$ | $p$-value[d] |
|---|---|---|---|---|---|
| *ADAR* | chr1:126167425 (*BLOC1S6*) | 0.121 (0.104) | 0.125 (0.118) | 0.49 | $8.02\text{x}10^{-4}$ |
| | chr6:39368241 (*UQCRFS1*) | 0.192 (0.145) | 0.186 (0.129) | 0.51 | $1.70\text{x}10^{-4}$ |
| | chr15:110910484 (*CCNYL1*) | 0.141 (0.115) | 0.167 (0.124) | 0.46 | $9.12\text{x}10^{-5}$ |
| | chr16:26512555 (*OXCT1*) | 0.088 (0.091) | 0.342 (0.176) | 0.21 | $1.23\text{x}10^{-3}$ |
| | chr16:26512645 (*OXCT1*) | 0.076 (0.080) | 0.254 (0.151) | 0.23 | $1.16\text{x}10^{-2}$ |
| *OXCT1* | chr1:126167425 (*BLOC1S6*) | 0.008 (0.026) | 0.313 (0.162) | 0.03 | $3.40\text{x}10^{-1}$ |
| | chr6:39368241 (*UQCRFS1*) | 0.005 (0.022) | 0.412 (0.177) | 0.01 | $3.80\text{x}10^{-1}$ |
| | chr15:110910484 (*CCNYL1*) | 0.000 (0.016) | 0.387 (0.170) | 0.00 | $5.00\text{x}10^{-1}$ |
| | chr16:26512555 (*OXCT1*) | 0.885 (0.541) | 0.256 (0.125) | 0.78 | $7.26\text{x}10^{-14}$ |
| | chr16:26512645 (*OXCT1*) | 0.338 (0.253) | 0.251 (0.141) | 0.57 | $1.37\text{x}10^{-5}$ |

[a] The "local" genomic variance REML estimate (Standard error)
[b] The "background" genomic variance REML estimate (Standard error)
[c] The ratio $\hat{\sigma}^2_{g_{local}} \; / \; \hat{\sigma}^2_{g_{local}} + \hat{\sigma}^2_{g_{BG}}$
[d] $p$-value testing $H_0 : \sigma^2_{g_{local}} = 0$

region of interest flanking *ADAR* to be the 27 SNPs identified previously (Fig 4.1B), and similarly defined our region of interest flanking *OXCT1* to be 25 SNPs within 500Kb on either side of the chr16:26512555 editing site. Like before, we modeled RNA editing levels using a REML-based univariate GBLUP model, but used a "local" polygenic random effect (aggregate effect of SNPs within a region of interest) and "background" polygenic random effect (aggregate effect due to all SNPs other than the SNPs of interest) (see section 4.5). With this model, we decomposed the genomic variance into local and background components and used a LRT to determine whether inclusion of the local polygenic effect fits the data any better than omitting it (Table 4.2). We observed suggestive evidence that the aggregate effect of SNPs near *ADAR* contributes to editing level variation (LRT $p$-values: $9.12\text{x}10^{-5}$—$1.16\text{x}10^{-2}$), but were unable to confidently quantify the proportion of genomic variance it explains due uncertainty in local and background genomic variance estimates. Our best estimate showed that roughly 50% of the genomic variance in RNA editing activity for *BLOC1S6*, *UQCRFS1*, and *CCNYL1* sites can explained by SNPs flanking *ADAR*, while that estimate reduces to 20% for *OXCT1* editing activity.

### 4.3.3 Suggestive evidence for a shared genetic architecture between RNA editing activity and complex traits

Given that variation in RNA editing activity across sites is potentially attributable to SNPs flanking *ADAR*, we sought to infer whether *ADAR*-flanking SNPs also contribute to variation in complex traits. Using the same two-polygenic univariate model as before, and the full MSUPRP with phenotype and SNP genotype data (N = 940), we tested whether variance among 67 growth, meat quality, and carcass composition (GMQCC) traits could be attributed to 27 SNPs flanking *ADAR*. Surprisingly, 15/67 traits (more than expected by chance) showed nominal evidence ($p$-value $\leq 0.05$) that *ADAR*-flanking SNPs contribute to their variance (Table B.2), with average daily gain (ADG) possessing the strongest evidence ($\hat{\sigma}^2_{g_{local}}$ [SE]: 0.14 [0.08]; $\hat{\sigma}^2_{g_{BG}}$ [SE]: 0.27 [0.05]; $p$-value = 2.9x10$^{-4}$). In contrast, only 1/67 GMQCC traits showed nominal evidence that *OXCT1*-flanking SNPs contribute to their variance.

We next utilized a bivariate model to decompose the covariance between site-specific editing activity and GMQCC traits into genomic and residual components (see section 4.5) (Table 4.3). All animals with RNA editing records ($n_1 \sim 168$) were among the larger set of MSUPRP animals with GMQCC records ($n_2 \sim 940$), enabling us to model the residual covariance between editing activity and GMQCC traits. For each of the five RNA editing sites, we inferred the genomic covariance between editing levels and 67 GMQCC traits, totaling 335 tests. We observed no significant genomic covariances after multiple test correction ($p$-value $\leq 0.05$ / 335 = 1.5x10$^{-4}$), but observed that $p$-values begin deviating from what is expected at around $p$-value = 0.1 (Fig 4.2). This provides small and subtle evidence that genomic covariances between RNA editing activity and complex traits exist, but perhaps at a magnitude that we are underpowered to detect either due to insufficient sample size or imperfect linkage-disequilibrium between SNPs and causal variants. The most significant genomic covariances estimated were between longissimus muscle moisture content (moisture) and the chr16:26,512,645 edited site within *OXCT1* ($\hat{\rho}_g$ [SE] = -0.70 [0.20]; $p$-value = 0.004), and between 45-min carcass temperature (temp_45m) and the chr15:110,910,484 edited site within CCNYL1 ($\hat{\rho}_g$[SE] = 0.70 [0.20]; $p$-value = 0.001). A full list of all genomic

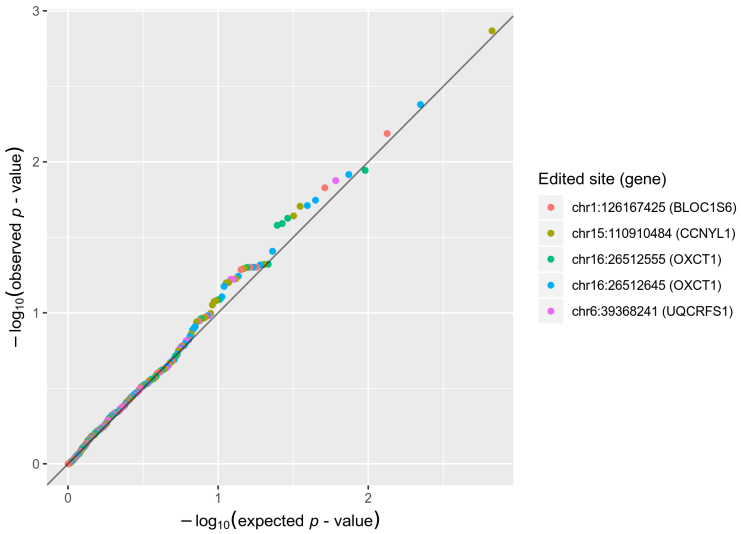


Figure 4.2: Quantile-quantile plot testing for genome-wide genomic covariances between site-specific editing levels and complex traits

covariances estimated are shown in Table B.3.

Interestingly, ADG, which showed the highest evidence of local genomic variance attributable to *ADAR*-flanking SNPs, showed modest evidence of a genomic covariance with editing activity at chr15:110,910,484, given a LRT $p$-value of 0.1 (Table B.3) Given that genomic covariances provide no information about where in the genome genetic effects are shared between traits, we sought to infer local genomic covariances attributable to *ADAR*-flanking SNPs using a two-polygenic bivariate model (see section 4.5). As before, when considering multiple tests, we observed no significant local genomic covariances attributable to *ADAR*-flanking SNPs. As expected however, ADG, along with phenotypically related growth traits, showed the highest evidence of an *ADAR*-localized genomic covariance with editing activity (Table 4.4). All our top *ADAR*-localized genomic covariance estimates were negative (except for Days to 105kg); if these are true positive signals, it suggests genetic variants near *ADAR* contribute to a negative relationship between RNA editing activity (particularly at chr15:110910484) and growth traits such as average daily gain, body weight at 22 weeks, empty body lipid, total body fat tissue, etc. Even if this is true, additional work would be needed to determine if *ADAR* variants exhibit pleiotropic effects between RNA editing activity and growth traits, or if separate causal variants affecting editing activity and growth traits are simply in

Table 4.3: Top genomic covariance estimates between site-specific RNA editing levels and growth, meat quality, and carcass composition traits[a]

| Editing site (gene) | Trait[b] | $\hat{\rho}_g$ (SE)[c] | $\hat{\sigma}_{g_1 g_2}$ [d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$ [e] | $\hat{\sigma}_{p_1 p_2}$ [f] | $p$-value[g] |
|---|---|---|---|---|---|---|
| chr1:126,167,425 (*BLOC16S*) | b | 0.63 (0.21) | 0.19 | -0.04 | 0.16 | 0.006 |
| | temp_45m | 0.67 (0.22) | 0.17 | -0.02 | 0.15 | 0.015 |
| | moisture | -0.48 (0.23) | -0.14 | 0.12 | -0.02 | 0.050 |
| chr15:110,910,484 (*CCNYL1*) | temp_45m | 0.70 (0.20) | 0.19 | -0.21 | -0.02 | 0.001 |
| | conn_tiss | -0.55 (0.23) | -0.14 | 0.05 | -0.08 | 0.020 |
| | fftoln | -0.51 (0.19) | -0.18 | 0.09 | -0.09 | 0.023 |
| chr16:26,512,555 (*OXCT1*) | moisture | -0.52 (0.18) | -0.22 | 0.03 | -0.19 | 0.011 |
| | picnic | 0.62 (0.16) | 0.33 | -0.26 | 0.06 | 0.024 |
| | color | -0.47 (0.19) | -0.18 | -0.02 | -0.19 | 0.026 |
| chr16:26,512,645 (*OXCT1*) | moisture | -0.70 (0.20) | -0.22 | 0.05 | -0.17 | 0.004 |
| | lrf_22wk | 0.53 (0.21) | 0.18 | -0.01 | 0.17 | 0.012 |
| | fat | 0.52 (0.19) | 0.21 | 0.05 | 0.26 | 0.018 |
| chr6:39,368,241 (*UQCRFS1*) | boston | -0.59 (0.14) | -0.32 | -0.09 | -0.40 | 0.013 |
| | temp_24h | -0.48 (0.20) | -0.14 | -0.34 | -0.48 | 0.060 |
| | wt_3wk | -0.69 (0.31) | -0.09 | 0.04 | -0.05 | 0.060 |

[a] Shown are the top three genomic covariance estimates (ranked by $p$-value) for each RNA editing site

[b] b = b* objective color; temp_45m = carcass temperature at 45m; conn_tiss = connective tissue sensory panel analysis; fftoln = fat-free total lean tissue; picnic = picnic shoulder cut weight; color = subjective color; lrf_22wk = last rib back fat at 22 weeks; fat = fat percentage; boston = boston shoulder cut weight; wt_3wk = body weight at 3 weeks

[c] Genetic correlation estimate (Standard error), where $\rho_g = \sigma_{g_1 g_2} \Big/ \sqrt{\sigma^2_{g_1} \sigma^2_{g_2}}$

[d] Genomic covariance REML estimate

[e] Residual covariance REML estimate

[f] Phenotypic covariance

[g] $P$-value testing $H_0 : \sigma_{g_1 g_2}$

linkage-disequilibrium.

## 4.4 Discussion

To date, few RNA editing studies have estimated genome-wide parameters such as heritability, including a previous study that evaluated regulation of 5-HT2C receptor editing activity [103]. Heritability estimates are valuable in that they estimate the degree that the aggregate effect of all causal variants (including those with relatively weak effects) influence phenotypes such as RNA editing levels. In comparison, traditional GWA techniques including those used to identify edQTL, are known to be underpowered such that only relatively large effect loci are detectable [104]. Consistent with previous studies [95, 96, 97, 98, 99, 36], among our sample of five significantly heritable RNA editing sites we observe the strongest genetic contributions of editing activity to be cis-acting. However, only editing sites within *OXCT1* were found to have genome-wide significant cis-acting edQTL, suggesting the remaining three sites are under alternative genetic control. We

Table 4.4: Top local genomic covariance estimates attributable to *ADAR*-flanking SNPs between site-specific RNA editing and growth, meat quality, and carcass composition traits[a]

| Editing site (gene) | Trait[b] | trait $\hat{h}^2_{g_{local}}$ [c] | $\hat{\rho}_{g_{local}}$ (SE) [d] | $\hat{\sigma}_{g_1 g_{2local}}$ [e] | $p$-value [f] |
|---|---|---|---|---|---|
| chr15:110910484 (*CCNYL1*) | ADG | 0.14 | -0.87 (0.16) | -0.17 | 0.029 |
| chr15:110910484 (*CCNYL1*) | wt_22wk | 0.06 | -0.86 (0.21) | -0.09 | 0.046 |
| chr15:110910484 (*CCNYL1*) | Days | 0.09 | 0.75 (0.27) | 0.09 | 0.052 |
| chr15:110910484 (*CCNYL1*) | mtfat | 0.04 | -0.88 (0.22) | -0.07 | 0.054 |
| chr15:110910484 (*CCNYL1*) | tofat | 0.09 | -0.79 (0.24) | -0.10 | 0.061 |
| chr15:110910484 (*CCNYL1*) | fftoln | 0.03 | -0.86 (0.24) | -0.07 | 0.064 |
| chr15:110910484 (*CCNYL1*) | mtpro | 0.05 | -0.78 (0.27) | -0.07 | 0.070 |
| chr6:39368241 (*UQCRFS1*) | Days | 0.08 | 0.65 (0.33) | 0.08 | 0.146 |
| chr6:39368241 (*UQCRFS1*) | wt_22wk | 0.06 | -0.73 (0.28) | -0.09 | 0.146 |
| chr15:110910484 (*CCNYL1*) | bf10_22wk | 0.05 | -0.67 (0.35) | -0.06 | 0.149 |

[a] Top 10 local genomic covariance estimates, ranked by $p$-value.
[b] ADG = average daily gain; wt_22wk = body weight at 22 weeks; Days = Days to 105kg; mtfat = empty body lipid; tofat = total body fat tissue; fftoln = fat-free total lean tissue; mtpro = empty body protein; bf10_22wk = 10th rib back fat at 22 weeks
[c] Estimated genomic heritability of the trait, localized to *ADAR*-flanking SNPs, as estimated from a two-polygenic bivariate model
[d] Local genetic correlation (Standard error)
[e] Local genomic covariance REML estimate
[f] Phenotypic covariance
[g] $P$-value testing $H_0 : \sigma_{g_1 g_{2local}}$

further provide suggestive evidence that variants near *ADAR* may be contributing a proportion of editing level variation across multiple editing sites, using models that decompose the aggregate effect of all causal variants to a region of interest.

Genomic covariances provide a high-level understanding of whether two traits are under similar genetic control. Under the cascading hypothesis that genetic variation (A) contributes to editing level variation (B), which contributes to complex trait variation (C) (A $\rightarrow$ B $\rightarrow$ C; an example of "vertical pleiotropy" [100]), it is required that the genomic covariance between editing levels and complex traits be non-null (this assumes sufficient linkage-disequilibrium between SNP markers and causal variants). In pursuit of such evidence, we were unable to infer non-null genomic covariances among 335 pairwise tests (5 RNA editing sites by 67 complex traits), but observed subtle evidence that genomic covariance $p$-values begin deviating from what is expected at around $p$-value = 0.1. This could indicate small-magnitude genomic covariances between RNA editing activity and complex traits, possibly due to imperfect linkage-disequilibrium between SNP markers and causal variants. We further localize genomic covariances to SNPs flanking *ADAR* to find suggestive evidence that the *ADAR* locus contributes to a negative relationship between RNA editing activity (particularly at the chr15:110910484 editing site) and numerous phenotypically

related growth traits (Table 4.4). While many of our top genomic covariance signals did not correspond with top *ADAR*-localized genomic covariance signals (comparing Table 4.3 and 4.4), this is not unordinary given that genomic covariances reflect co-localization of causal variants, on average across the whole genome. Curiously, we noticed that several editing-trait pairs with top *ADAR*-localized genomic covariance evidence had genome-wide genomic covariance $p$-values that deviated from what was expected (Table B.3; Figure 4.2). For example, ADG and the editing level at chr15:110910484 had a genome-wide genomic covariance LRT $p$-value of 0.1, while fat-free total lean tissue (fftoln) and chr15:110910484 had a LRT $p$-value of 0.02.

Despite finding suggestive evidence that variants near *ADAR* may influence both RNA editing activity and numerous growth traits, our inability to obtain significant evidence of this hypothesis after multiple test correction warrants further study; a larger sample size is suggested to obtain sufficient statistical power and more precise estimates of local genomic variances/covariances. We also note that our analysis of RNA editing activity was limited to five sites. As suggested from an earlier study [101], perhaps relatively few RNA editing sites show considerable variation in editing levels from individual to individual, and perhaps fewer still exhibit variation attributable to genetic variation. Still, our ability to detect heritable RNA editing sites in LD muscle tissue was likely affected to by two major factors: 1) skeletal muscle tissue is known to exhibit relatively low RNA editing activity [35], and 2) the RNA sequencing depth of our analysis cohort (N = 168) was relatively modest (~63M reads for 24/168 animals and ~23M reads for the remaining 144 animals). This second point meant we were limited to surveying edited genes that are relatively highly expressed in LD muscle tissue. Another limitation affecting this study was that genotypes at edited sites within the analysis cohort were unobserved. However, for each of the five sites considered, it is unlikely that variation in the editing level (proportion of reads containing a G) simply reflected allele content variation at a SNP because 1) all five sites possessed suggestive edQTL signals at *ADAR* and 2) editing level genomic heritability estimates were relatively low ($<$ 0.6); by definition, the heritability of allele content at a SNP is exactly one [105].

In this study we have utilized a combination of univariate and bivariate mixed models to

decompose variation in RNA editing activity and variation in complex traits into shared genetic sources. This approach is highly suggested to influence our understanding of RNA editing regulation and the degree that genetically regulated RNA editing activity could influence complex traits. As data from more animals and more tissues becomes available, we foresee the answer to whether RNA editing can directly link genetic variation to complex trait variation becoming clearer.

## 4.5 Materials and Methods

### 4.5.1 Sequencing data

To discover candidate RNA editing sites for downstream genetic analysis, a "discovery cohort" consisting of three adult animals with matched whole genome sequencing and RNA sequencing from LD muscle tissue were used. Two animals were Yorkshire pigs from the Functional Annotation of Animal Genomes Project (FAANG; `https://www.animalgenome.org/community/FAANG/`) and the third was an $F_2$ pig from Michigan State University's pig resource population (MSUPRP), originating from four $F_0$ Duroc sires and 15 $F_0$ Pietrain dams [85].

Whole genome sequencing of the two FAANG animals was done using 100bp paired-end reads; one animal was sequenced at a depth of ~489M reads and the other sequenced at ~564M reads. Whole genome sequencing of the $F_2$ pig was done using 150bp paired-end reads, totaling ~249M reads. LD Muscle RNA sequencing from the FAANG animals was done using 100bp paired-end strand-specific reads, with one animal sequenced at a depth of ~56M cDNA reads and the other sequenced at ~42M cDNA reads. Finally, LD RNA seq from the $F_2$ pig consisted of ~104M 150bp paired-end, strand-specific cDNA reads. More details regarding library prep and sequencing of the $F_2$ animal can be found in Funkhouser et al. [62].

At each candidate RNA editing site, editing levels were estimated among a subset of the MSUPRP possessing LD muscle RNA sequencing (N = 168). RNA sequencing of these animals is detailed in full in Velez-Irizarry et al. [106], resulting in a depth of ~63M strand-specific cDNA reads for 24/168 animals and ~23M strand-specific cDNA reads for the remaining 144 animals.

### 4.5.2 Genotyping data

To analyze the genetic architecture of each RNA editing site, we utilized the MSUPRP, a population that has been genotyped using the Illumina PorcineSNP60 BeadChip with SNPs mapped to the Sscrofa11.1 genome assembly. Genotype data pre-processing is detailed in an earlier study [106]. Briefly, using all MSUPRP genotyped animals (N = 940), the following SNPs were removed from analysis: monomorphic and non-autosomal SNPs, SNPs with evident mendelian error, and SNPs with a minor allele frequency less than 0.01. This resulted in 43,130 markers used in all genetic analyses.

### 4.5.3 Phenotypes

Phenotypes from the MSUPRP have been described in previous studies [85, 106, 107]. A total of 67 traits (29 growth traits, 20 carcass composition traits, and 18 meat quality traits) were tested to be genetically correlated with 5 heritable editing levels. Brief descriptions and summary statistics for each trait can be found in [106].

### 4.5.4 Sequencing data preparation and RNA editing detection

Using the "discovery cohort" consisting of three animals with matched WGS and RNA sequencing, preparation of sequencing data and discovery of RNA editing sites was largely consistent with Funkhouser et al. [62]. For both WGS and RNAseq, trimmomatic version 0.38 [108] was used to remove low quality bases at the 3' end, retaining minimum length sequences of 56 bps. Six bases at the 5' end were also removed from cDNA reads to remove any artifactual bases introduced during cDNA synthesis [76]. Trimmed DNA reads were mapped to the Sus Scrofa 11.1 reference assembly with Bowtie version 2.3.2 and trimmed RNA reads were mapped to the same reference with TopHat version 2.1.1. DNA and RNA reads that had more than one recorded alignment were removed from further analysis. Prior to variant calling, RNAseq alignments were split such that plus-strand transcript alignments were separated from minus-strand transcript alignments.

To detect RNA editing sites for downstream genetic analysis, variant calling was performed using Samtools version 1.7 and bcftools 1.9.64, whereby, for each animal, variants were jointly called between DNA and strand-specific RNA alignments. Resulting variant calling data is then processed by editTools (`https://github.com/funkhou9/editTools`), a suite of compiled R functions designed to rapidly screen variant calling data for RNA editing evidence. A DNA-to-RNA mismatch, indicative of an RNA editing event, was detected according to the following criteria: 1) the genotype is homozygous according to 95% of DNA reads, 2) 10 or more reads were used to call the genotype, 3) at least 5 cDNA reads differed from the genotype, and 4) the cDNA reads have a Phred-scaled strand-bias *p*-value of 20 or less. Once LD muscle DNA-to-RNA mismatches were identified in each of the three discovery animals, editing sites were retained for downstream analysis if 1) they were the of the canonical A-to-G form indicative of ADAR activity, 2) they were detectable in both the $F_2$ animal and at least one of the FAANG animals and 3) the genotype at the RNA editing site was homozygous reference. This resulted in 104 putative ADAR-catalyzed RNA editing sites for downstream analysis.

### 4.5.5 Editing level estimation

At each of the 104 putative RNA editing sites previously identified, editing levels were estimated within the "analysis cohort", a subset of MSUPRP pigs with LD muscle RNA sequencing data (N = 168). RNA sequencing from the analysis cohort was trimmed and mapped in the same way as RNA sequencing from the discovery cohort. For each of the 168 animals, variant calling at each of the 104 putative RNA editing sites was performed with Samtools version 1.7 and bcftools 1.9.64; at each site, editing levels were obtained by diving the number of high-quality (base quality >= 25) reads supporting the edited allele by the total number of high-quality reads. Editing levels were discarded if they were estimated with less than 10 high-quality reads.

To further identify RNA editing sites with variable editing levels across animals, we retained an RNA editing site for analysis if 1) it was detectable in at least 10/168 MSUPRP animals, and 2) the editing level was not fixed across animals and showed at least weak evidence of gaussian variance

(shapiro wilk $p$-value > $1x10^{-10}$). This resulted in 47 RNA editing sites, suitable for genomic heritability estimation.

### 4.5.6 Univariate variance component estimation and GWA

A genomic best linear unbiased prediction (GBLUP) model was used to decompose site-specific editing level variance into genomic and residual components. The model can be expressed as:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{x}\beta_{\text{sex}} + \boldsymbol{g} + \boldsymbol{\varepsilon} \tag{4.1}$$

where $\boldsymbol{y}$ is a vector of estimated editing levels for each animal (centered and scaled to mean 0 and unit variance) at one of 47 RNA editing sites, $\mu$ is an overall mean, $\boldsymbol{x}$ is an indicator of the sex of each animal, $\boldsymbol{g}$ a vector of polygenic values, and $\boldsymbol{\varepsilon}$ are residuals. The overall mean and sex-specific deviation from the mean ($\beta_{\text{sex}}$) are assumed fixed values, while polygenic values and residuals are assumed random and distributed as: $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \boldsymbol{I}\sigma^2_{\varepsilon}\right)$ and $\boldsymbol{g} \sim N\left(\mathbf{0}, \boldsymbol{G}\sigma^2_{g}\right)$, where genomic relationships $\boldsymbol{G} = \frac{ZZ'}{tr(ZZ')/n}$, and $\boldsymbol{Z}$ is a centered genotype matrix (with animals in rows and SNPs in columns), centered by subtracting each column by its sample mean. Estimates of interest, $\hat{\sigma}^2_g$ and $\hat{\sigma}^2_{\varepsilon}$, and standard errors thereof were derived using REML and the inverse of the information matrix, respectively.

For each model fit (eq. 4.1), GWA was performed by transforming predicted additive genomic values $\hat{\boldsymbol{g}}$ to estimates of additive SNP effects $\hat{\boldsymbol{b}}$ and their (co)variances $Var(\hat{\boldsymbol{b}})$ [109, 110]. The test statistic used to test the null hypothesis of no association between allele dosage at SNP $j$ and editing level were obtained with $T_j = \frac{\hat{b}_j}{\sqrt{Var\left(\hat{b}_j\right)}} \sim N(0,1)$. P-values from such a test have been shown to be equivalent to a test in which a single SNP is associated with the trait, while modeling a random polygenic effect (akin to EMMAX) [109, 110, 111]. Functions to fit equation 4.1, and transform genomic values to SNP effects and variances are provided in the gwaR package (`https://github.com/steibelj/gwaR`).

To estimate the genomic variance localized to a site of interest, we used a two polygenic effect model:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{x}\beta_{\text{sex}} + \boldsymbol{g}_{local} + \boldsymbol{g}_{BG} + \boldsymbol{\varepsilon} \tag{4.2}$$

where $\boldsymbol{g}_{local}$ are genomic values arising from selected SNP markers, and $\boldsymbol{g}_{BG}$ are background genomic values arising from all markers except the selected SNP markers. It is assumed $\boldsymbol{g}_{local} \sim N\left(\mathbf{0}, \boldsymbol{G}_{local}\sigma^2_{g_{local}}\right)$ and $\boldsymbol{g}_{BG} \sim N\left(\mathbf{0}, \boldsymbol{G}_{BG}\sigma^2_{g_{BG}}\right)$, where genomic relationships $\boldsymbol{G}_{local}$ and $\boldsymbol{G}_{BG}$ are derived using their corresponding SNP sets.

To test for non-null variance components of interest (such as $\sigma^2_g$ or $\sigma^2_{g_{BG}}$), a chi-squared test statistic from a likelihood ratio test is computed: $LRT_{\sigma^2_k} = 2\left(\mathcal{L} - \mathcal{L}_{\sigma^2_k=0}\right)$, where $\mathcal{L}$ is the log likelihood evaluated at the REML estimate for the full model (either equation 4.1 when testing global genomic variance components or equation 4.2 when testing localized genomic variance components) and $\mathcal{L}_{\sigma^2_k=0}$ is the log likelihood for a reduced model in which the $k^{th}$ variance component of interest is removed. It has been shown that $LRT_{\sigma^2_k}$ asymptotically follows a mixture of $\chi^2_1$ and $\chi^2_0$ distributions [112], therefore the $p$-value from such a test was $\frac{1-F_1\left(LRT_{\sigma^2_k}\right)}{2}$, where $F_1()$ is the chi-squared cumulative distribution function with 1 degree of freedom.

### 4.5.7 Bivariate analysis to estimate genomic covariances

Jointly modeling site-specific editing levels and higher-order phenotypes was done using trait-specific means, sex-specific effects, polygenic effects and residuals:

$$\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{1}\mu_1\\ \mathbf{1}\mu_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{x}_1\beta_{sex_1}\\ \boldsymbol{x}_2\beta_{sex_2}\end{bmatrix} + \begin{bmatrix}\boldsymbol{g}_1\\ \boldsymbol{g}_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_1\\ \boldsymbol{\varepsilon}_2\end{bmatrix},$$

where $\boldsymbol{y}_1 = \left\{y_{1_i}\right\}_{i=1}^{n_1 \sim 168}$ are editing levels at one of 5 editing sites (those pre-determined to possess a significant polygenic effect) and $\boldsymbol{y}_2 = \left\{y_{2_i}\right\}_{i=1}^{n_2 \sim 940}$ are phenotypes at one of 67 growth,

carcass composition, or meat quality traits. The two traits (an editing level and a higher order phenotype) are modeled to be jointly distributed as:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{1}\mu_1 + \boldsymbol{x}_1\beta_{sex_1} \\ \mathbf{1}\mu_2 + \boldsymbol{x}_2\beta_{sex_2} \end{bmatrix}, \begin{bmatrix} \boldsymbol{G}_1\sigma^2_{g_1} + \boldsymbol{I}_{n_1}\sigma^2_{\varepsilon_1} & \boldsymbol{G}_{1,2}\sigma_{g_1g_2} + \boldsymbol{B}_{n_1,n_2}\sigma_{\varepsilon_1\varepsilon_2} \\ \boldsymbol{G}_{2,1}\sigma_{g_1g_2} + \boldsymbol{B}_{n_2,n_1}\sigma_{\varepsilon_1\varepsilon_2} & \boldsymbol{G}_2\sigma^2_{g_2} + \boldsymbol{I}_{n_2}\sigma^2_{\varepsilon_2} \end{bmatrix}\right),$$

where for example, $\boldsymbol{I}_{n_1}$ is the $n_1$-by-$n_1$ identity matrix and $\boldsymbol{B}_{n_1,n_2}$ is an $n_1$-by-$n_2$ logical matrix consisting of 0s and 1s used to link common animals between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ records. $\boldsymbol{G}_{1,2}$ are the genomic relationships between $\boldsymbol{y}_1$ animals (in rows) and $\boldsymbol{y}_2$ (in columns) and $\boldsymbol{G}_1$ are the genomic relationships between $\boldsymbol{y}_1$ animals only. The genetic correlation is defined as $\rho_g = \frac{\sigma_{g_1g_2}}{\sqrt{\sigma^2_{g_1}\sigma^2_{g_2}}}$. Estimates of individual variance components and covariances were obtained using REML, and standard errors of genetic correlation estimates were approximated using the delta method [113]. Inferring non-null genomic covariances or correlations was done using $LRT_{\sigma_{g_1g_2}} = 2\left(\mathcal{L} - \mathcal{L}_{\sigma_{g_1g_2}=0}\right)$, with a $p$-value calculated from $\left(1 - F_1\left(LRT_{\sigma_{g_1g_2}}\right)\right)$. Estimating local genomic covariances were performed similarly, only (co)variance structures used to estimate trait-specific variance components and cross-trait covariances utilized "local" and "background" genomic relationship matrices.

# CHAPTER 5

# CONCLUSION

## 5.1 Gene-by-sex interactions and ideas for future analyses

Among the four complex traits studied in chapter 2, they broadly fall into two categories: i) subtle G×S interactions (such as height, BMI, and bone-mineral density), and ii) large-magnitude G×S interactions (waist-to-hip ratio). For measurements in the second category, they may be considered a different trait when measured in males than when measured in females. In other words, if the factors (genetic or not) that contribute to a measurable outcome are dramatically different between sexes, then one can properly define the male outcome differently from the female outcome. To date, few additional human traits are known to fall into the second category. As a caveat to this discussion, it is possible that small sex-specific effects at genomic regions possessing suggestive G×S interactions could simply reflect poor linkage-disequilibrium between SNPs and QTL. This is possible to check using higher-density (~13 million SNP) imputed data.

To further explore how G×S interactions may influence population-level phenotypic variance by creating mean and variance differences between sexes, it will be interesting to examine minor allele frequencies (MAF) at each detected G×S interaction. Under a single causal variant model in which sex-specific environmental variances are identical, and assuming Hardy-Weinberg equilibrium, a G×S interaction with a high MAF will mainly create differences in variances between sexes. On the other hand, a lower MAF G×S interaction will create differences in means between sexes (as well as differences in variances). It will be worthwhile to formulate the exact relationship between G×E causal variant allele frequency and its effect on mean/variance differences between environments, which will depend on certain assumptions but be applicable to any G×E scenario where allele frequencies are identical between environments.

In Rawlik et al. [14], using a bivariate mixed model and data from the UK Biobank they estimated genome-wide sex-specific additive genetic variances. If additive genetic variances differ

between sexes, this indicates additive SNP effects differ between sexes (because autosomal allele frequencies between sexes are assumed the same). This alone does not indicate whether male-specific and female-specific SNP effects uniformly differ by a proportionality constant, nor does it indicate where in the genome sex-specific effects differ. It would be interesting to formally decompose sex-specific additive variances into approximately independent genomic regions [114], or decompose "cross-sex" genetic covariances to such regions as an alternative means to map G×S interactions.

One plausible biological mechanism explaining why G×S interactions create observable differences between males and females is that some eQTL may have sex-dependent function [115] (only regulate transcript abundance in males, or only in females). This is consistent with the observation that many large effect G×S interactions have an effect in one sex but little to no effect in the other. To further explore this hypothesis, it would be worthwhile to determine if G×S interactions are enriched in G×S interacting eQTL.

## 5.2 Present limitations to modeling RNA editing activity and ideas for future functional genetics studies

In this work, we assess site-specific RNA editing activity by estimating editing levels at each RNA editing site. Here we define the true editing level at an RNA editing site to be the proportion of transcripts (in a population of transcripts transcribed from the same locus) containing the edited inosine variant. In practice, we only estimate editing levels using sequencing data, with each estimate subject to measurement error. If measurement error is substantial, this could negatively impact power to detect global and local genomic variance components contributing to editing level variance. This encourages RNA editing levels to be estimated using relatively deep sequencing to limit editing level estimation error.

In chapter 4, we report modest evidence that covariance between editing activity (as measured in skeletal muscle) and complex traits is driven by a shared genetic architecture. One potential reason for this could be that the growth, carcass composition, and meat quality traits studied are

under differing molecular control. For instance, it is possible that variation in these traits can be attributed to RNA editing activity in a different tissue other than skeletal muscle tissue (RNA editing activity measured in skeletal muscle may be lowly correlated with RNA editing activity in other tissues [35]). It will be particularly interesting to re-visit the hypothesis that genetic variation contributes to covariance between RNA editing activity and complex traits as data from more (related) individuals and tissues becomes increasingly available.

In functional genetics studies such as RNA editing studies, it is fairly common to identify eQTL (or edQTL [99]) that co-localize with phenotypic QTL. This provides evidence that a DNA segment contributes to variation in gene expression and variation in phenotype, however it does not necessarily imply that the DNA segment contributes to covariance between gene expression and phenotype. In chapter 4, we find evidence that SNPs flanking *ADAR* contribute to variation in RNA editing activity and variation in growth traits such as average daily gain, but only detect modest evidence that *ADAR*-flanking SNPs contribute to covariance between RNA editing activity and growth traits. Ultimately, it may be worth re-visiting many co-localizations between phenotypic QTL and edQTL (or sQTL, eQTL, etc.) using bi-variate models to determine if the co-localized DNA region contributes to covariance.

## 5.3 Overall conclusions

Here, multiple perspectives were used to address the long-standing question: how does genetics contribute to phenotypic variation? Under a quantitative genetic perspective, we have provided evidence that numerous small magnitude G×S interactions may together contribute to broad-sense heritability among traits such as human height, BMI, bone-mineral density and waist-to-hip ratio. From a functional genetic perspective, we have investigated the degree that RNA editing, as measured from skeletal muscle tissue, may serve as a plausible biological mechanism linking genetic variation with complex trait variation in pig populations.

This work illustrates the well-accepted belief that traditional GWAS methods—single marker regression (SMR)—are severely underpowered to detect many QTL; in chapter 2, we provided

evidence that SMR may be underpowered to detect typical G×S interactions, even with relatively large sample sizes (N ~ 250,000). In chapter 4, we found RNA editing activity for numerous sites to be highly heritable, yet only a couple editing sites showed genome-wide significant editing QTL (edQTL), suggesting numerous RNA editing sites may be influenced by small additive genetic effects undetectable by SMR (given current sample sizes and SNP markers).

In total, this work encourages the use of local Bayesian regressions to further study G×S and other G×E interactions among unstructured human populations for which large sample sizes exist. It also encourages molecular phenotypes such as RNA editing to be studied using multi-variate models to decompose covariance between molecular phenotypes and complex traits to genetic sources.

# APPENDICES

# APPENDIX A

# CHAPTER 2 SUPPLEMENTARY MATERIAL

Table A.1: Sex-specific phenotype statistics

| trait | statistic | male | female |
|---|---|---|---|
| height | Sample size | 119190 | 139738 |
| height | Sample mean | 176 | 163 |
| height | Sample SD | 6.76 | 6.2 |
| height | 25% Quantile | 172 | 159 |
| height | 75% Quantile | 181 | 167 |
| WHR | Sample size | 119153 | 139681 |
| WHR | Sample mean | 0.935 | 0.816 |
| WHR | Sample SD | 0.0651 | 0.07 |
| WHR | 25% Quantile | 0.892 | 0.766 |
| WHR | 75% Quantile | 0.974 | 0.861 |
| bhmd | Sample size | 106662 | 124970 |
| bhmd | Sample mean | 0.574 | 0.516 |
| bhmd | Sample SD | 0.145 | 0.118 |
| bhmd | 25% Quantile | 0.482 | 0.434 |
| bhmd | 75% Quantile | 0.647 | 0.587 |
| BMI | Sample size | 119061 | 139591 |
| BMI | Sample mean | 27.8 | 27 |
| BMI | Sample SD | 4.22 | 5.13 |
| BMI | 25% Quantile | 24.9 | 23.4 |
| BMI | 75% Quantile | 30 | 29.6 |

Height units: cm
BMD units: g/cm2
BMI units: Kg/m2

Table A.2: Inferred G×S interactions using sex-specific window variances. Listed are all windows with PPDiff$_{\sigma^2_{g_{j^*}}} \geq 0.9$

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | PPM$_{\sigma^2_{g_{j^*}}}$ [c] | PPF$_{\sigma^2_{g_{j^*}}}$ [c] | PPDiff$_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs1535515 | height | 0.0000211 | 0.0001174 | 0.8186207 | 0.9988506 | 0.9563218 | LRRC8C LRRC8D |
| rs580251 | height | 0.0000134 | 0.0000901 | 0.7464368 | 0.9958621 | 0.936092 | LRRC8C LRRC8D |
| rs519989 | height | 0.0000125 | 0.0000826 | 0.7232184 | 0.9903448 | 0.9165517 | LRRC8C LRRC8D |
| rs12064668 | height | 0.0000134 | 0.0000901 | 0.7464368 | 0.9958621 | 0.936092 | LRRC8C LRRC8D |
| rs10737711 | height | 0.0000134 | 0.0000901 | 0.7464368 | 0.9958621 | 0.936092 | LRRC8C LRRC8D |
| rs6688061 | height | 0.0000194 | 0.0000907 | 0.8337931 | 0.997931 | 0.9225287 | LRRC8C LRRC8D |
| rs55668929 | height | 0.0000194 | 0.0000907 | 0.8337931 | 0.997931 | 0.9225287 | LRRC8C LRRC8D |
| rs1544926 | height | 0.0000763 | 0.0000034 | 0.9832184 | 0.4181609 | 0.9554023 | COL23A1 |
| rs10903280 | height | 0.0000765 | 0.0000042 | 0.983908 | 0.4586207 | 0.9514943 | COL23A1 |
| rs72819017 | height | 0.000076 | 0.0000032 | 0.982069 | 0.3947126 | 0.9549425 | COL23A1 |
| rs57478839 | height | 0.0000765 | 0.0000042 | 0.983908 | 0.4586207 | 0.9514943 | COL23A1 |
| rs35519588 | height | 0.0000693 | 0.000001 | 0.9643678 | 0.1514943 | 0.9455172 | COL23A1 |
| rs890802 | height | 0.0000763 | 0.0000034 | 0.9832184 | 0.4181609 | 0.9554023 | COL23A1 |
| rs61739424 | height | 0.0000693 | 0.000001 | 0.9643678 | 0.1514943 | 0.9455172 | COL23A1 |
| rs2913847 | height | 0.0000763 | 0.0000034 | 0.9832184 | 0.4181609 | 0.9554023 | COL23A1 |
| rs1388358 | height | 0.0001218 | 0.0000114 | 0.9931034 | 0.5034483 | 0.9445977 | ANXA1 RORB |
| rs2172162 | height | 0.0001218 | 0.0000114 | 0.9931034 | 0.5034483 | 0.9445977 | ANXA1 RORB |
| rs76907378 | height | 0.0001169 | 0.0000098 | 0.9903448 | 0.4004598 | 0.9388506 | ANXA1 RORB |
| rs7020553 | height | 0.0001169 | 0.0000098 | 0.9903448 | 0.4004598 | 0.9388506 | ANXA1 RORB |
| rs11143787 | height | 0.0001169 | 0.0000098 | 0.9903448 | 0.4004598 | 0.9388506 | ANXA1 RORB |
| rs11021216 | height | 0.0000955 | 0.0000087 | 0.9848276 | 0.5333333 | 0.9397701 | SESN3 FAM76B |
| rs11021219 | height | 0.0000955 | 0.0000087 | 0.9848276 | 0.5333333 | 0.9397701 | SESN3 FAM76B |
| rs10831376 | height | 0.0000857 | 0.0000076 | 0.9751724 | 0.4908046 | 0.923908 | SESN3 FAM76B |
| rs2636063 | height | 0.0000026 | 0.0000664 | 0.1724138 | 0.9326437 | 0.9135632 | FAM189A1 |
| rs2672705 | height | 0.0000048 | 0.0000669 | 0.3411494 | 0.9505747 | 0.9137931 | FAM189A1 |
| rs79512105 | height | 0.0000025 | 0.0000655 | 0.1868966 | 0.9250575 | 0.9032184 | FAM189A1 |
| rs77268983 | height | 0.0000938 | 0.0000028 | 0.9657471 | 0.2570115 | 0.9434483 | SMAD6 SMAD3 |
| rs12593707 | height | 0.0000938 | 0.0000028 | 0.9657471 | 0.2570115 | 0.9434483 | SMAD6 SMAD3 |
| rs1895886 | height | 0.0000653 | 0.0000073 | 0.9845977 | 0.6068966 | 0.9041379 | FAM69C CNDP2 |
| rs747175 | height | 0.0000738 | 0.0000086 | 0.9931034 | 0.6682759 | 0.9241379 | FAM69C CNDP2 |
| rs1365249 | height | 0.0000653 | 0.0000073 | 0.9845977 | 0.6068966 | 0.9041379 | CNDP2 |
| rs2278161 | height | 0.0000653 | 0.0000073 | 0.9845977 | 0.6068966 | 0.9041379 | CNDP2 |
| rs653004 | height | 0.0000906 | 0.0000026 | 0.9712644 | 0.2965517 | 0.9298851 | SIK1 FLJ41733 LINC00322 |
| rs4818928 | height | 0.0000906 | 0.0000026 | 0.9712644 | 0.2965517 | 0.9298851 | SIK1 FLJ41733 LINC00322 |
| rs1003792 | height | 0.0000906 | 0.0000026 | 0.9712644 | 0.2965517 | 0.9298851 | SIK1 FLJ41733 LINC00322 |
| rs12627203 | height | 0.0000906 | 0.0000026 | 0.9712644 | 0.2965517 | 0.9298851 | SIK1 FLJ41733 LINC00322 |
| rs2071931 | WHR | 0.0000898 | 0.0002293 | 0.9783908 | 1 | 0.9229885 | H6PD |

Table A.2 (cont'd)

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | $\text{PPM}_{\sigma^2_{g_{j^*}}}$ [c] | $\text{PPF}_{\sigma^2_{g_{j^*}}}$ [c] | $\text{PPDiff}_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs7517657 | WHR | 0.0000451 | 0.0002563 | 0.9144828 | 0.9993103 | 0.962069 | LOC284688 METTL11B |
| rs1332955 | WHR | 0.0000647 | 0.0002944 | 0.9696552 | 0.9997701 | 0.9726437 | LOC284688 METTL11B |
| rs80290375 | WHR | 0.0000744 | 0.0003468 | 0.9374713 | 0.9995402 | 0.9390805 | LOC284688 GORAB |
| rs6427245 | WHR | 0.0000647 | 0.0002944 | 0.9696552 | 0.9997701 | 0.9726437 | LOC284688 GORAB |
| rs7537355 | WHR | 0.0001064 | 0.00041 | 0.9887356 | 0.9997701 | 0.9650575 | LOC284688 GORAB |
| rs12139302 | WHR | 0.0001064 | 0.00041 | 0.9887356 | 0.9997701 | 0.9650575 | LOC284688 GORAB |
| rs7522128 | WHR | 0.0000815 | 0.0003638 | 0.96 | 0.9997701 | 0.9537931 | GORAB LOC284688 |
| rs61838774 | WHR | 0.0000013 | 0.0001621 | 0.1308046 | 0.9751724 | 0.965977 | LYPLAL1 RNU5F-1 |
| rs2168333 | WHR | 0.0000075 | 0.0006972 | 0.4475862 | 1 | 1 | LYPLAL1 RNU5F-1 |
| rs12747505 | WHR | 0.0000029 | 0.0002979 | 0.2473563 | 0.9965517 | 0.9908046 | LYPLAL1 RNU5F-1 |
| rs12724708 | WHR | 0.0000058 | 0.0004579 | 0.3565517 | 0.9983908 | 0.9931034 | LYPLAL1 RNU5F-1 |
| rs6541227 | WHR | 0.000007 | 0.0006346 | 0.4273563 | 1 | 0.9997701 | LYPLAL1 RNU5F-1 |
| rs17005614 | WHR | 0.000002 | 0.0002048 | 0.1797701 | 0.9641379 | 0.9448276 | LYPLAL1 RNU5F-1 |
| rs12030989 | WHR | 0.000007 | 0.0006346 | 0.4273563 | 1 | 0.9997701 | LYPLAL1 RNU5F-1 |
| rs2820436 | WHR | 0.0000076 | 0.0007028 | 0.4574713 | 1 | 1 | LYPLAL1 RNU5F-1 |
| rs2605100 | WHR | 0.0000077 | 0.0007022 | 0.4671264 | 1 | 1 | LYPLAL1 RNU5F-1 |
| rs1538749 | WHR | 0.0000079 | 0.0007128 | 0.4802299 | 1 | 1 | LYPLAL1 RNU5F-1 |
| rs12022722 | WHR | 0.000008 | 0.0007184 | 0.4898851 | 1 | 1 | LYPLAL1 RNU5F-1 |
| rs2605110 | WHR | 0.0000077 | 0.0007022 | 0.4671264 | 1 | 1 | LYPLAL1 RNU5F-1 |
| rs2061154 | WHR | 0.000008 | 0.0007184 | 0.4898851 | 1 | 1 | LYPLAL1 RNU5F-1 |
| rs2791545 | WHR | 0.000013 | 0.0004375 | 0.5735632 | 0.9993103 | 0.9942529 | LYPLAL1 RNU5F-1 |
| rs3923113 | WHR | 0.0000126 | 0.0010649 | 0.5643678 | 1 | 1 | GRB14 COBLL1 |
| rs10195252 | WHR | 0.0000126 | 0.0010649 | 0.5643678 | 1 | 1 | COBLL1 GRB14 |
| rs1128249 | WHR | 0.0000132 | 0.0010696 | 0.6135632 | 1 | 1 | COBLL1 GRB14 |
| rs75297654 | WHR | 0.0000085 | 0.0007424 | 0.403908 | 1 | 1 | COBLL1 |
| rs17244632 | WHR | 0.0000096 | 0.0007438 | 0.4790805 | 1 | 1 | COBLL1 |
| rs13067911 | WHR | 0.0000027 | 0.0001533 | 0.1885057 | 0.9977011 | 0.9931034 | PPARG TSEN2 |
| rs4684859 | WHR | 0.0000039 | 0.000157 | 0.3298851 | 0.9983908 | 0.9937931 | PPARG TSEN2 |
| rs73029213 | WHR | 0.0000024 | 0.0001479 | 0.3204598 | 0.9958621 | 0.9914943 | PPARG TSEN2 |
| rs17036788 | WHR | 0.0000059 | 0.0001633 | 0.5314943 | 0.9990805 | 0.9935632 | PPARG TSEN2 |
| rs6795735 | WHR | 0.000017 | 0.0005836 | 0.7310345 | 1 | 1 | ADAMTS9-AS2 |
| rs4132228 | WHR | 0.0000164 | 0.0005834 | 0.6977011 | 1 | 1 | ADAMTS9-AS2 MIR548A2 |
| rs4607103 | WHR | 0.0000195 | 0.0005915 | 0.8091954 | 1 | 1 | ADAMTS9-AS2 MIR548A2 |
| rs7433808 | WHR | 0.0000195 | 0.0005915 | 0.8091954 | 1 | 1 | ADAMTS9-AS2 MIR548A2 |
| rs7638389 | WHR | 0.0000195 | 0.0005915 | 0.8091954 | 1 | 1 | ADAMTS9-AS2 MIR548A2 |
| rs2194094 | WHR | 0.0000164 | 0.0005834 | 0.6977011 | 1 | 1 | ADAMTS9-AS2 MIR548A2 |
| rs60960425 | WHR | 0.0000171 | 0.000287 | 0.7967816 | 0.9951724 | 0.983908 | RPL32P3 |
| rs79763737 | WHR | 0.0000015 | 0.0002008 | 0.111954 | 0.9708046 | 0.9590805 | EFCAB12 |
| rs16861373 | WHR | 0.0000066 | 0.0004297 | 0.3889655 | 0.9995402 | 0.9947126 | PLXND1 |

Table A.2 (cont'd)

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | $\text{PPM}_{\sigma^2_{g_{j^*}}}$ [c] | $\text{PPF}_{\sigma^2_{g_{j^*}}}$ [c] | $\text{PPDiff}_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs79870266 | WHR | 0.0000066 | 0.0004297 | 0.3889655 | 0.9995402 | 0.9947126 | PLXND1 |
| rs9833879 | WHR | 0.0000066 | 0.0004297 | 0.3889655 | 0.9995402 | 0.9947126 | PLXND1 TMCC1 |
| rs2306374 | WHR | 0.000001 | 0.0000661 | 0.1227586 | 0.9222989 | 0.9022989 | MRAS |
| rs9818870 | WHR | 0.000001 | 0.0000661 | 0.1227586 | 0.9222989 | 0.9022989 | MRAS |
| rs4301033 | WHR | 0.0000008 | 0.0000786 | 0.0763218 | 0.9501149 | 0.9367816 | TSC22D2 LOC646903 |
| rs73162462 | WHR | 0.0000008 | 0.0000786 | 0.0763218 | 0.9501149 | 0.9367816 | TSC22D2 LOC646903 |
| rs62271364 | WHR | 0.0000008 | 0.0000786 | 0.0763218 | 0.9501149 | 0.9367816 | TSC22D2 LOC646903 |
| rs4450871 | WHR | 0.0000002 | 0.0001683 | 0.0266667 | 1 | 1 | CYTL1 MSX1 |
| rs13133548 | WHR | 0.0000019 | 0.0002404 | 0.1754023 | 0.9687356 | 0.9558621 | FAM13A |
| rs3822072 | WHR | 0.0000019 | 0.0002404 | 0.1754023 | 0.9687356 | 0.9558621 | FAM13A |
| rs13147493 | WHR | 0.0000019 | 0.0002404 | 0.1754023 | 0.9687356 | 0.9558621 | FAM13A |
| rs974801 | WHR | 0.000005 | 0.0000716 | 0.3668966 | 0.9616092 | 0.9064368 | TET2 |
| rs9884482 | WHR | 0.000005 | 0.0000716 | 0.3668966 | 0.9616092 | 0.9064368 | TET2 |
| rs2285720 | WHR | 0.000005 | 0.0000716 | 0.3668966 | 0.9616092 | 0.9064368 | TET2 |
| rs10488872 | WHR | 0.0000742 | 0.0000025 | 0.9774713 | 0.2917241 | 0.9448276 | PAPSS1 DKK2 |
| rs10000444 | WHR | 0.0000742 | 0.0000025 | 0.9774713 | 0.2917241 | 0.9448276 | PAPSS1 DKK2 |
| rs17037679 | WHR | 0.0000742 | 0.0000025 | 0.9774713 | 0.2917241 | 0.9448276 | PAPSS1 DKK2 |
| rs6818614 | WHR | 0.0000742 | 0.0000025 | 0.9774713 | 0.2917241 | 0.9448276 | PAPSS1 DKK2 |
| rs28399230 | WHR | 0.0000742 | 0.0000025 | 0.9774713 | 0.2917241 | 0.9448276 | PAPSS1 DKK2 |
| rs13156948 | WHR | 0.0000016 | 0.000066 | 0.0788506 | 0.9696552 | 0.9570115 | IRX1 LOC340094 |
| rs6867983 | WHR | 0.0000192 | 0.0003815 | 0.4404598 | 1 | 0.9981609 | MAP3K1 ANKRD55 |
| rs3936510 | WHR | 0.0000188 | 0.0003812 | 0.4091954 | 1 | 0.9981609 | MAP3K1 ANKRD55 |
| rs9687846 | WHR | 0.0000188 | 0.0003812 | 0.4091954 | 1 | 0.9981609 | MAP3K1 ANKRD55 |
| rs37521 | WHR | 0.0000033 | 0.000093 | 0.2901149 | 0.9645977 | 0.9243678 | PLK2 ACTBL2 |
| rs10073521 | WHR | 0.0000139 | 0.0001164 | 0.4147126 | 0.9701149 | 0.9002299 | TNFAIP8 |
| rs17145265 | WHR | 0.0000163 | 0.0001422 | 0.5262069 | 0.9935632 | 0.9363218 | TNFAIP8 |
| rs55682871 | WHR | 0.0000156 | 0.0001184 | 0.4924138 | 0.9811494 | 0.9050575 | TNFAIP8 |
| rs7704120 | WHR | 0.0000049 | 0.0001374 | 0.476092 | 0.9983908 | 0.9912644 | STC2 NKX2-5 |
| rs6879065 | WHR | 0.0000049 | 0.0001374 | 0.476092 | 0.9983908 | 0.9912644 | STC2 NKX2-5 |
| rs1023617 | WHR | 0.0000049 | 0.0001374 | 0.476092 | 0.9983908 | 0.9912644 | STC2 NKX2-5 |
| rs3836828 | WHR | 0.0000029 | 0.0001343 | 0.3045977 | 0.9981609 | 0.9908046 | STC2 |
| rs9502498 | WHR | 0.0000141 | 0.0001057 | 0.4783908 | 0.9954023 | 0.9301149 | RREB1 LY86 |
| rs4960245 | WHR | 0.0000136 | 0.0001 | 0.4445977 | 0.9931034 | 0.9133333 | RREB1 LY86 |
| Affx-37047069 | WHR | 0.0000133 | 0.0000978 | 0.4186207 | 0.9921839 | 0.9057471 | RREB1 LY86 |
| rs56005336 | WHR | 0.0000853 | 0.0006116 | 0.9726437 | 1 | 0.9503448 | GRM4 HMGA1 |
| rs76412020 | WHR | 0.0000855 | 0.0006109 | 0.9703448 | 1 | 0.9494253 | GRM4 HMGA1 |
| rs114355919 | WHR | 0.0000853 | 0.0006116 | 0.9726437 | 1 | 0.9503448 | GRM4 HMGA1 |
| rs7742369 | WHR | 0.0000853 | 0.0006116 | 0.9726437 | 1 | 0.9503448 | GRM4 HMGA1 |
| rs10947487 | WHR | 0.0000734 | 0.0005684 | 0.9452874 | 1 | 0.9349425 | HMGA1 GRM4 |

Table A.2 (cont'd)

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | $PPM_{\sigma^2_{g_{j^*}}}$ [c] | $PPF_{\sigma^2_{g_{j^*}}}$ [c] | $PPDiff_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs117525671 | WHR | 0.0000734 | 0.0005684 | 0.9452874 | 1 | 0.9349425 | HMGA1 GRM4 |
| rs1776897 | WHR | 0.000087 | 0.0006141 | 0.9758621 | 1 | 0.9503448 | HMGA1 GRM4 |
| rs2780226 | WHR | 0.000087 | 0.0006141 | 0.9758621 | 1 | 0.9503448 | HMGA1 GRM4 |
| rs114344942 | WHR | 0.0000853 | 0.0006116 | 0.9726437 | 1 | 0.9503448 | HMGA1 |
| rs139876191 | WHR | 0.0000734 | 0.0005684 | 0.9452874 | 1 | 0.9349425 | HMGA1 |
| rs35381162 | WHR | 0.0000734 | 0.0005684 | 0.9452874 | 1 | 0.9349425 | HMGA1 |
| rs1150781 | WHR | 0.000087 | 0.0006141 | 0.9758621 | 1 | 0.9503448 | C6orf1 |
| rs6918981 | WHR | 0.0000767 | 0.0005756 | 0.9517241 | 1 | 0.9363218 | C6orf1 NUDT3 |
| rs4711750 | WHR | 0.0000625 | 0.0022238 | 0.9533333 | 1 | 1 | VEGFA LOC100132354 |
| rs6899540 | WHR | 0.0000455 | 0.0011492 | 0.8533333 | 1 | 1 | VEGFA LOC100132354 |
| rs6905288 | WHR | 0.0000567 | 0.0022249 | 0.92 | 1 | 1 | VEGFA LOC100132354 |
| rs9472126 | WHR | 0.0000455 | 0.0011492 | 0.8533333 | 1 | 1 | VEGFA LOC100132354 |
| rs1885659 | WHR | 0.0000194 | 0.0004076 | 0.5958621 | 0.9997701 | 0.9744828 | VEGFA LOC100132354 |
| rs2396081 | WHR | 0.0000231 | 0.0007352 | 0.6774713 | 1 | 1 | VEGFA LOC100132354 |
| rs12526378 | WHR | 0.0000194 | 0.0004076 | 0.5958621 | 0.9997701 | 0.9744828 | VEGFA LOC100132354 |
| rs36184164 | WHR | 0.0000058 | 0.000163 | 0.3032184 | 0.9864368 | 0.9370115 | VEGFA LOC100132354 |
| rs4236084 | WHR | 0.0000194 | 0.0004076 | 0.5958621 | 0.9997701 | 0.9744828 | VEGFA LOC100132354 |
| rs10046368 | WHR | 0.0000005 | 0.00042 | 0.0604598 | 1 | 1 | VEGFA LOC100132354 |
| rs17789218 | WHR | 0.0000259 | 0.0002546 | 0.5377011 | 0.9825287 | 0.9121839 | SIM1 LOC728012 |
| rs2503097 | WHR | 0.0000434 | 0.000459 | 0.8036782 | 1 | 1 | SIM1 LOC728012 |
| rs743011 | WHR | 0.0000259 | 0.0002546 | 0.5377011 | 0.9825287 | 0.9121839 | SIM1 LOC728012 |
| rs2073267 | WHR | 0.0000434 | 0.000459 | 0.8036782 | 1 | 1 | SIM1 LOC728012 |
| rs7756047 | WHR | 0.0000431 | 0.000418 | 0.7921839 | 1 | 0.9990805 | SIM1 LOC728012 |
| rs6937293 | WHR | 0.0000457 | 0.0004656 | 0.8390805 | 1 | 1 | SIM1 LOC728012 |
| rs972275 | WHR | 0.0000301 | 0.0006924 | 0.7956322 | 1 | 1 | CENPW RSPO3 |
| rs2800725 | WHR | 0.0000301 | 0.0006924 | 0.7956322 | 1 | 1 | CENPW RSPO3 |
| rs2800734 | WHR | 0.0000289 | 0.0006845 | 0.757931 | 1 | 1 | CENPW RSPO3 |
| rs1936799 | WHR | 0.0000306 | 0.0006836 | 0.7983908 | 1 | 1 | CENPW RSPO3 |
| rs1936801 | WHR | 0.0000284 | 0.0006838 | 0.7462069 | 1 | 1 | CENPW RSPO3 |
| rs9491696 | WHR | 0.0000284 | 0.0006838 | 0.7462069 | 1 | 1 | RSPO3 |
| rs2745353 | WHR | 0.0000284 | 0.0006838 | 0.7462069 | 1 | 1 | RSPO3 |
| rs6932207 | WHR | 0.0000284 | 0.0006838 | 0.7462069 | 1 | 1 | RSPO3 |
| rs41285262 | WHR | 0.0000293 | 0.0006924 | 0.7671264 | 1 | 1 | RSPO3 |
| rs1892172 | WHR | 0.0000284 | 0.0006838 | 0.7462069 | 1 | 1 | RSPO3 |
| rs13202608 | WHR | 0.0000267 | 0.0006791 | 0.6834483 | 1 | 1 | RSPO3 |
| rs4620145 | WHR | 0.0000284 | 0.0006838 | 0.7462069 | 1 | 1 | RSPO3 |
| rs6569474 | WHR | 0.0000284 | 0.0006838 | 0.7462069 | 1 | 1 | RSPO3 |
| rs72961013 | WHR | 0.0003262 | 0.0018148 | 1 | 1 | 1 | RNF146 RSPO3 |
| rs7766444 | WHR | 0.0000109 | 0.0003721 | 0.4668966 | 0.997931 | 0.9917241 | LOC645434 LOC100132735 |

Table A.2 (cont'd)

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | $PPM_{\sigma^2_{g_{j^*}}}$ [c] | $PPF_{\sigma^2_{g_{j^*}}}$ [c] | $PPDiff_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs9376422 | WHR | 0.0000107 | 0.0004179 | 0.5241379 | 1 | 1 | LOC645434 LOC100132735 |
| rs74623604 | WHR | 0.0000107 | 0.0004179 | 0.5241379 | 1 | 1 | LOC645434 LOC100132735 |
| rs2908521 | WHR | 0.0000103 | 0.0004166 | 0.497931 | 1 | 1 | LOC645434 LOC100132735 |
| rs651837 | WHR | 0.0000103 | 0.0004166 | 0.497931 | 1 | 1 | LOC645434 LOC100132735 |
| rs632057 | WHR | 0.0000105 | 0.0004162 | 0.5285057 | 1 | 1 | LOC645434 LOC100132735 |
| rs668459 | WHR | 0.0000103 | 0.0004166 | 0.497931 | 1 | 1 | LOC645434 LOC100132735 |
| rs72976928 | WHR | 0.0000103 | 0.0004166 | 0.497931 | 1 | 1 | LOC645434 LOC100132735 |
| rs628751 | WHR | 0.0000086 | 0.0003786 | 0.4236782 | 1 | 0.9993103 | LOC645434 LOC100132735 |
| rs592423 | WHR | 0.0000093 | 0.0003973 | 0.4496552 | 1 | 0.9993103 | LOC645434 LOC100132735 |
| rs6570354 | WHR | 0.0000091 | 0.0003796 | 0.4696552 | 1 | 0.9993103 | LOC645434 LOC100132735 |
| rs12526447 | WHR | 0.000002 | 0.0000758 | 0.1816092 | 0.9806897 | 0.9616092 | ESR1 |
| rs7772579 | WHR | 0.0000078 | 0.0000771 | 0.3942529 | 0.9873563 | 0.9344828 | ESR1 |
| rs1999805 | WHR | 0.0000019 | 0.0000756 | 0.1606897 | 0.9788506 | 0.9611494 | ESR1 |
| rs2152750 | WHR | 0.0000019 | 0.0000756 | 0.1606897 | 0.9788506 | 0.9611494 | ESR1 |
| rs1361024 | WHR | 0.0000022 | 0.000076 | 0.2034483 | 0.9818391 | 0.962069 | ESR1 |
| rs2504069 | WHR | 0.0000099 | 0.0000913 | 0.4554023 | 0.991954 | 0.9452874 | ESR1 |
| rs1055144 | WHR | 0.0000032 | 0.0003506 | 0.2645977 | 1 | 0.9988506 | MIR148A NPVF RNU6-16P |
| rs12700666 | WHR | 0.0000032 | 0.0003506 | 0.2645977 | 1 | 0.9988506 | MIR148A RNU6-16P |
| rs12700667 | WHR | 0.0000032 | 0.0003506 | 0.2645977 | 1 | 0.9988506 | MIR148A RNU6-16P |
| rs73068463 | WHR | 0.0000068 | 0.0004224 | 0.4613793 | 1 | 1 | SNX10 |
| rs74979045 | WHR | 0.0000046 | 0.0004204 | 0.3036782 | 1 | 1 | SNX10 |
| rs1534696 | WHR | 0.000005 | 0.0004217 | 0.3416092 | 1 | 1 | SNX10 |
| rs7798433 | WHR | 0.0000046 | 0.0004204 | 0.3036782 | 1 | 1 | SNX10 |
| rs1358503 | WHR | 0.0000021 | 0.0000716 | 0.3089655 | 0.9885057 | 0.9662069 | SEMA3C HGF |
| rs35736598 | WHR | 0.0000015 | 0.0000669 | 0.2282759 | 0.9816092 | 0.9581609 | SEMA3C HGF |
| Affx-30952281 | WHR | 0.0000021 | 0.0000716 | 0.3089655 | 0.9885057 | 0.9662069 | SEMA3C HGF |
| rs10091014 | WHR | 0.000009 | 0.0002402 | 0.5581609 | 0.9954023 | 0.9777011 | NKX2-6 STC1 |
| rs568890 | WHR | 0.0000129 | 0.0003106 | 0.8089655 | 1 | 1 | NKX2-6 STC1 |
| rs6983481 | WHR | 0.0000045 | 0.0002015 | 0.2570115 | 0.9634483 | 0.9333333 | NKX2-6 STC1 |
| rs67846476 | WHR | 0.000009 | 0.0002402 | 0.5581609 | 0.9954023 | 0.9777011 | NKX2-6 STC1 |
| rs445114 | WHR | 0.0000065 | 0.0000898 | 0.3170115 | 0.9793103 | 0.916092 | LOC727677 POU5F1B PCAT1 |
| rs622856 | WHR | 0.0000065 | 0.0000898 | 0.3170115 | 0.9793103 | 0.916092 | LOC727677 POU5F1B PCAT1 |
| rs444318 | WHR | 0.0000081 | 0.0000953 | 0.4547126 | 0.9889655 | 0.9326437 | LOC727677 POU5F1B PCAT1 |
| rs17464492 | WHR | 0.0000065 | 0.0000898 | 0.3170115 | 0.9793103 | 0.916092 | LOC727677 POU5F1B PCAT1 |
| rs12541832 | WHR | 0.0000065 | 0.0000898 | 0.3170115 | 0.9793103 | 0.916092 | LOC727677 POU5F1B PCAT1 |
| rs13281615 | WHR | 0.000007 | 0.0000919 | 0.3597701 | 0.9827586 | 0.9236782 | LOC727677 POU5F1B PCAT1 |
| rs11783615 | WHR | 0.0000072 | 0.0000919 | 0.3935632 | 0.9845977 | 0.9245977 | LOC727677 POU5F1B PCAT1 |
| rs10991415 | WHR | 0.0000026 | 0.0001057 | 0.1963218 | 0.9590805 | 0.9347126 | ABCA1 |
| rs2472377 | WHR | 0.000012 | 0.000134 | 0.5333333 | 0.991954 | 0.9636782 | ABCA1 |

Table A.2 (cont'd)

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | $PPM_{\sigma^2_{g_{j^*}}}$ [c] | $PPF_{\sigma^2_{g_{j^*}}}$ [c] | $PPDiff_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs2515609 | WHR | 0.0000023 | 0.0001059 | 0.162069 | 0.957931 | 0.9354023 | ABCA1 |
| rs10991417 | WHR | 0.0000048 | 0.0001225 | 0.3388506 | 0.9862069 | 0.9664368 | ABCA1 |
| rs62568211 | WHR | 0.0000193 | 0.000143 | 0.5604598 | 0.9885057 | 0.9537931 | ABCA1 |
| rs10760322 | WHR | 0.0000027 | 0.0000812 | 0.2818391 | 0.9857471 | 0.9675862 | LHX2 NEK6 |
| rs943484 | WHR | 0.0000042 | 0.000082 | 0.3450575 | 0.9882759 | 0.9636782 | LHX2 NEK6 |
| rs76204549 | WHR | 0.0000026 | 0.0000792 | 0.2103448 | 0.9827586 | 0.9636782 | LHX2 NEK6 |
| rs117790707 | WHR | 0.0000002 | 0.0000771 | 0.0294253 | 0.9728736 | 0.965977 | LHX2 NEK6 |
| rs12002771 | WHR | 0.0000038 | 0.0000811 | 0.3131034 | 0.9862069 | 0.9643678 | LHX2 NEK6 |
| rs10986225 | WHR | 0.000003 | 0.0000803 | 0.242069 | 0.9841379 | 0.9641379 | LHX2 NEK6 |
| rs1998951 | WHR | 0.0000046 | 0.0000823 | 0.3813793 | 0.9889655 | 0.9634483 | LHX2 NEK6 |
| rs78393198 | WHR | 0.0000026 | 0.0000792 | 0.2103448 | 0.9827586 | 0.9636782 | LHX2 NEK6 |
| Affx-2640174 | WHR | 0.000003 | 0.0001563 | 0.2595402 | 0.971954 | 0.948046 | FGFR2 MIR5694 |
| rs2244506 | WHR | 0.0000101 | 0.0002068 | 0.4526437 | 0.9977011 | 0.9845977 | FGFR2 MIR5694 |
| rs7907754 | WHR | 0.000003 | 0.0001563 | 0.2595402 | 0.971954 | 0.948046 | FGFR2 MIR5694 |
| rs2254069 | WHR | 0.000003 | 0.0001563 | 0.2595402 | 0.971954 | 0.948046 | FGFR2 MIR5694 |
| rs4436487 | WHR | 0.000003 | 0.0001563 | 0.2595402 | 0.971954 | 0.948046 | FGFR2 MIR5694 |
| rs56297542 | WHR | 0.0000073 | 0.0000971 | 0.3986207 | 0.9563218 | 0.9009195 | FGFR2 MIR5694 |
| rs1907282 | WHR | 0.0000022 | 0.0000957 | 0.1898851 | 0.9404598 | 0.9068966 | FGFR2 MIR5694 |
| rs7089185 | WHR | 0.0000022 | 0.0000957 | 0.1898851 | 0.9404598 | 0.9068966 | FGFR2 MIR5694 |
| rs10788149 | WHR | 0.0000058 | 0.0000981 | 0.3445977 | 0.9542529 | 0.905977 | FGFR2 MIR5694 |
| rs578270 | WHR | 0.0000182 | 0.0001261 | 0.7381609 | 0.9887356 | 0.9252874 | FRMD8 |
| rs2073800 | WHR | 0.0000182 | 0.0001261 | 0.7381609 | 0.9887356 | 0.9252874 | FRMD8 |
| rs512715 | WHR | 0.0000249 | 0.0001482 | 0.8351724 | 0.9990805 | 0.9478161 | NEAT1 |
| rs673753 | WHR | 0.0000183 | 0.0001264 | 0.7450575 | 0.9891954 | 0.9264368 | NEAT1 MIR612 |
| rs1784859 | WHR | 0.0000182 | 0.0001261 | 0.7381609 | 0.9887356 | 0.9252874 | NEAT1 MALAT1 MIR612 |
| rs1783210 | WHR | 0.0000182 | 0.0001261 | 0.7381609 | 0.9887356 | 0.9252874 | MALAT1 MIR612 |
| rs1784100 | WHR | 0.0000182 | 0.0001261 | 0.7381609 | 0.9887356 | 0.9252874 | MALAT1 MIR612 |
| rs11263641 | WHR | 0.0000207 | 0.0002343 | 0.7227586 | 0.9997701 | 0.9908046 | CCND1 MYEOV |
| rs74471298 | WHR | 0.0000075 | 0.0001749 | 0.3406897 | 0.9963218 | 0.9777011 | CCND1 MYEOV |
| rs10160464 | WHR | 0.0000075 | 0.0001727 | 0.3533333 | 0.9963218 | 0.976092 | LOC100505834 CCND1 MYEOV |
| rs4980785 | WHR | 0.0000083 | 0.0001809 | 0.4335632 | 0.9990805 | 0.9832184 | LOC100505834 CCND1 MYEOV |
| rs7105934 | WHR | 0.0000079 | 0.00018 | 0.3885057 | 0.9990805 | 0.9832184 | LOC100505834 CCND1 MYEOV |
| rs11233117 | WHR | 0.0000084 | 0.0000803 | 0.4733333 | 0.9765517 | 0.9112644 | ANO1 FGF3 |
| rs2343876 | WHR | 0.0000243 | 0.0003416 | 0.5390805 | 1 | 0.9914943 | SSPN ITPR2 |
| rs112251480 | WHR | 0.0000229 | 0.0003321 | 0.4482759 | 1 | 0.9885057 | SSPN ITPR2 |
| rs718314 | WHR | 0.0000232 | 0.0003606 | 0.5234483 | 1 | 0.997931 | ITPR2 SSPN |
| rs931384 | WHR | 0.0000232 | 0.0003606 | 0.5234483 | 1 | 0.997931 | ITPR2 SSPN |
| rs2171522 | WHR | 0.0000241 | 0.0003647 | 0.5606897 | 1 | 0.9981609 | ITPR2 SSPN |
| rs10842713 | WHR | 0.0000233 | 0.0003561 | 0.5321839 | 1 | 0.9981609 | ITPR2 SSPN |

Table A.2 (cont'd)

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | $PPM_{\sigma^2_{g_{j^*}}}$ [c] | $PPF_{\sigma^2_{g_{j^*}}}$ [c] | $PPDiff_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs598322 | WHR | 0.0000156 | 0.0001329 | 0.6898851 | 0.9967816 | 0.9528736 | HCAR1 KNTC1 HCAR2 |
| rs4759364 | WHR | 0.0000175 | 0.000169 | 0.7521839 | 0.9990805 | 0.9756322 | HCAR1 KNTC1 HCAR2 |
| rs10847452 | WHR | 0.0000166 | 0.0001329 | 0.7222989 | 0.9974713 | 0.9494253 | HCAR1 KNTC1 HCAR2 |
| rs10773433 | WHR | 0.0000166 | 0.0001329 | 0.7222989 | 0.9974713 | 0.9494253 | HCAR1 KNTC1 HCAR2 |
| rs1798219 | WHR | 0.0000175 | 0.000169 | 0.7521839 | 0.9990805 | 0.9756322 | HCAR1 KNTC1 HCAR2 |
| rs586573 | WHR | 0.0000166 | 0.0001329 | 0.7222989 | 0.9974713 | 0.9494253 | HCAR1 KNTC1 HCAR2 |
| rs118091133 | WHR | 0.0000166 | 0.0001329 | 0.7222989 | 0.9974713 | 0.9494253 | HCAR1 HCAR3 |
| rs1798192 | WHR | 0.0000166 | 0.0001329 | 0.7222989 | 0.9974713 | 0.9494253 | HCAR1 HCAR3 |
| rs1696352 | WHR | 0.0000166 | 0.0001329 | 0.7222989 | 0.9974713 | 0.9494253 | HCAR1 HCAR3 |
| Affx-7035109 | WHR | 0.0000166 | 0.0001329 | 0.7222989 | 0.9974713 | 0.9494253 | HCAR1 HCAR3 |
| rs71456771 | WHR | 0.0000175 | 0.000169 | 0.7521839 | 0.9990805 | 0.9756322 | HCAR1 HCAR3 |
| rs10847570 | WHR | 0.0000156 | 0.0001367 | 0.677931 | 0.9967816 | 0.9443678 | HCAR1 HCAR3 |
| rs1316952 | WHR | 0.0000438 | 0.0006033 | 0.8462069 | 1 | 0.9997701 | DNAH10 |
| rs11057396 | WHR | 0.0000438 | 0.0006033 | 0.8462069 | 1 | 0.9997701 | DNAH10 CCDC92 |
| rs11057401 | WHR | 0.0000438 | 0.0006033 | 0.8462069 | 1 | 0.9997701 | CCDC92 |
| rs10773048 | WHR | 0.0000315 | 0.0005557 | 0.7068966 | 1 | 0.9997701 | CCDC92 |
| rs4765127 | WHR | 0.0000438 | 0.0006033 | 0.8462069 | 1 | 0.9997701 | ZNF664 ZNF664-FAM101A |
| rs74551816 | WHR | 0.0001551 | 0.0000048 | 0.9818391 | 0.382069 | 0.9641379 | NOVA1 STXBP6 |
| rs12432376 | WHR | 0.0001741 | 0.0000074 | 1 | 0.5524138 | 0.9942529 | NOVA1 STXBP6 |
| rs11627916 | WHR | 0.0001741 | 0.0000074 | 1 | 0.5524138 | 0.9942529 | NOVA1 STXBP6 |
| rs7161009 | WHR | 0.0001673 | 0.0000056 | 0.9997701 | 0.437931 | 0.9926437 | NOVA1 STXBP6 |
| rs4983099 | WHR | 0.0001546 | 0.000004 | 0.9811494 | 0.2878161 | 0.9627586 | NOVA1 STXBP6 |
| rs8021667 | WHR | 0.0001673 | 0.0000056 | 0.9997701 | 0.437931 | 0.9926437 | NOVA1 STXBP6 |
| rs1955872 | WHR | 0.0001587 | 0.0000058 | 0.9850575 | 0.431954 | 0.9645977 | NOVA1 STXBP6 |
| rs116145925 | WHR | 0.0000014 | 0.0000747 | 0.1425287 | 0.9632184 | 0.9356322 | UBE2I |
| rs115466201 | WHR | 0.0000014 | 0.0000747 | 0.1425287 | 0.9632184 | 0.9356322 | UBE2I |
| rs2286973 | WHR | 0.0000339 | 0.0001226 | 0.8763218 | 0.9986207 | 0.916092 | CLEC16A |
| rs12930396 | WHR | 0.000004 | 0.0000937 | 0.2655172 | 0.9501149 | 0.9209195 | LITAF LOC388210 RMI2 |
| rs4131548 | WHR | 0.000004 | 0.0000937 | 0.2655172 | 0.9501149 | 0.9209195 | LITAF LOC388210 RMI2 |
| rs17777180 | WHR | 0.0000031 | 0.0005946 | 0.2905747 | 1 | 1 | CMIP |
| rs4889326 | WHR | 0.0000019 | 0.000591 | 0.2151724 | 1 | 1 | CMIP |
| rs2925979 | WHR | 0.0000016 | 0.0005902 | 0.177931 | 1 | 1 | CMIP |
| rs11865332 | WHR | 0.0000016 | 0.0005902 | 0.177931 | 1 | 1 | CMIP |
| rs9892297 | WHR | 0.0000262 | 0.0001187 | 0.8524138 | 0.9997701 | 0.9370115 | TNFSF12 POLR2A |
| rs62070804 | WHR | 0.0000004 | 0.0000887 | 0.0521839 | 0.9689655 | 0.9609195 | ABHD15 |
| rs3110647 | WHR | 0.0000023 | 0.0000731 | 0.2696552 | 0.9563218 | 0.9285057 | HNF1B DDX52 |
| rs34064336 | WHR | 0.0000007 | 0.0000694 | 0.0556322 | 0.9273563 | 0.9158621 | HNF1B DDX52 |
| rs4080890 | WHR | 0.0000153 | 0.0001631 | 0.5944828 | 0.9986207 | 0.9747126 | KCNJ2 |
| rs1605750 | WHR | 0.0000188 | 0.0001683 | 0.654023 | 0.9990805 | 0.9717241 | KCNJ2 |

Table A.2 (cont'd)

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | $\text{PPM}_{\sigma^2_{g_{j^*}}}$ [c] | $\text{PPF}_{\sigma^2_{g_{j^*}}}$ [c] | $\text{PPDiff}_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs16975820 | WHR | 0.0000119 | 0.0001567 | 0.4781609 | 0.9981609 | 0.9731034 | KCNJ2 |
| rs17779747 | WHR | 0.0000188 | 0.0001683 | 0.654023 | 0.9990805 | 0.9717241 | KCNJ2 |
| rs1594476 | WHR | 0.0000119 | 0.0001567 | 0.4781609 | 0.9981609 | 0.9731034 | KCNJ2 |
| rs3810068 | WHR | 0.0000026 | 0.0003594 | 0.174023 | 1 | 1 | EMILIN2 SMCHD1 |
| rs684320 | WHR | 0.0000048 | 0.0003602 | 0.3062069 | 1 | 0.9997701 | EMILIN2 |
| rs9954931 | WHR | 0.0000048 | 0.0003602 | 0.3062069 | 1 | 0.9997701 | EMILIN2 |
| rs679153 | WHR | 0.0000048 | 0.0003602 | 0.3062069 | 1 | 0.9997701 | EMILIN2 |
| rs623561 | WHR | 0.0000052 | 0.0003603 | 0.34 | 1 | 0.9997701 | EMILIN2 |
| rs4800269 | WHR | 0.0000093 | 0.0000589 | 0.577931 | 0.988046 | 0.9036782 | AQP4-AS1 AQP4 LOC728606 |
| rs12454712 | WHR | 0.0000087 | 0.0001024 | 0.3604598 | 0.9958621 | 0.9648276 | BCL2 |
| rs4940576 | WHR | 0.0000087 | 0.0001024 | 0.3604598 | 0.9958621 | 0.9648276 | BCL2 |
| rs11661511 | WHR | 0.0000087 | 0.0001024 | 0.3604598 | 0.9958621 | 0.9648276 | BCL2 |
| rs2288404 | WHR | 0.0000024 | 0.0000727 | 0.268046 | 0.9698851 | 0.934023 | INSR |
| rs1799815 | WHR | 0.0000006 | 0.0000702 | 0.0512644 | 0.9455172 | 0.9268966 | INSR |
| rs10408374 | WHR | 0.000001 | 0.0000713 | 0.1367816 | 0.9563218 | 0.9312644 | INSR |
| rs34465381 | WHR | 0.0000126 | 0.0000984 | 0.6216092 | 0.9947126 | 0.9236782 | HAUS8 CPAMD8 |
| rs7259285 | WHR | 0.0000182 | 0.0001711 | 0.7668966 | 1 | 0.9889655 | HAUS8 CPAMD8 |
| rs7260259 | WHR | 0.0000098 | 0.0000928 | 0.5264368 | 0.9903448 | 0.9163218 | HAUS8 |
| Affx-15620245 | WHR | 0.0000158 | 0.0001675 | 0.7227586 | 1 | 0.9885057 | HAUS8 |
| rs6512161 | WHR | 0.0000158 | 0.0001675 | 0.7227586 | 1 | 0.9885057 | HAUS8 |
| rs7259348 | WHR | 0.0000098 | 0.0000928 | 0.5264368 | 0.9903448 | 0.9163218 | HAUS8 |
| rs6102059 | WHR | 0.0000011 | 0.0000756 | 0.0827586 | 0.9588506 | 0.9303448 | MAFB LOC339568 |
| rs1936963 | WHR | 0.0000168 | 0.0002126 | 0.5498851 | 1 | 0.9749425 | TSHZ2 |
| rs2741366 | WHR | 0.0000135 | 0.0002118 | 0.4832184 | 1 | 0.9765517 | TSHZ2 |
| rs73140232 | WHR | 0.0000112 | 0.0001844 | 0.3977011 | 0.9990805 | 0.9462069 | TSHZ2 |
| rs2000339 | WHR | 0.000016 | 0.000208 | 0.483908 | 1 | 0.968046 | TSHZ2 |
| rs6013630 | WHR | 0.0000134 | 0.0001853 | 0.5002299 | 0.9990805 | 0.9455172 | TSHZ2 |
| rs2800999 | WHR | 0.0000201 | 0.0002224 | 0.6908046 | 1 | 0.9786207 | TSHZ2 |
| rs1293430 | WHR | 0.0000177 | 0.0001917 | 0.6374713 | 0.9993103 | 0.9450575 | TSHZ2 |
| rs2256596 | BMD | 0.0002546 | 0.0000714 | 1 | 0.9432184 | 0.9374713 | RREB1 |
| rs35742417 | BMD | 0.0002546 | 0.0000714 | 1 | 0.9432184 | 0.9374713 | RREB1 |
| rs2714341 | BMD | 0.0002347 | 0.0000711 | 0.9981609 | 0.9406897 | 0.9117241 | RREB1 SSR1 |
| rs9403141 | BMD | 0.0001358 | 0.0000005 | 0.9581609 | 0.0855172 | 0.9468966 | MCHR2 PRDM13 |
| rs7451306 | BMD | 0.0001358 | 0.0000005 | 0.9581609 | 0.0855172 | 0.9468966 | MCHR2 PRDM13 |
| rs17428220 | BMD | 0.0002729 | 0.0001293 | 0.9933333 | 0.8126437 | 0.9045977 | EVX1 HIBADH |
| rs776746 | BMD | 0.0000308 | 0.000213 | 0.7701149 | 1 | 0.9457471 | CYP3A5 |
| rs45442295 | BMD | 0.0000128 | 0.0001952 | 0.4517241 | 0.9995402 | 0.9344828 | CYP3A7 CYP3A7-CYP3AP1 |
| rs45446698 | BMD | 0.0000091 | 0.000193 | 0.177931 | 0.9995402 | 0.9342529 | CYP3A4 CYP3A7 |
| rs6945984 | BMD | 0.0000308 | 0.000213 | 0.7701149 | 1 | 0.9457471 | CYP3A4 CYP3A7 |

Table A.2 (cont'd)

| Focal SNP[a] | trait | $\hat{\sigma}^2_{g_{m_{j^*}}}$ [b] | $\hat{\sigma}^2_{g_{f_{j^*}}}$ [b] | $PPM_{\sigma^2_{g_{j^*}}}$ [c] | $PPF_{\sigma^2_{g_{j^*}}}$ [c] | $PPDiff_{\sigma^2_{g_{j^*}}}$ [d] | nearby genes[e] |
|---|---|---|---|---|---|---|---|
| rs12333983 | BMD | 0.0000308 | 0.000213 | 0.7701149 | 1 | 0.9457471 | CYP3A4 CYP3A7 |
| rs3735451 | BMD | 0.0000308 | 0.000213 | 0.7701149 | 1 | 0.9457471 | CYP3A4 |
| rs6956344 | BMD | 0.0000308 | 0.000213 | 0.7701149 | 1 | 0.9457471 | CYP3A4 |
| rs2242480 | BMD | 0.0000308 | 0.000213 | 0.7701149 | 1 | 0.9457471 | CYP3A4 |
| rs8176719 | BMD | 0.0006001 | 0.0000182 | 1 | 0.794023 | 1 | ABO |
| rs687621 | BMD | 0.0005978 | 0.0000176 | 1 | 0.7783908 | 1 | ABO |
| rs657152 | BMD | 0.0006001 | 0.0000182 | 1 | 0.794023 | 1 | ABO |
| rs514659 | BMD | 0.0005978 | 0.0000176 | 1 | 0.7783908 | 1 | ABO |
| rs643434 | BMD | 0.0006001 | 0.0000182 | 1 | 0.794023 | 1 | ABO |
| rs612169 | BMD | 0.0005978 | 0.0000176 | 1 | 0.7783908 | 1 | ABO |
| rs581107 | BMD | 0.0005984 | 0.0000115 | 1 | 0.623908 | 1 | ABO |
| rs505922 | BMD | 0.0005978 | 0.0000176 | 1 | 0.7783908 | 1 | ABO |
| rs507666 | BMD | 0.0005833 | 0.000007 | 1 | 0.4397701 | 1 | ABO |
| rs651007 | BMD | 0.0005833 | 0.000007 | 1 | 0.4397701 | 1 | ABO SURF6 |
| rs579459 | BMD | 0.0005833 | 0.000007 | 1 | 0.4397701 | 1 | ABO SURF6 |
| rs495828 | BMD | 0.0005833 | 0.000007 | 1 | 0.4397701 | 1 | ABO SURF6 |
| rs635634 | BMD | 0.0005833 | 0.000007 | 1 | 0.4397701 | 1 | ABO SURF6 |
| rs56196860 | BMD | 0.0001283 | 0.0000016 | 0.9590805 | 0.0643678 | 0.9365517 | FKBP4 |
| rs1038196 | BMD | 0.000028 | 0.0001236 | 0.8227586 | 0.997931 | 0.9165517 | HMGA2 |
| rs1351394 | BMD | 0.000028 | 0.0001236 | 0.8227586 | 0.997931 | 0.9165517 | HMGA2 |
| rs1042725 | BMD | 0.000028 | 0.0001236 | 0.8227586 | 0.997931 | 0.9165517 | HMGA2 |
| rs8756 | BMD | 0.000028 | 0.0001236 | 0.8227586 | 0.997931 | 0.9165517 | HMGA2 |
| rs12424086 | BMD | 0.0000265 | 0.0001235 | 0.7912644 | 0.9970115 | 0.9197701 | HMGA2 LLPH |
| rs34029815 | BMD | 0.0000265 | 0.0001235 | 0.7912644 | 0.9970115 | 0.9197701 | HMGA2 LLPH |
| rs947211 | BMI | 9.52E-05 | 0.0000153 | 0.9898851 | 0.7687356 | 0.9022989 | SLC41A1 RAB7L1 |
| rs1775146 | BMI | 9.52E-05 | 0.0000153 | 0.9898851 | 0.7687356 | 0.9022989 | SLC41A1 RAB7L1 |
| rs13119835 | BMI | 2.31E-05 | 0.000109 | 0.8822989 | 0.9983908 | 0.9057471 | NDST3 |
| rs4833565 | BMI | 2.31E-05 | 0.000109 | 0.8822989 | 0.9983908 | 0.9057471 | NDST3 |
| rs10781293 | BMI | 2.00E-06 | 0.0001038 | 0.1848276 | 0.9448276 | 0.9121839 | PIP5K1B |
| rs10869623 | BMI | 2.00E-06 | 0.0001038 | 0.1848276 | 0.9448276 | 0.9121839 | PIP5K1B |
| rs941714 | BMI | 6.10E-06 | 0.0000667 | 0.514023 | 0.9763218 | 0.9094253 | MIR656 MEG9 |
| rs3742407 | BMI | 7.00E-06 | 0.0000687 | 0.5871264 | 0.9841379 | 0.9151724 | MEG9 |
| rs2295654 | BMI | 7.00E-06 | 0.0000687 | 0.5871264 | 0.9841379 | 0.9151724 | MEG9 |
| rs4906037 | BMI | 6.10E-06 | 0.0000667 | 0.514023 | 0.9763218 | 0.9094253 | DIO3OS MEG9 |
| rs2400968 | BMI | 7.00E-06 | 0.0000687 | 0.5871264 | 0.9841379 | 0.9151724 | DIO3OS MEG9 |
| rs235348 | BMI | 2.30E-06 | 0.0000694 | 0.2967816 | 0.9604598 | 0.9275862 | TSPEAR UBE2G2 |
| rs690333 | BMI | 1.16E-05 | 0.0000728 | 0.5243678 | 0.9694253 | 0.9006897 | TSPEAR UBE2G2 |

[a] Focal SNP is defined as the center SNP $j$ in window $j^*$.

[b] Proportion of variance explained by sex-specific SNP effects within window $j^*$.

[c] Posterior probability that sex-specific effects are nonzero.

[d] Posterior probability that effects differ between sexes.

[e] Nearest genes identified through Axiom UKB WCSG annotations, release 34.

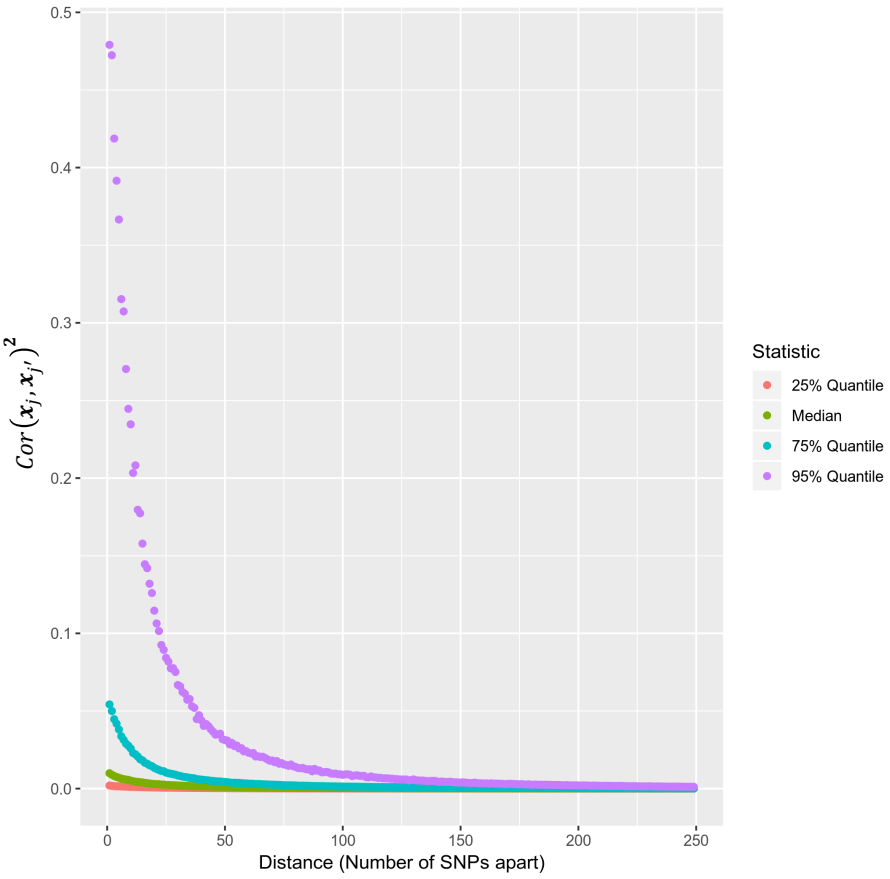


Figure A.1: LD statistics across distances

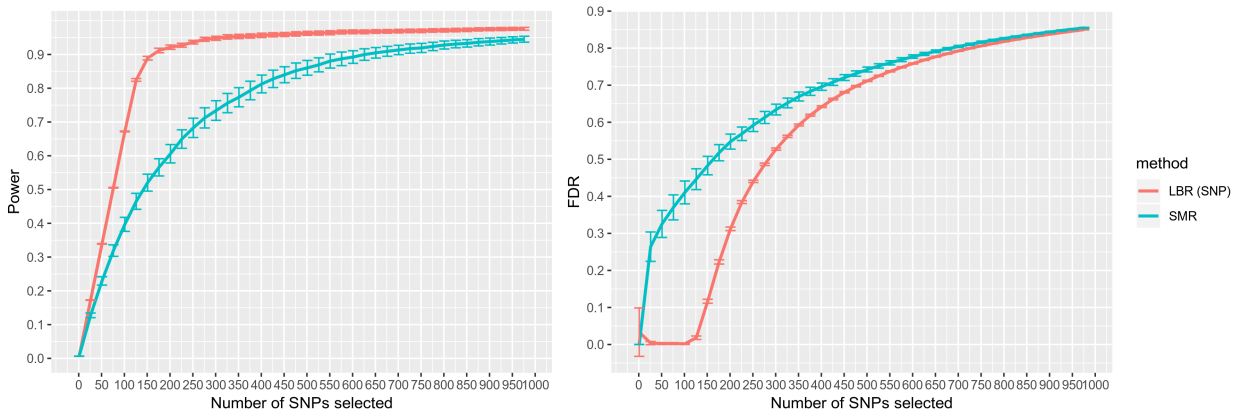


Figure A.2: Estimated power and false-discovery rate for discovering observed SNPs with effects in at least one sex

Estimated power (left) and FDR (right) shown as a function of the number of SNPs selected. Each point represents a sample average and error bars represent 95% confidence intervals, each derived using 30 Monte Carlo replicates. LBR (SNP): local Bayesian regression, utilizing $\text{PP}_{\text{SNP}_j}$. SMR: single-marker regression, utilizing the F-test-based $p$-value.

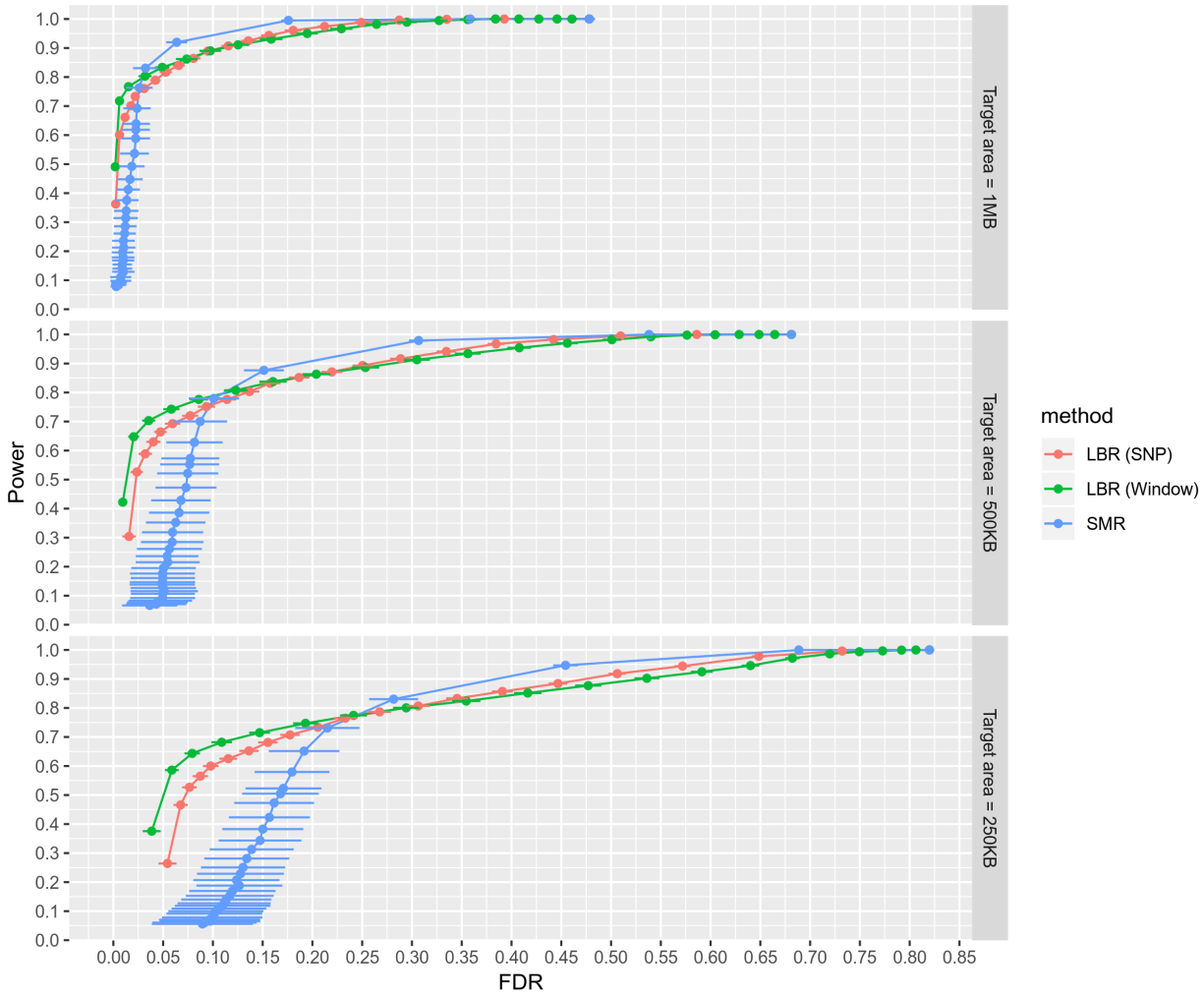


Figure A.3: Power vs false-discovery rate for discovering genomic regions containing masked causal variants

Here power is defined as the expected proportion of causal variants that are being tagged by at least one selected SNP $j$ or window $j^*$. False discovery rate is defined as the proportion of selected SNPs or windows that are not tagging any causal variants. Each point is an estimate and error bars for both axes represent 95% confidence intervals. Point estimates and intervals were derived using 30 Monte Carlo replicates. Each facet corresponds to a different "target area", a fixed width around each causal variant that defines the set of SNPs effectively tagging it. LBR (SNP): uses the $\text{PP}_{\text{SNP}_j}$ metric spanning 1-0. LBR (Window): uses the maximum between $\text{PPM}_{\sigma^2_{g_{j^*}}}$ and $\text{PPF}_{\sigma^2_{g_{j^*}}}$ spanning 1-0. SMR: uses the F-test-based $p$-value spanning (on the -$\log_{10}$ scale) 30-0.

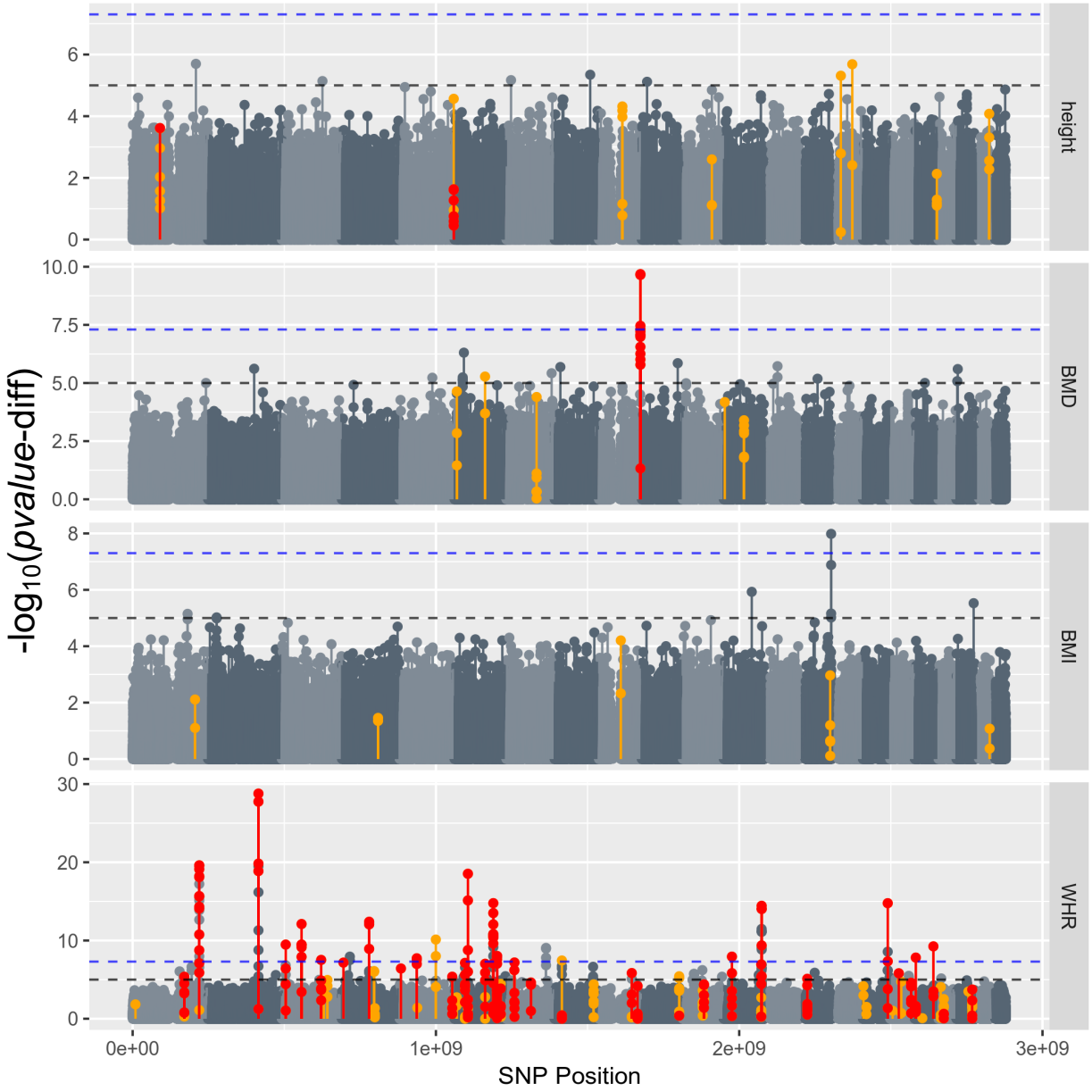


Figure A.4: Comparison between SMR and LBR for discovering G×S interactions
Manhattan plot showing *pvalue*-diff for each analyzed SNP. SNPs are colored yellow if they were focal SNPs with a PPDiff $_{\sigma^2_{g_{j*}}}$ ≥ 0.9 and colored red if they were focal SNPs with a PPDiff $_{\sigma^2_{g_{j*}}}$ ≥ 0.95. The dashed horizontal lines denote p-diff thresholds of $1x10^{-5}$ and $5x10^{-8}$.

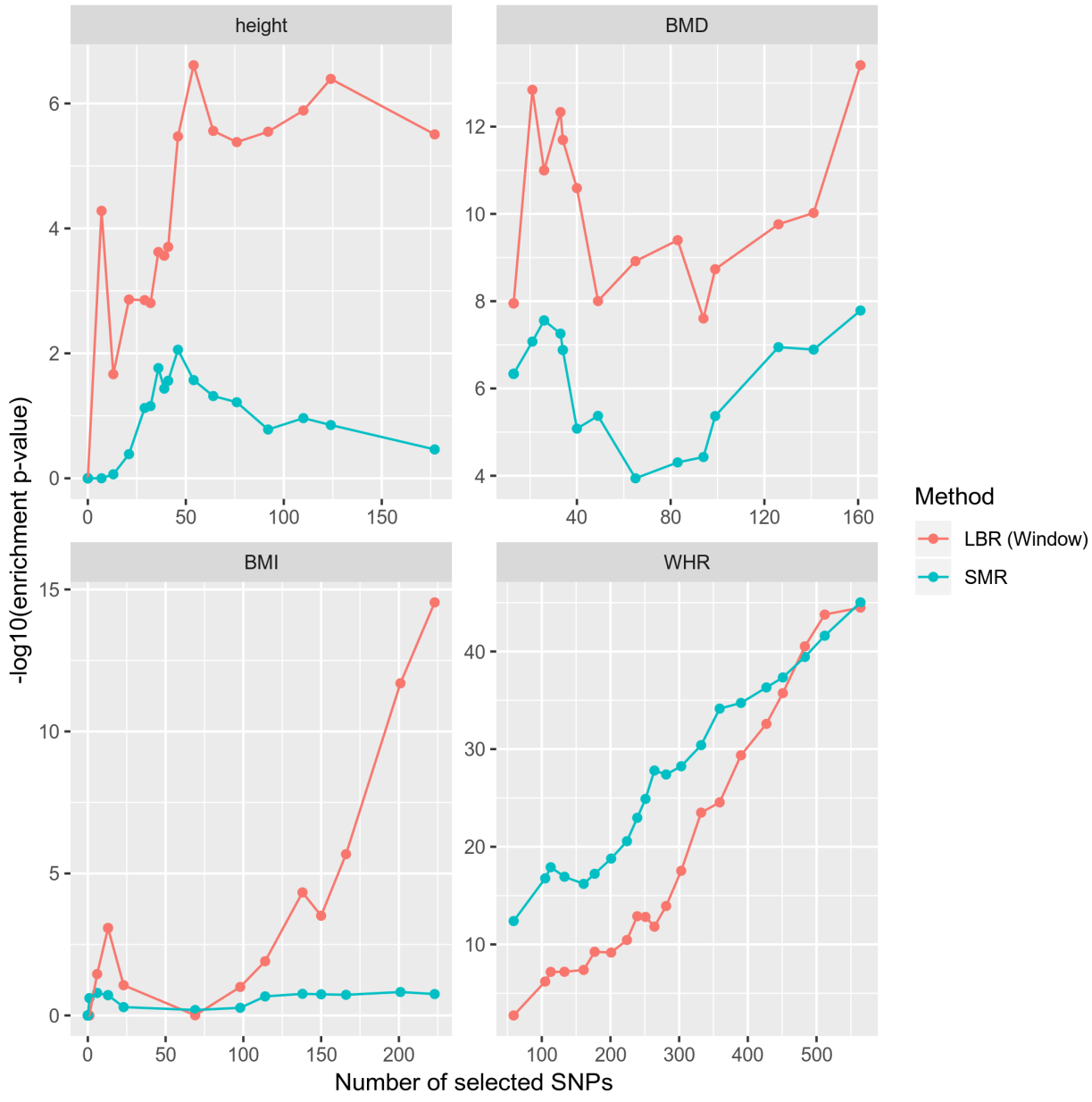


Figure A.5: eQTL enrichment as a function of the number of SNPs selected LBR (Window): uses the PPDiff $_{\sigma^2_{g_{j*}}}$ metric. SMR: uses the *pvalue*-diff metric.

# APPENDIX B

# CHAPTER 4 SUPPLEMENTARY MATERIAL

Table B.1: Heritability estimates for all RNA editing sites with sample size

| Editing site position:strand | N[a] | $\hat{\sigma}_g^{2\,b}$ | $SE\left(\hat{\sigma}_g^2\right)^c$ | $\hat{\sigma}_\varepsilon^{2\,d}$ | $SE\left(\hat{\sigma}_\varepsilon^2\right)^e$ | $\hat{h}^{2f}$ | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| 16:26512555:minus | 139 | 0.576 | 0.221 | 0.454 | 0.149 | 0.56 | 0.0000819 |
| 15:110910484:plus | 166 | 0.386 | 0.171 | 0.609 | 0.134 | 0.388 | 0.0000859 |
| 6:39368241:minus | 166 | 0.411 | 0.176 | 0.599 | 0.135 | 0.407 | 0.000157 |
| 16:26512645:minus | 159 | 0.335 | 0.166 | 0.647 | 0.139 | 0.341 | 0.000234 |
| 1:126167425:minus | 165 | 0.31 | 0.16 | 0.686 | 0.138 | 0.311 | 0.00088 |
| X:60918926:plus | 34 | 0.696 | 0.551 | 0.23 | 0.442 | 0.752 | 0.0302 |
| 6:115902177:plus | 76 | 0.325 | 0.278 | 0.689 | 0.254 | 0.32 | 0.0662 |
| 8:67587714:minus | 89 | 0.333 | 0.241 | 0.689 | 0.218 | 0.326 | 0.0698 |
| 14:91530951:plus | 119 | 0.112 | 0.14 | 0.842 | 0.171 | 0.117 | 0.0752 |
| 2:30295839:minus | 117 | 0.193 | 0.175 | 0.812 | 0.181 | 0.192 | 0.0758 |
| 6:93743264:minus | 141 | 0.189 | 0.152 | 0.816 | 0.159 | 0.188 | 0.0786 |
| 10:56281853:minus | 150 | 0.122 | 0.129 | 0.884 | 0.152 | 0.122 | 0.136 |
| 7:96470770:plus | 34 | 0.512 | 0.524 | 0.378 | 0.44 | 0.575 | 0.137 |
| 6:48557105:minus | 150 | 0.0979 | 0.123 | 0.901 | 0.15 | 0.098 | 0.151 |
| X:60918925:plus | 32 | 0.467 | 0.586 | 0.526 | 0.542 | 0.47 | 0.209 |
| 18:11662642:plus | 54 | 0.196 | 0.324 | 0.816 | 0.338 | 0.194 | 0.263 |
| 1:126837057:minus | 11 | 0.904 | 1.74 | 0.00000325 | 1.64 | 1 | 0.272 |
| 16:26513687:minus | 117 | 0.067 | 0.132 | 0.919 | 0.174 | 0.068 | 0.288 |
| 6:62205750:minus | 86 | 0.115 | 0.53 | 0.93 | 0.214 | 0.11 | 0.291 |
| 5:57668385:plus | 23 | 0.243 | 0.633 | 0.67 | 0.614 | 0.266 | 0.3 |
| 15:77692835:plus | 81 | 0.0832 | 0.186 | 0.899 | 0.228 | 0.0847 | 0.325 |
| 6:75654298:minus | 36 | 0.166 | 0.325 | 0.565 | 0.345 | 0.227 | 0.358 |
| 18:11664433:plus | 36 | 0.106 | 0.416 | 0.909 | 0.468 | 0.105 | 0.362 |
| 13:140729295:minus | 68 | 0.069 | 0.219 | 0.933 | 0.265 | 0.0688 | 0.365 |
| 4:91383948:minus | 10 | 0.876 | 2.06 | 0.126 | 1.49 | 0.874 | 0.366 |
| 14:87895203:minus | 54 | 0.057 | 0.259 | 0.9 | 0.301 | 0.0595 | 0.379 |
| 6:159410531:minus | 29 | 0.0676 | 0.609 | 0.95 | 0.599 | 0.0664 | 0.438 |
| 14:39851072:minus | 33 | 0.0375 | 0.429 | 0.994 | 0.53 | 0.0363 | 0.474 |
| 18:17492309:plus | 150 | 0.00431 | 0.089 | 0.998 | 0.146 | 0.0043 | 0.477 |
| 1:128090247:minus | 74 | 0.0083 | 0.163 | 0.974 | 0.241 | 0.00845 | 0.48 |
| 10:29217567:plus | 34 | 0.00000698 | 0.399 | 1.02 | 0.473 | 0.00000685 | 0.5 |
| 13:111022773:minus | 60 | 0.00000731 | 0.221 | 0.999 | 0.288 | 0.00000732 | 0.5 |
| 13:111022791:minus | 79 | 1.97E-09 | 0.172 | 1.01 | 0.237 | 1.94E-09 | 0.5 |
| 13:111023101:minus | 80 | 1.86E-09 | 0.154 | 0.927 | 0.214 | 2.01E-09 | 0.5 |
| 13:146223066:plus | 18 | 0.000000764 | 0.613 | 0.819 | 0.688 | 0.000000933 | 0.5 |
| 13:39922407:plus | 16 | 0.000229 | 0.831 | 1.03 | 0.919 | 0.000222 | 0.5 |
| 13:86543564:minus | 68 | 1.82E-09 | 0.182 | 0.943 | 0.248 | 1.92E-09 | 0.5 |
| 14:111666396:plus | 42 | 7.92E-08 | 0.319 | 1.02 | 0.392 | 7.76E-08 | 0.5 |
| 15:110909979:plus | 117 | 1.93E-09 | 0.111 | 1 | 0.173 | 1.92E-09 | 0.5 |
| 15:59690693:plus | 88 | 0.000000047 | 0.135 | 0.904 | 0.194 | 0.000000052 | 0.5 |
| 16:23629281:minus | 138 | 1.83E-09 | 0.094 | 0.997 | 0.154 | 1.84E-09 | 0.5 |
| 6:168677788:plus | 15 | 0.382 | 0.971 | 0.472 | 1.05 | 0.448 | 0.5 |
| 6:62207065:minus | 43 | 1.51E-09 | 0.322 | 0.978 | 0.384 | 1.55E-09 | 0.5 |
| 6:93743195:minus | 124 | 2.01E-09 | 0.099 | 0.965 | 0.159 | 2.09E-09 | 0.5 |
| 7:3053788:plus | 40 | 1.49E-09 | 0.365 | 1.03 | 0.436 | 1.45E-09 | 0.5 |
| 9:115531653:minus | 31 | 4.66E-08 | 0.411 | 1.02 | 0.499 | 4.59E-08 | 0.5 |
| X:60916145:plus | 24 | 1.48E-09 | 0.578 | 1 | 0.665 | 1.47E-09 | 0.5 |

[a] Sample size (the number of animals with a detectable editing level)
[b] REML estimated genomic variance component
[c] Standard error of genomic variance estimate
[d] REML estimated residual variance component
[e] Standard error of residual variance estimate
[f] Genomic heritability estimate
[g] $p$-value from a likelihood ratio test, testing $H_0 : \sigma_g^2 = 0$.

Table B.2: *ADAR*-localized genomic variance estimates for 67 carcass composition, meat quality, and growth traits

| trait[a] | N[b] | $\hat{\sigma}^2_{g_{local}}$ [c] | $SE\left(\hat{\sigma}^2_{g_{local}}\right)$ [d] | $\hat{\sigma}^2_{g_{BG}}$ [e] | $SE\left(\hat{\sigma}^2_{g_{BG}}\right)$ [f] | $p$-value[g] |
|---|---|---|---|---|---|---|
| ADG | 936 | 0.144 | 0.0849 | 0.266 | 0.0545 | 0.000293 |
| last_lum | 932 | 0.0265 | 0.0237 | 0.358 | 0.0615 | 0.000459 |
| tofat | 936 | 0.0773 | 0.051 | 0.247 | 0.0513 | 0.000774 |
| bf10_22wk | 940 | 0.0325 | 0.0259 | 0.315 | 0.0528 | 0.00118 |
| car_bf10 | 927 | 0.0275 | 0.0227 | 0.287 | 0.0495 | 0.00264 |
| mtpro | 936 | 0.0355 | 0.0281 | 0.3 | 0.0539 | 0.0032 |
| belly | 933 | 0.0343 | 0.0284 | 0.205 | 0.0515 | 0.00912 |
| wt_22wk | 940 | 0.0541 | 0.0395 | 0.252 | 0.0548 | 0.00967 |
| Days | 940 | 0.0677 | 0.0467 | 0.231 | 0.0528 | 0.0109 |
| WBS | 923 | 0.0145 | 0.0166 | 0.259 | 0.0582 | 0.0154 |
| lrf_22wk | 940 | 0.0166 | 0.0168 | 0.368 | 0.0576 | 0.0219 |
| fftoln | 936 | 0.0328 | 0.0278 | 0.212 | 0.0532 | 0.0268 |
| mtfat | 936 | 0.0383 | 0.0309 | 0.221 | 0.0538 | 0.0276 |
| car_wt | 934 | 0.0254 | 0.0226 | 0.142 | 0.0442 | 0.0424 |
| bf10_16wk | 940 | 0.0163 | 0.0171 | 0.402 | 0.0621 | 0.0483 |
| lrf_16wk | 940 | 0.00887 | 0.0126 | 0.411 | 0.0659 | 0.0651 |
| farm_wt | 934 | 0.0211 | 0.0202 | 0.189 | 0.0498 | 0.0685 |
| lma_22wk | 940 | 0.0124 | 0.0153 | 0.339 | 0.064 | 0.0967 |
| bf10_19wk | 940 | 0.00681 | 0.0105 | 0.393 | 0.0602 | 0.0972 |
| lrf_13wk | 940 | 0.00817 | 0.0122 | 0.382 | 0.0648 | 0.115 |
| wt_19wk | 940 | 0.0169 | 0.0182 | 0.325 | 0.0626 | 0.123 |
| bf10_10wk | 940 | 0.00545 | 0.00957 | 0.362 | 0.0632 | 0.126 |
| first_rib | 845 | 0.00615 | 0.0104 | 0.203 | 0.0529 | 0.133 |
| lrf_10wk | 940 | 0.00541 | 0.0101 | 0.321 | 0.0622 | 0.164 |

Table B.2 (cont'd)

| trait[a] | N[b] | $\hat{\sigma}^2_{g_{local}}$ [c] | $SE\left(\hat{\sigma}^2_{g_{local}}\right)$ [d] | $\hat{\sigma}^2_{g_{BG}}$ [e] | $SE\left(\hat{\sigma}^2_{g_{BG}}\right)$ [f] | $p$-value[g] |
|---|---|---|---|---|---|---|
| lrf_19wk | 940 | 0.0041 | 0.00833 | 0.49 | 0.0679 | 0.181 |
| lma_19wk | 940 | 0.0107 | 0.014 | 0.315 | 0.0618 | 0.185 |
| wt_16wk | 940 | 0.0095 | 0.0129 | 0.255 | 0.0564 | 0.204 |
| temp_24h | 931 | 0.00549 | 0.00953 | 0.179 | 0.0488 | 0.205 |
| lma_16wk | 940 | 0.00488 | 0.0096 | 0.274 | 0.0583 | 0.207 |
| off_flavor | 928 | 0.00274 | 0.00605 | 0.0312 | 0.0254 | 0.211 |
| ph_24h | 913 | 0.00315 | 0.00748 | 0.192 | 0.0512 | 0.219 |
| conn_tiss | 928 | 0.00364 | 0.00779 | 0.135 | 0.0435 | 0.248 |
| juiciness | 928 | 0.00428 | 0.00767 | 0.0695 | 0.0337 | 0.252 |
| tenderness | 928 | 0.00439 | 0.0092 | 0.265 | 0.0582 | 0.262 |
| last_rib | 933 | 0.00259 | 0.00753 | 0.251 | 0.0538 | 0.284 |
| bf10_13wk | 940 | 0.00225 | 0.00717 | 0.363 | 0.0608 | 0.325 |
| overtend | 928 | 0.00246 | 0.00764 | 0.273 | 0.0591 | 0.354 |
| firm | 918 | 0.00147 | 0.00604 | 0.146 | 0.0453 | 0.397 |
| lma_13wk | 940 | 0.00147 | 0.00693 | 0.332 | 0.0637 | 0.409 |
| wt_birth | 940 | 0.000834 | 0.00557 | 0.213 | 0.0534 | 0.464 |
| b | 887 | 0.000294 | 0.00596 | 0.356 | 0.0645 | 0.476 |
| temp_45m | 933 | 0.00027 | 0.00534 | 0.2 | 0.0521 | 0.476 |
| wt_3wk | 940 | 2.06E-09 | 0.00388 | 0.0603 | 0.0313 | 0.5 |
| wt_6wk | 939 | 0.00000685 | 0.00484 | 0.166 | 0.0473 | 0.5 |
| wt_10wk | 940 | 2.06E-09 | 0.00515 | 0.225 | 0.0535 | 0.5 |
| lma_10wk | 940 | 2.06E-09 | 0.00545 | 0.3 | 0.0598 | 0.5 |
| wt_13wk | 940 | 0.0000122 | 0.00553 | 0.283 | 0.0596 | 0.5 |
| dress_ptg | 934 | 0.00000018 | 0.00542 | 0.25 | 0.0572 | 0.5 |
| color | 931 | 0.00000206 | 0.00531 | 0.227 | 0.0549 | 0.5 |

Table B.2 (cont'd)

| trait[a] | N[b] | $\hat{\sigma}^2_{g_{local}}$ [c] | $SE\left(\hat{\sigma}^2_{g_{local}}\right)$ [d] | $\hat{\sigma}^2_{g_{BG}}$ [e] | $SE\left(\hat{\sigma}^2_{g_{BG}}\right)$ [f] | $p$-value[g] |
|---|---|---|---|---|---|---|
| L | 887 | 2.06E-09 | 0.00586 | 0.342 | 0.0658 | 0.5 |
| a | 887 | 0.00000111 | 0.00572 | 0.408 | 0.0673 | 0.5 |
| cook_yield | 924 | 4.91E-08 | 0.00552 | 0.286 | 0.0598 | 0.5 |
| marb | 932 | 2.06E-09 | 0.0059 | 0.372 | 0.0671 | 0.5 |
| driploss | 932 | 0.00000275 | 0.00524 | 0.239 | 0.0551 | 0.5 |
| ph_45m | 920 | 2.06E-09 | 0.00442 | 0.107 | 0.0399 | 0.5 |
| pH_dec | 900 | 0.000000122 | 0.00418 | 0.0735 | 0.0347 | 0.5 |
| car_length | 933 | 0.000000163 | 0.00595 | 0.393 | 0.0684 | 0.5 |
| num_ribs | 655 | 0.00000781 | 0.00743 | 0.361 | 0.0811 | 0.5 |
| car_lma | 928 | 2.06E-09 | 0.00565 | 0.469 | 0.0687 | 0.5 |
| ham | 933 | 2.06E-09 | 0.00545 | 0.265 | 0.0582 | 0.5 |
| loin | 933 | 2.06E-09 | 0.0054 | 0.254 | 0.0571 | 0.5 |
| boston | 933 | 2.06E-09 | 0.0064 | 0.403 | 0.0726 | 0.5 |
| picnic | 933 | 2.06E-09 | 0.00697 | 0.519 | 0.0825 | 0.5 |
| spareribs | 930 | 2.06E-09 | 0.00601 | 0.389 | 0.0688 | 0.5 |
| moisture | 922 | 6.58E-08 | 0.00558 | 0.312 | 0.0616 | 0.5 |
| fat | 922 | 2.06E-09 | 0.00592 | 0.483 | 0.0717 | 0.5 |
| protein | 921 | 2.06E-09 | 0.00575 | 0.354 | 0.065 | 0.5 |

[a] More information about each trait can be found in Velez-Irizarry et al. [106]

[b] Sample size

[c] REML estimated ADAR-localized genomic variance component

[d] Standard error of ADAR-localized genomic variance estimate

[e] REML estimated background genomic variance component

[f] Standard error of background genomic variance estimate

[g] $p$-value from a likelihood ratio test, testing $H_0 : \sigma^2_{g_{local}} = 0$.

Table B.3: Genomic covariance estimates between site-specific RNA editing levels and 67 carcass composition, meat quality, and growth traits

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\left(\hat{\rho}_g\right)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| temp_45m | 15:110910484:plus | 0.6978532 | 0.2006619 | 0.1947621 | -0.2112079 | -0.0164458 | 0.0013573 |
| moisture | 16:26512645:minus | -0.7000733 | 0.2000639 | -0.2168411 | 0.0504626 | -0.1663786 | 0.0041762 |
| b | 1:126167425:minus | 0.6289766 | 0.2138839 | 0.194918 | -0.036493 | 0.158425 | 0.0064947 |
| moisture | 16:26512555:minus | -0.5179504 | 0.177277 | -0.2196051 | 0.02667 | -0.1929351 | 0.0113916 |
| lrf_22wk | 16:26512645:minus | 0.5314482 | 0.2058477 | 0.1795119 | -0.0102409 | 0.1692709 | 0.0121452 |
| boston | 6:39368241:minus | -0.5933886 | 0.1416324 | -0.3154157 | -0.0886732 | -0.4040889 | 0.0133171 |
| temp_45m | 1:126167425:minus | 0.667827 | 0.2164199 | 0.1734387 | -0.0248664 | 0.1485723 | 0.0148532 |
| fat | 16:26512645:minus | 0.5161289 | 0.1897534 | 0.2086495 | 0.0494572 | 0.2581067 | 0.0179607 |
| last_lum | 16:26512645:minus | 0.5145507 | 0.1952752 | 0.1913932 | -0.0039213 | 0.1874719 | 0.0194878 |
| conn_tiss | 15:110910484:plus | -0.5527452 | 0.2268368 | -0.1374818 | 0.0535365 | -0.0839453 | 0.0197101 |
| fftoln | 15:110910484:plus | -0.5108137 | 0.1923533 | -0.1818749 | 0.0911418 | -0.0907331 | 0.0228199 |
| picnic | 16:26512555:minus | 0.6174721 | 0.162787 | 0.326789 | -0.2618334 | 0.0649556 | 0.0236605 |
| color | 16:26512555:minus | -0.4711909 | 0.1923727 | -0.1755983 | -0.0175018 | -0.1931001 | 0.0256422 |
| L | 16:26512555:minus | 0.4781228 | 0.1821823 | 0.2071758 | -0.0436777 | 0.1634981 | 0.0263383 |
| last_rib | 16:26512645:minus | 0.4859347 | 0.2266758 | 0.1353587 | 0.0587771 | 0.1941358 | 0.0391113 |
| wt_3wk | 16:26512555:minus | -0.6893225 | 0.283316 | -0.1189598 | 0.0715114 | -0.0474485 | 0.0476917 |
| mtfat | 15:110910484:plus | -0.4472025 | 0.2039263 | -0.1522525 | 0.0757822 | -0.0764703 | 0.0477578 |
| car_bf10 | 16:26512645:minus | 0.3805601 | 0.2066879 | 0.1216296 | 0.0658033 | 0.1874328 | 0.0481823 |
| moisture | 1:126167425:minus | -0.4750748 | 0.2306446 | -0.1433086 | 0.1190167 | -0.0242919 | 0.0498548 |
| pH_dec | 16:26512645:minus | 0.5150611 | 0.3327587 | 0.0772333 | -0.1163237 | -0.0390904 | 0.0498863 |
| ph_45m | 1:126167425:minus | 0.5916669 | 0.2981662 | 0.1010036 | -0.1692225 | -0.0682189 | 0.0500024 |
| num_ribs | 16:26512555:minus | -0.4653506 | 0.200344 | -0.2025083 | 0.1169612 | -0.0855471 | 0.0500832 |
| b | 15:110910484:plus | 0.4728684 | 0.2177794 | 0.1524001 | -0.0375163 | 0.1148838 | 0.0508777 |
| pH_dec | 1:126167425:minus | 0.545542 | 0.3473278 | 0.0745804 | -0.0602521 | 0.0143283 | 0.0517325 |
| car_wt | 16:26512645:minus | 0.5843297 | 0.2535415 | 0.1309091 | -0.1174667 | 0.0134424 | 0.0572019 |
| dress_ptg | 15:110910484:plus | 0.3978575 | 0.2058523 | 0.136633 | -0.0464345 | 0.0901985 | 0.0594525 |
| temp_24h | 6:39368241:minus | -0.4792902 | 0.2032275 | -0.1400908 | -0.3415478 | -0.4816387 | 0.0599582 |
| wt_3wk | 6:39368241:minus | -0.6890714 | 0.3143074 | -0.094669 | 0.0404951 | -0.0541739 | 0.0599692 |
| driploss | 15:110910484:plus | -0.3867181 | 0.2107094 | -0.1298377 | 0.1519669 | 0.0221292 | 0.0630321 |
| wt_10wk | 15:110910484:plus | -0.4181734 | 0.2175119 | -0.1252656 | -0.0017598 | -0.1270255 | 0.0632891 |
| belly | 16:26512645:minus | 0.5505696 | 0.2374497 | 0.1403632 | 0.0305953 | 0.1709586 | 0.0667876 |
| marb | 16:26512645:minus | 0.4073535 | 0.2089169 | 0.1462183 | 0.0804837 | 0.226702 | 0.0783371 |
| temp_24h | 16:26512645:minus | 0.5809202 | 0.2723509 | 0.1205994 | -0.1429541 | -0.0223547 | 0.0813252 |
| b | 16:26512555:minus | 0.3182276 | 0.1857647 | 0.1465983 | -0.0996418 | 0.0469565 | 0.0820527 |
| lma_19wk | 15:110910484:plus | -0.3544394 | 0.1928252 | -0.1416331 | 0.0641738 | -0.0774593 | 0.0831723 |
| lma_22wk | 15:110910484:plus | -0.3548259 | 0.185235 | -0.1522617 | 0.0392898 | -0.1129718 | 0.0841729 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\,(\hat{\rho}_g)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| wt_3wk | 15:110910484:plus | -0.5972737 | 0.3112136 | -0.0873063 | -0.0039575 | -0.0912638 | 0.0886598 |
| ADG | 15:110910484:plus | -0.3313927 | 0.1919747 | -0.134748 | 0.1008835 | -0.0338645 | 0.1009422 |
| bf10_19wk | 6:39368241:minus | 0.2911484 | 0.181434 | 0.1274551 | -0.1708141 | -0.043359 | 0.1041676 |
| bf10_22wk | 16:26512645:minus | 0.3444655 | 0.2183995 | 0.1121352 | 0.0114755 | 0.1236107 | 0.104852 |
| a | 1:126167425:minus | -0.4000349 | 0.2477763 | -0.1218053 | 0.1353083 | 0.013503 | 0.1066612 |
| ph_45m | 15:110910484:plus | 0.5515001 | 0.1747256 | 0.0979464 | -0.1123954 | -0.014449 | 0.1081383 |
| L | 15:110910484:plus | -0.3501275 | 0.1985502 | -0.1386411 | 0.1738201 | 0.035179 | 0.1085107 |
| a | 16:26512555:minus | -0.2911679 | 0.1825653 | -0.1408192 | 0.147053 | 0.0062338 | 0.109653 |
| ham | 15:110910484:plus | 0.3728721 | 0.2201093 | 0.1199787 | -0.1438646 | -0.0238859 | 0.1128409 |
| wt_6wk | 6:39368241:minus | -0.4467673 | 0.2535047 | -0.1049723 | 0.0962471 | -0.0087252 | 0.1140483 |
| wt_6wk | 15:110910484:plus | -0.3815983 | 0.2228853 | -0.1076214 | -0.0741062 | -0.1817275 | 0.1148975 |
| first_rib | 16:26512645:minus | 0.399654 | 0.2403603 | 0.111272 | 0.0607663 | 0.1720384 | 0.1233643 |
| WBS | 16:26512645:minus | -0.3636745 | 0.2209916 | -0.1186282 | -0.0154713 | -0.1340995 | 0.1266919 |
| farm_wt | 16:26512645:minus | 0.4520142 | 0.2501198 | 0.1157098 | -0.0985005 | 0.0172093 | 0.129934 |
| wt_22wk | 15:110910484:plus | -0.3195801 | 0.2055025 | -0.1153862 | 0.0797386 | -0.0356475 | 0.1379282 |
| mtpro | 16:26512645:minus | 0.3398009 | 0.2242867 | 0.1089872 | -0.0133729 | 0.0956143 | 0.1434817 |
| num_ribs | 1:126167425:minus | -0.3870412 | 0.2455344 | -0.1281953 | 0.1351493 | 0.0069539 | 0.1510288 |
| wt_6wk | 16:26512645:minus | -0.3778147 | 0.2514977 | -0.0942609 | -0.0085057 | -0.1027666 | 0.1524548 |
| a | 6:39368241:minus | -0.2674827 | 0.1846067 | -0.1169186 | -0.0956302 | -0.2125488 | 0.1528527 |
| lrf_19wk | 16:26512645:minus | 0.3072594 | 0.2080163 | 0.1187616 | 0.0643692 | 0.1831308 | 0.1605653 |
| wt_10wk | 16:26512645:minus | -0.3241411 | 0.2409166 | -0.0928159 | 0.1103734 | 0.0175575 | 0.1652651 |
| Days | 15:110910484:plus | 0.3004014 | 0.2046399 | 0.1084417 | -0.0099376 | 0.0985042 | 0.1662346 |
| protein | 1:126167425:minus | 0.2926056 | 0.2187942 | 0.10401 | -0.0876923 | 0.0163177 | 0.1663039 |
| firm | 16:26512645:minus | 0.3378975 | 0.2632971 | 0.080186 | -0.0424556 | 0.0377304 | 0.1711678 |
| ph_45m | 6:39368241:minus | 0.3912599 | 0.2785512 | 0.079414 | -0.0531326 | 0.0262814 | 0.1741369 |
| bf10_13wk | 15:110910484:plus | 0.28291 | 0.2116243 | 0.100724 | -0.1323884 | -0.0316644 | 0.1787045 |
| b | 16:26512645:minus | 0.28794 | 0.2252757 | 0.0978759 | -0.0264436 | 0.0714323 | 0.1879715 |
| a | 15:110910484:plus | -0.2780321 | 0.206455 | -0.105727 | -0.0030079 | -0.1087349 | 0.1933367 |
| conn_tiss | 16:26512555:minus | -0.3209072 | 0.2378448 | -0.0917414 | 0.0243247 | -0.0674166 | 0.1940133 |
| car_lma | 16:26512555:minus | 0.2720148 | 0.1734739 | 0.1442266 | -0.1099755 | 0.0342512 | 0.2045772 |
| bf10_19wk | 16:26512645:minus | 0.2837844 | 0.2172744 | 0.0988265 | 0.0141858 | 0.1130123 | 0.2048029 |
| off_flavor | 1:126167425:minus | -0.4994814 | 0.4187853 | -0.0537841 | 0.0470309 | -0.0067532 | 0.2088798 |
| juiciness | 1:126167425:minus | -0.3695545 | 0.3075291 | -0.0655989 | 0.0782349 | 0.012636 | 0.2111407 |
| ph_45m | 16:26512645:minus | 0.4301392 | 0.3073378 | 0.0739628 | -0.0103846 | 0.0635782 | 0.2132017 |
| ph_24h | 15:110910484:plus | 0.2956059 | 0.2298783 | 0.0887444 | -0.1305047 | -0.0417603 | 0.2174268 |
| last_lum | 1:126167425:minus | 0.2688917 | 0.2150534 | 0.0982519 | -0.0501248 | 0.0481271 | 0.2227989 |
| pH_dec | 6:39368241:minus | 0.3892706 | 0.3144992 | 0.0639188 | -0.0038634 | 0.0600554 | 0.2259253 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\,(\hat{\rho}_g)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| bf10_16wk | 16:26512645:minus | 0.2581885 | 0.2112177 | 0.0957501 | -0.0083846 | 0.0873655 | 0.2270376 |
| lrf_16wk | 6:39368241:minus | 0.2205575 | 0.1900252 | 0.0964238 | -0.0872119 | 0.0092119 | 0.2331844 |
| car_bf10 | 6:39368241:minus | 0.2082403 | 0.1949637 | 0.0770797 | -0.0462581 | 0.0308216 | 0.2338157 |
| tenderness | 15:110910484:plus | -0.2758242 | 0.2151229 | -0.0944788 | 0.0445027 | -0.049976 | 0.2353996 |
| last_lum | 15:110910484:plus | 0.2533347 | 0.2031338 | 0.0979362 | -0.0214882 | 0.076448 | 0.237471 |
| wt_16wk | 15:110910484:plus | -0.2590085 | 0.2138979 | -0.0880929 | -0.0200387 | -0.1081316 | 0.2385872 |
| protein | 16:26512555:minus | 0.2240531 | 0.1945021 | 0.1001846 | -0.1833907 | -0.083206 | 0.2408811 |
| overtend | 16:26512645:minus | 0.288481 | 0.2282291 | 0.0928696 | -0.0039821 | 0.0888875 | 0.2416114 |
| driploss | 6:39368241:minus | -0.2436748 | 0.2204878 | -0.0803892 | 0.1009261 | 0.0205368 | 0.2466252 |
| overtend | 15:110910484:plus | -0.2732181 | 0.2152852 | -0.0938704 | 0.0415702 | -0.0523002 | 0.2468283 |
| belly | 15:110910484:plus | 0.28095 | 0.2248886 | 0.0866608 | 0.0313898 | 0.1180506 | 0.24926 |
| fftoln | 6:39368241:minus | -0.272801 | 0.2156793 | -0.0927051 | 0.0060356 | -0.0866695 | 0.251919 |
| belly | 1:126167425:minus | 0.2893782 | 0.2463631 | 0.0800935 | 0.0049732 | 0.0850667 | 0.2525818 |
| car_wt | 16:26512555:minus | 0.2947202 | 0.2305088 | 0.0920975 | -0.0953904 | -0.0032929 | 0.2643088 |
| car_length | 16:26512555:minus | 0.2231852 | 0.1880947 | 0.1072539 | -0.0550028 | 0.0522511 | 0.2646737 |
| tofat | 15:110910484:plus | -0.217367 | 0.2033011 | -0.079918 | 0.0369009 | -0.0430171 | 0.2704865 |
| temp_24h | 1:126167425:minus | -0.3140209 | 0.2579774 | -0.0743201 | -0.3131702 | -0.3874903 | 0.2727324 |
| pH_dec | 15:110910484:plus | 0.4554498 | 0.347066 | 0.0635814 | -0.0123335 | 0.0512479 | 0.2730641 |
| loin | 16:26512555:minus | 0.2595433 | 0.2109098 | 0.0980741 | -0.0656056 | 0.0324685 | 0.2744468 |
| ph_24h | 16:26512645:minus | -0.2828384 | 0.2572533 | -0.0736194 | 0.110557 | 0.0369376 | 0.2748747 |
| cook_yield | 15:110910484:plus | 0.2489873 | 0.2145452 | 0.0866483 | -0.0049773 | 0.0816709 | 0.2786352 |
| wt_19wk | 16:26512555:minus | -0.2211012 | 0.190224 | -0.1007887 | -0.0151277 | -0.1159164 | 0.2793971 |
| fat | 16:26512555:minus | 0.2073381 | 0.179037 | 0.1033649 | 0.1796151 | 0.28298 | 0.2803456 |
| car_wt | 15:110910484:plus | 0.2998882 | 0.2578546 | 0.0736429 | -0.1189525 | -0.0453096 | 0.2828107 |
| wt_19wk | 15:110910484:plus | -0.2276545 | 0.2056989 | -0.0870598 | 0.0222916 | -0.0647682 | 0.2887758 |
| tenderness | 16:26512645:minus | 0.25802 | 0.2312266 | 0.0818423 | 0.0116808 | 0.093523 | 0.293002 |
| lma_13wk | 6:39368241:minus | 0.2182071 | 0.2074296 | 0.0825468 | -0.0036772 | 0.0788697 | 0.293991 |
| marb | 15:110910484:plus | 0.232211 | 0.2061084 | 0.0898224 | 0.0101539 | 0.0999763 | 0.2948951 |
| spareribs | 6:39368241:minus | -0.238048 | 0.1937929 | -0.1013694 | -0.0328957 | -0.1342652 | 0.2970763 |
| dress_ptg | 16:26512645:minus | 0.2471977 | 0.2438372 | 0.0725209 | -0.0376742 | 0.0348467 | 0.2976703 |
| lrf_16wk | 16:26512645:minus | 0.2246823 | 0.2145423 | 0.083247 | 0.0754287 | 0.1586757 | 0.2983286 |
| spareribs | 15:110910484:plus | -0.2544918 | 0.2070737 | -0.0984231 | -0.0042541 | -0.1026772 | 0.3023104 |
| pH_dec | 16:26512555:minus | 0.3315993 | 0.308353 | 0.0647022 | -0.0946616 | -0.0299593 | 0.3060452 |
| wt_6wk | 1:126167425:minus | -0.2834482 | 0.2751467 | -0.0637884 | 0.0046975 | -0.0590909 | 0.3072649 |
| num_ribs | 16:26512645:minus | -0.272252 | 0.2404619 | -0.0944863 | -0.0739905 | -0.1684768 | 0.308162 |
| farm_wt | 16:26512555:minus | 0.2696127 | 0.2195012 | 0.0935755 | -0.0893406 | 0.0042349 | 0.3085265 |
| bf10_13wk | 1:126167425:minus | 0.2158451 | 0.2287056 | 0.0713476 | -0.1296212 | -0.0582736 | 0.3131989 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\,(\hat{\rho}_g)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| car_wt | 1:126167425:minus | 0.2779068 | 0.2741068 | 0.064099 | -0.0963426 | -0.0322436 | 0.3171421 |
| first_rib | 6:39368241:minus | 0.2480959 | 0.2370649 | 0.0738355 | -0.0609145 | 0.012921 | 0.3209069 |
| num_ribs | 6:39368241:minus | -0.2255063 | 0.2144594 | -0.0900992 | -0.175821 | -0.2659202 | 0.3276659 |
| last_rib | 1:126167425:minus | -0.2083497 | 0.2375975 | -0.0618438 | 0.0742384 | 0.0123946 | 0.3291292 |
| mtfat | 6:39368241:minus | -0.2193661 | 0.2128225 | -0.0763235 | -0.0632297 | -0.1395532 | 0.3333426 |
| L | 1:126167425:minus | 0.2341815 | 0.2315646 | 0.07841 | -0.0489712 | 0.0294389 | 0.3353028 |
| belly | 16:26512555:minus | 0.2350459 | 0.2183797 | 0.0834237 | 0.008415 | 0.0918387 | 0.3360095 |
| ADG | 6:39368241:minus | -0.1929044 | 0.1931165 | -0.0792743 | -0.0961642 | -0.1754385 | 0.3402255 |
| lrf_19wk | 6:39368241:minus | 0.1750989 | 0.1847332 | 0.0811234 | -0.0362444 | 0.044879 | 0.3415369 |
| color | 15:110910484:plus | 0.2234357 | 0.2261328 | 0.0715226 | -0.1235983 | -0.0520756 | 0.3437511 |
| L | 16:26512645:minus | 0.2289738 | 0.2281236 | 0.0786197 | -0.031651 | 0.0469687 | 0.3448456 |
| lrf_10wk | 6:39368241:minus | 0.1852931 | 0.2079282 | 0.0697556 | 0.0491578 | 0.1189133 | 0.3502217 |
| conn_tiss | 6:39368241:minus | -0.2302646 | 0.2528379 | -0.0566242 | -0.0961793 | -0.1528035 | 0.3524339 |
| lma_22wk | 16:26512645:minus | 0.2412204 | 0.2342659 | 0.0794976 | 0.002409 | 0.0819066 | 0.3548819 |
| tofat | 16:26512645:minus | 0.2179066 | 0.2352458 | 0.06675 | -0.0054048 | 0.0613452 | 0.3563514 |
| bf10_22wk | 15:110910484:plus | 0.1791731 | 0.2057897 | 0.0644749 | -0.0200097 | 0.0444652 | 0.36103 |
| mtfat | 16:26512555:minus | -0.208964 | 0.2079981 | -0.0811153 | -0.0141015 | -0.0952168 | 0.3614119 |
| lma_10wk | 6:39368241:minus | 0.1840969 | 0.2106209 | 0.0668068 | -0.0048988 | 0.061908 | 0.3640684 |
| WBS | 16:26512555:minus | 0.1969153 | 0.2084988 | 0.0783108 | -0.047497 | 0.0308138 | 0.3677198 |
| off_flavor | 15:110910484:plus | -0.3819488 | 0.4347374 | -0.040543 | 0.0357121 | -0.0048309 | 0.3743531 |
| farm_wt | 1:126167425:minus | 0.2414641 | 0.2603512 | 0.0622988 | -0.0848597 | -0.0225609 | 0.3754812 |
| last_lum | 6:39368241:minus | 0.183068 | 0.2002154 | 0.0732312 | 0.0313499 | 0.1045812 | 0.3781264 |
| loin | 15:110910484:plus | -0.217189 | 0.2285626 | -0.0687994 | 0.0482158 | -0.0205836 | 0.3820069 |
| wt_birth | 16:26512555:minus | 0.2111275 | 0.2151731 | 0.0768658 | -0.0217238 | 0.055142 | 0.3831235 |
| firm | 6:39368241:minus | 0.2050648 | 0.2530462 | 0.0529088 | -0.0544969 | -0.0015881 | 0.3873636 |
| wt_10wk | 1:126167425:minus | -0.2133004 | 0.2557446 | -0.0561465 | -0.0435573 | -0.0997038 | 0.3892369 |
| protein | 15:110910484:plus | 0.1782856 | 0.2122012 | 0.0661004 | -0.0170241 | 0.0490763 | 0.3906322 |
| bf10_10wk | 1:126167425:minus | 0.1713848 | 0.2254814 | 0.0593093 | -0.0821068 | -0.0227975 | 0.3966245 |
| Days | 16:26512555:minus | 0.1830439 | 0.1978819 | 0.0770067 | 0.0044836 | 0.0814903 | 0.4004955 |
| bf10_22wk | 6:39368241:minus | 0.149477 | 0.1947515 | 0.0586909 | -0.0851445 | -0.0264536 | 0.406966 |
| marb | 16:26512555:minus | 0.1619657 | 0.1868475 | 0.0774346 | 0.1086784 | 0.1861131 | 0.4127097 |
| protein | 6:39368241:minus | -0.1594778 | 0.1967758 | -0.0666801 | 0.1319136 | 0.0652335 | 0.4145938 |
| mtpro | 1:126167425:minus | 0.1737786 | 0.2237524 | 0.0577126 | -0.0202291 | 0.0374835 | 0.4149834 |
| juiciness | 16:26512645:minus | 0.2575322 | 0.3225344 | 0.0436639 | 0.0137643 | 0.0574282 | 0.4175586 |
| lma_10wk | 15:110910484:plus | -0.1654816 | 0.2136592 | -0.0581469 | -0.0701406 | -0.1282875 | 0.4184943 |
| tenderness | 6:39368241:minus | -0.1774821 | 0.20296 | -0.0658282 | -0.1548923 | -0.2207204 | 0.4194451 |
| lma_22wk | 6:39368241:minus | -0.1695973 | 0.2034665 | -0.0680204 | 0.084712 | 0.0166917 | 0.426322 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\,(\hat{\rho}_g)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| color | 1:126167425:minus | -0.2001555 | 0.2596138 | -0.052825 | -0.0123827 | -0.0652077 | 0.4275635 |
| bf10_19wk | 1:126167425:minus | 0.1635779 | 0.2164784 | 0.0586532 | -0.0306371 | 0.0280161 | 0.4306564 |
| b | 6:39368241:minus | 0.1636274 | 0.207116 | 0.0624275 | 0.0542659 | 0.1166934 | 0.4306727 |
| conn_tiss | 1:126167425:minus | -0.2156408 | 0.2938451 | -0.0441841 | 0.0627977 | 0.0186136 | 0.4372271 |
| bf10_16wk | 6:39368241:minus | 0.1472387 | 0.1921934 | 0.062496 | -0.0589849 | 0.0035111 | 0.4383716 |
| first_rib | 1:126167425:minus | 0.2100715 | 0.2651923 | 0.0536197 | 0.0957624 | 0.1493821 | 0.4442219 |
| mtfat | 16:26512645:minus | -0.1913746 | 0.2368362 | -0.0587899 | -0.0399691 | -0.098759 | 0.4466127 |
| lma_10wk | 16:26512645:minus | -0.1650294 | 0.2328298 | -0.0534743 | 0.0766793 | 0.023205 | 0.4503357 |
| ham | 1:126167425:minus | 0.1889276 | 0.2504659 | 0.0542388 | -0.097682 | -0.0434432 | 0.4528543 |
| car_wt | 6:39368241:minus | 0.1986988 | 0.2475285 | 0.0531838 | -0.1582138 | -0.10503 | 0.4536542 |
| wt_3wk | 1:126167425:minus | -0.2665715 | 0.3630597 | -0.0362631 | -0.080621 | -0.1168841 | 0.4542367 |
| bf10_10wk | 16:26512645:minus | -0.1467791 | 0.2142593 | -0.0549716 | 0.0570213 | 0.0020497 | 0.4571014 |
| picnic | 16:26512645:minus | 0.2417437 | 0.2162563 | 0.1000864 | -0.0471375 | 0.0529489 | 0.4578164 |
| car_bf10 | 1:126167425:minus | 0.1445919 | 0.225077 | 0.0449817 | -0.021583 | 0.0233987 | 0.4589563 |
| wt_13wk | 15:110910484:plus | -0.15945 | 0.2136491 | -0.0559987 | -0.1231484 | -0.1791472 | 0.4590514 |
| ham | 16:26512645:minus | 0.1812063 | 0.2405252 | 0.0553937 | -0.0554881 | -0.0000944 | 0.4606909 |
| ph_24h | 16:26512555:minus | -0.1701174 | 0.2279145 | -0.0569059 | 0.1328594 | 0.0759534 | 0.4657904 |
| lrf_22wk | 6:39368241:minus | 0.1288604 | 0.1883409 | 0.0541613 | -0.1375891 | -0.0834278 | 0.4674333 |
| car_length | 15:110910484:plus | -0.1524818 | 0.2069982 | -0.0605154 | 0.0543489 | -0.0061665 | 0.467747 |
| lma_19wk | 6:39368241:minus | -0.1558365 | 0.2129435 | -0.0575692 | 0.1094468 | 0.0518776 | 0.4682166 |
| ADG | 16:26512555:minus | -0.1478755 | 0.1897236 | -0.0682718 | -0.0023605 | -0.0706322 | 0.475393 |
| driploss | 16:26512645:minus | -0.1686144 | 0.2485945 | -0.0478864 | -0.0228969 | -0.0707833 | 0.4754865 |
| lma_10wk | 1:126167425:minus | -0.1626418 | 0.2430544 | -0.0488912 | -0.0108961 | -0.0597873 | 0.4764479 |
| lrf_10wk | 1:126167425:minus | 0.1544691 | 0.2357872 | 0.0501495 | -0.0215955 | 0.028554 | 0.4768412 |
| ADG | 16:26512645:minus | 0.1692704 | 0.2251502 | 0.0586575 | -0.0609666 | -0.0023091 | 0.4814227 |
| boston | 16:26512555:minus | 0.204255 | 0.1877892 | 0.1017839 | -0.0454032 | 0.0563807 | 0.4899681 |
| juiciness | 15:110910484:plus | -0.2069683 | 0.3062559 | -0.0382808 | 0.0854201 | 0.0471394 | 0.4907493 |
| bf10_13wk | 16:26512645:minus | 0.140599 | 0.2168903 | 0.0502488 | 0.0249588 | 0.0752076 | 0.4923723 |
| wt_birth | 1:126167425:minus | 0.1949795 | 0.2633105 | 0.0507076 | -0.1353645 | -0.0846569 | 0.495174 |
| ham | 16:26512555:minus | 0.1482028 | 0.2081321 | 0.0590085 | -0.0477163 | 0.0112922 | 0.4961451 |
| ham | 6:39368241:minus | -0.1564239 | 0.2163433 | -0.0546501 | -0.0164437 | -0.0710938 | 0.4984195 |
| last_lum | 16:26512555:minus | 0.1355909 | 0.189248 | 0.0634519 | 0.0217243 | 0.0851762 | 0.5004708 |
| juiciness | 16:26512555:minus | 0.2011498 | 0.2909916 | 0.042879 | 0.0443604 | 0.0872393 | 0.5095759 |
| fat | 15:110910484:plus | 0.1393573 | 0.1957201 | 0.0610172 | -0.0312756 | 0.0297416 | 0.5107377 |
| lrf_13wk | 1:126167425:minus | 0.1289383 | 0.2167077 | 0.0479946 | -0.0905471 | -0.0425525 | 0.5122901 |
| juiciness | 6:39368241:minus | 0.1987993 | 0.3046339 | 0.0372065 | -0.0603744 | -0.0231679 | 0.5155103 |
| wt_22wk | 6:39368241:minus | -0.1303561 | 0.200148 | -0.0504709 | -0.1340778 | -0.1845487 | 0.5275101 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\,(\hat{\rho}_g)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| fat | 1:126167425:minus | 0.1411743 | 0.2170744 | 0.0547474 | -0.0664766 | -0.0117293 | 0.5383842 |
| last_rib | 15:110910484:plus | -0.1229844 | 0.2166575 | -0.0410274 | 0.2114687 | 0.1704413 | 0.5401597 |
| fftoln | 16:26512555:minus | -0.1432472 | 0.2129389 | -0.0549381 | 0.0163072 | -0.0386309 | 0.540564 |
| WBS | 1:126167425:minus | 0.1624697 | 0.2543049 | 0.046345 | -0.0686191 | -0.0222741 | 0.5423175 |
| lma_10wk | 16:26512555:minus | -0.1152547 | 0.2033807 | -0.0476796 | 0.0312974 | -0.0163822 | 0.5474477 |
| bf10_19wk | 15:110910484:plus | 0.1205604 | 0.1997647 | 0.0473525 | 0.0171634 | 0.0645159 | 0.5543059 |
| moisture | 6:39368241:minus | -0.132149 | 0.2170971 | -0.0469339 | 0.0282772 | -0.0186566 | 0.555355 |
| boston | 1:126167425:minus | 0.2069861 | 0.2260322 | 0.0756742 | -0.0572619 | 0.0184123 | 0.5576765 |
| ADG | 1:126167425:minus | 0.1323108 | 0.2277962 | 0.0449251 | 0.0049455 | 0.0498706 | 0.5673039 |
| car_length | 16:26512645:minus | -0.1282113 | 0.2235696 | -0.0467479 | 0.0245737 | -0.0221742 | 0.5676653 |
| boston | 15:110910484:plus | -0.1702409 | 0.2073931 | -0.0688635 | -0.0423637 | -0.1112272 | 0.571351 |
| WBS | 15:110910484:plus | 0.1363886 | 0.2286853 | 0.0444778 | 0.0055971 | 0.0500749 | 0.5722162 |
| farm_wt | 6:39368241:minus | 0.1446576 | 0.2332404 | 0.0436757 | -0.147997 | -0.1043213 | 0.5722967 |
| temp_45m | 6:39368241:minus | -0.1502822 | 0.2395513 | -0.0432328 | -0.1496475 | -0.1928803 | 0.5736385 |
| spareribs | 16:26512645:minus | 0.1481132 | 0.2231766 | 0.054075 | 0.0254644 | 0.0795395 | 0.5754834 |
| ph_24h | 6:39368241:minus | 0.1386448 | 0.2404243 | 0.0403208 | -0.0342431 | 0.0060777 | 0.5763378 |
| off_flavor | 6:39368241:minus | -0.2306106 | 0.4246238 | -0.0262392 | 0.0078805 | -0.0183587 | 0.5811339 |
| bf10_13wk | 6:39368241:minus | 0.1057578 | 0.2033075 | 0.040972 | -0.0966894 | -0.0557173 | 0.5837899 |
| wt_13wk | 16:26512555:minus | -0.1132874 | 0.2060728 | -0.0461116 | 0.0090628 | -0.0370488 | 0.584874 |
| ph_24h | 1:126167425:minus | -0.1496535 | 0.2723411 | -0.0361461 | -0.069358 | -0.1055041 | 0.5885506 |
| last_rib | 16:26512555:minus | 0.105973 | 0.209809 | 0.0395416 | 0.1143323 | 0.1538739 | 0.5890126 |
| wt_22wk | 16:26512555:minus | -0.1136835 | 0.1989365 | -0.0482446 | -0.0234481 | -0.0716926 | 0.5930439 |
| Days | 1:126167425:minus | -0.1283173 | 0.2409273 | -0.0392281 | 0.0860409 | 0.0468128 | 0.5943441 |
| cook_yield | 6:39368241:minus | 0.1223936 | 0.2198685 | 0.0425102 | -0.0943446 | -0.0518344 | 0.5966366 |
| ph_45m | 16:26512555:minus | 0.1516595 | 0.275161 | 0.0361816 | 0.0260771 | 0.0622587 | 0.6006761 |
| bf10_22wk | 1:126167425:minus | 0.1042059 | 0.2219571 | 0.0347694 | -0.006106 | 0.0286634 | 0.6009449 |
| fftoln | 16:26512645:minus | -0.1379222 | 0.2432724 | -0.0414282 | -0.0112717 | -0.0526999 | 0.6021818 |
| dress_ptg | 16:26512555:minus | 0.107475 | 0.2133221 | 0.0413156 | -0.0195704 | 0.0217452 | 0.6025657 |
| lma_16wk | 1:126167425:minus | 0.1251931 | 0.2412788 | 0.0382801 | -0.0461075 | -0.0078274 | 0.604718 |
| temp_45m | 16:26512555:minus | 0.1280856 | 0.2263802 | 0.0439548 | -0.0667698 | -0.0228149 | 0.6057148 |
| car_length | 1:126167425:minus | -0.1158889 | 0.2305331 | -0.0404105 | 0.0402431 | -0.0001674 | 0.609722 |
| cook_yield | 16:26512645:minus | 0.1276 | 0.2390478 | 0.0398805 | 0.0180708 | 0.0579513 | 0.6121718 |
| lrf_13wk | 6:39368241:minus | 0.0937985 | 0.201529 | 0.0380085 | 0.0321128 | 0.0701214 | 0.6156588 |
| loin | 1:126167425:minus | -0.1354001 | 0.2568535 | -0.0373797 | 0.0560479 | 0.0186681 | 0.6203779 |
| wt_22wk | 1:126167425:minus | 0.1193795 | 0.239603 | 0.0368454 | -0.0031671 | 0.0336783 | 0.6246045 |
| Days | 16:26512645:minus | -0.1222784 | 0.2384499 | -0.0380758 | 0.0480155 | 0.0099397 | 0.6274328 |
| a | 16:26512645:minus | -0.0996815 | 0.2155526 | -0.0377495 | -0.0624146 | -0.1001641 | 0.6303711 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\,(\hat{\rho}_g)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| bf10_19wk | 16:26512555:minus | -0.0830604 | 0.1810327 | -0.0403962 | 0.0630358 | 0.0226396 | 0.6310913 |
| wt_16wk | 16:26512555:minus | -0.0992865 | 0.208538 | -0.0391869 | 0.001633 | -0.037554 | 0.6396938 |
| cook_yield | 1:126167425:minus | -0.1149566 | 0.2451556 | -0.0344387 | -0.038007 | -0.0724457 | 0.6446724 |
| first_rib | 15:110910484:plus | 0.1225542 | 0.2467426 | 0.035025 | 0.0586411 | 0.0936661 | 0.6450172 |
| lrf_13wk | 16:26512555:minus | 0.081697 | 0.1877211 | 0.039355 | -0.0173667 | 0.0219883 | 0.6457789 |
| overtend | 16:26512555:minus | -0.1016838 | 0.2043616 | -0.0421201 | 0.0801851 | 0.038065 | 0.6469964 |
| picnic | 15:110910484:plus | 0.1352035 | 0.2066383 | 0.0585721 | -0.1663812 | -0.107809 | 0.649712 |
| temp_24h | 15:110910484:plus | -0.1186655 | 0.248899 | -0.0320029 | -0.0082482 | -0.0402512 | 0.650646 |
| lrf_22wk | 16:26512555:minus | 0.0813195 | 0.1835097 | 0.0382327 | -0.00958 | 0.0286527 | 0.6523264 |
| overtend | 6:39368241:minus | -0.1002467 | 0.2041207 | -0.0376138 | -0.1831348 | -0.2207486 | 0.6534165 |
| fat | 6:39368241:minus | 0.0884439 | 0.1819734 | 0.042917 | -0.1453352 | -0.1024182 | 0.6542608 |
| bf10_10wk | 6:39368241:minus | 0.0846975 | 0.2031756 | 0.0336485 | -0.0978635 | -0.064215 | 0.6545824 |
| firm | 16:26512555:minus | -0.1025534 | 0.2486785 | -0.0295026 | -0.0389664 | -0.068469 | 0.6583117 |
| moisture | 15:110910484:plus | -0.1024474 | 0.2222102 | -0.0351406 | -0.0149755 | -0.0501161 | 0.6635611 |
| car_lma | 1:126167425:minus | -0.1052799 | 0.2144879 | -0.040782 | 0.0543422 | 0.0135603 | 0.6707484 |
| wt_13wk | 16:26512645:minus | -0.1023502 | 0.2422701 | -0.0316575 | 0.1145418 | 0.0828843 | 0.6745644 |
| dress_ptg | 6:39368241:minus | 0.0924685 | 0.2311743 | 0.0294511 | 0.0188937 | 0.0483448 | 0.6779402 |
| picnic | 6:39368241:minus | 0.1272447 | 0.2107339 | 0.0538243 | -0.1748191 | -0.1209947 | 0.6805866 |
| car_lma | 15:110910484:plus | 0.0969451 | 0.1988965 | 0.0408374 | -0.0609388 | -0.0201014 | 0.6816932 |
| dress_ptg | 1:126167425:minus | 0.0973208 | 0.2539051 | 0.0275009 | -0.0244175 | 0.0030834 | 0.6855967 |
| lma_13wk | 1:126167425:minus | 0.0914911 | 0.2352427 | 0.0300887 | -0.0351742 | -0.0050855 | 0.6868016 |
| lma_13wk | 16:26512555:minus | -0.0776127 | 0.1993232 | -0.0341269 | 0.0150029 | -0.0191239 | 0.6931593 |
| lrf_19wk | 15:110910484:plus | 0.0751761 | 0.1931471 | 0.0329649 | 0.0100871 | 0.0430521 | 0.6976633 |
| lma_19wk | 1:126167425:minus | -0.0927744 | 0.2403673 | -0.0294269 | 0.0811806 | 0.0517537 | 0.6982174 |
| off_flavor | 16:26512645:minus | -0.176229 | 0.4665122 | -0.0175972 | 0.0548617 | 0.0372645 | 0.6997954 |
| Days | 6:39368241:minus | 0.0813298 | 0.2041485 | 0.0306457 | 0.2161869 | 0.2468326 | 0.703174 |
| lma_16wk | 16:26512645:minus | 0.0955141 | 0.2430472 | 0.0290555 | -0.0345935 | -0.005538 | 0.7050698 |
| lrf_10wk | 15:110910484:plus | 0.0773913 | 0.2197702 | 0.0272958 | 0.0709857 | 0.0982815 | 0.7103721 |
| boston | 16:26512645:minus | -0.1117029 | 0.2243104 | -0.0419084 | -0.00132 | -0.0432284 | 0.7234848 |
| tofat | 6:39368241:minus | -0.0664997 | 0.2014655 | -0.0251202 | -0.1065948 | -0.131715 | 0.7289428 |
| spareribs | 16:26512555:minus | 0.0797391 | 0.1926834 | 0.0379814 | -0.0006944 | 0.037287 | 0.7310941 |
| wt_19wk | 6:39368241:minus | -0.0719486 | 0.2067507 | -0.0279894 | -0.0852592 | -0.1132486 | 0.7346876 |
| bf10_16wk | 15:110910484:plus | 0.0661742 | 0.200787 | 0.0264741 | -0.0186399 | 0.0078342 | 0.741028 |
| driploss | 1:126167425:minus | -0.0763835 | 0.2559055 | -0.0210051 | 0.0749105 | 0.0539053 | 0.7420244 |
| lrf_22wk | 1:126167425:minus | 0.0650115 | 0.2145676 | 0.0233606 | -0.0474056 | -0.0240449 | 0.7437884 |
| farm_wt | 15:110910484:plus | 0.082839 | 0.2468476 | 0.0231778 | -0.0810389 | -0.0578611 | 0.7580161 |
| lrf_13wk | 16:26512645:minus | 0.0650751 | 0.2297954 | 0.0221296 | 0.1644947 | 0.1866243 | 0.7593929 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\,(\hat{\rho}_g)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| tofat | 1:126167425:minus | 0.0656985 | 0.2358232 | 0.0203366 | 0.00887 | 0.0292066 | 0.7643431 |
| fftoln | 1:126167425:minus | 0.0820297 | 0.25393 | 0.0231438 | 0.046007 | 0.0691508 | 0.7648345 |
| car_bf10 | 15:110910484:plus | 0.0536883 | 0.2037739 | 0.0189283 | 0.0332653 | 0.0521937 | 0.7687679 |
| mtfat | 1:126167425:minus | 0.0775479 | 0.2526576 | 0.022045 | 0.0260355 | 0.0480805 | 0.768823 |
| car_lma | 16:26512645:minus | -0.07139 | 0.2071697 | -0.0288831 | -0.087058 | -0.1159411 | 0.7717547 |
| wt_birth | 16:26512645:minus | -0.0847345 | 0.2572912 | -0.0227705 | -0.1079413 | -0.1307119 | 0.7728507 |
| bf10_22wk | 16:26512555:minus | 0.0517128 | 0.1864257 | 0.0232568 | 0.0192372 | 0.0424941 | 0.7736674 |
| lrf_13wk | 15:110910484:plus | -0.0524319 | 0.2019387 | -0.02117 | -0.0445553 | -0.0657253 | 0.7801575 |
| wt_10wk | 6:39368241:minus | -0.0618979 | 0.2273147 | -0.0196546 | -0.0692724 | -0.0889271 | 0.7841681 |
| loin | 16:26512645:minus | -0.0707188 | 0.2471477 | -0.0208961 | 0.0286673 | 0.0077712 | 0.7893274 |
| belly | 6:39368241:minus | 0.0630998 | 0.2358162 | 0.0192672 | 0.057424 | 0.0766912 | 0.7901433 |
| wt_6wk | 16:26512555:minus | -0.0645583 | 0.2382651 | -0.0199841 | -0.0065554 | -0.0265395 | 0.7965679 |
| tenderness | 16:26512555:minus | -0.0545148 | 0.2082312 | -0.0219981 | 0.0527379 | 0.0307398 | 0.8063238 |
| cook_yield | 16:26512555:minus | -0.0539747 | 0.2087991 | -0.0217456 | 0.0336472 | 0.0119016 | 0.8082364 |
| lrf_19wk | 1:126167425:minus | 0.0494301 | 0.2086892 | 0.0199124 | -0.0196591 | 0.0002533 | 0.8088257 |
| lma_22wk | 16:26512555:minus | 0.0509056 | 0.1974898 | 0.0229958 | 0.039394 | 0.0623897 | 0.8113698 |
| tofat | 16:26512555:minus | -0.0431893 | 0.1957682 | -0.0184248 | -0.0064753 | -0.0249 | 0.8219577 |
| lma_22wk | 1:126167425:minus | 0.0515857 | 0.2333947 | 0.0174564 | -0.0087097 | 0.0087467 | 0.8279009 |
| marb | 6:39368241:minus | 0.045541 | 0.2009511 | 0.0188274 | -0.1487871 | -0.1299597 | 0.8307879 |
| mtpro | 6:39368241:minus | 0.0403107 | 0.1954781 | 0.0160683 | -0.1077932 | -0.0917249 | 0.8311801 |
| lma_16wk | 6:39368241:minus | -0.0469076 | 0.2236622 | -0.0158017 | 0.0891573 | 0.0733556 | 0.8366246 |
| lrf_16wk | 1:126167425:minus | -0.0424866 | 0.2216924 | -0.0154942 | -0.0039458 | -0.0194401 | 0.8391594 |
| wt_birth | 15:110910484:plus | 0.0520304 | 0.2502773 | 0.0142693 | -0.1784088 | -0.1641395 | 0.8517117 |
| num_ribs | 15:110910484:plus | -0.046148 | 0.2352905 | -0.0169746 | -0.1309972 | -0.1479718 | 0.856189 |
| temp_24h | 16:26512555:minus | -0.0414201 | 0.2307567 | -0.0136441 | 0.1769013 | 0.1632572 | 0.858759 |
| mtpro | 15:110910484:plus | -0.0351686 | 0.2054784 | -0.0130801 | 0.0281308 | 0.0150508 | 0.8607926 |
| wt_13wk | 6:39368241:minus | 0.0382724 | 0.2164188 | 0.0136476 | -0.1217133 | -0.1080657 | 0.861163 |
| protein | 16:26512645:minus | -0.0381909 | 0.2288531 | -0.012975 | -0.1924175 | -0.2053924 | 0.8628197 |
| wt_16wk | 16:26512645:minus | -0.0406553 | 0.242841 | -0.0123087 | -0.0012041 | -0.0135129 | 0.8666131 |
| driploss | 16:26512555:minus | 0.0356692 | 0.2150485 | 0.0133817 | 0.0546035 | 0.0679852 | 0.8666922 |
| temp_45m | 16:26512645:minus | 0.048772 | 0.263157 | 0.0127654 | 0.0254079 | 0.0381732 | 0.8706103 |
| wt_birth | 6:39368241:minus | -0.0440749 | 0.2429699 | -0.0127384 | -0.0959784 | -0.1087168 | 0.872635 |
| bf10_10wk | 16:26512555:minus | -0.026251 | 0.1902752 | -0.0122417 | 0.0740019 | 0.0617602 | 0.8791091 |
| bf10_13wk | 16:26512555:minus | 0.026578 | 0.1840945 | 0.0125376 | 0.1185115 | 0.1310492 | 0.8799535 |
| car_lma | 6:39368241:minus | -0.0329626 | 0.1913465 | -0.0147553 | 0.0382447 | 0.0234894 | 0.8824579 |
| bf10_16wk | 1:126167425:minus | -0.0300976 | 0.2187554 | -0.0108983 | -0.0144716 | -0.02537 | 0.8859281 |
| color | 16:26512645:minus | -0.0344413 | 0.2520494 | -0.0097509 | -0.0440156 | -0.0537665 | 0.8901388 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$[b] | $SE\,(\hat{\rho}_g)$[c] | $\hat{\sigma}_{g_1 g_2}$[d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$[e] | $\hat{\sigma}_{p_1 p_2}$[f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| car_bf10 | 16:26512555:minus | -0.0224438 | 0.1941115 | -0.0090602 | 0.0960906 | 0.0870304 | 0.8950399 |
| lma_19wk | 16:26512645:minus | 0.0319783 | 0.2340198 | 0.0106262 | 0.044185 | 0.0548111 | 0.8971709 |
| first_rib | 16:26512555:minus | -0.0277208 | 0.2235411 | -0.009995 | 0.0842241 | 0.0742292 | 0.9048467 |
| WBS | 6:39368241:minus | 0.025388 | 0.2102882 | 0.0093911 | 0.1383409 | 0.147732 | 0.9072824 |
| firm | 1:126167425:minus | 0.0285841 | 0.2869703 | 0.0062276 | 0.1133976 | 0.1196252 | 0.9114846 |
| lma_19wk | 16:26512555:minus | 0.0226767 | 0.2007018 | 0.0098648 | 0.0242411 | 0.0341059 | 0.9115242 |
| picnic | 1:126167425:minus | 0.039421 | 0.2252794 | 0.0156001 | -0.0449129 | -0.0293128 | 0.9124973 |
| color | 6:39368241:minus | -0.0237271 | 0.2353243 | -0.0073135 | -0.0191298 | -0.0264433 | 0.9183845 |
| wt_22wk | 16:26512645:minus | 0.0247414 | 0.2337634 | 0.0079986 | -0.0444655 | -0.0364669 | 0.92508 |
| lrf_16wk | 16:26512555:minus | 0.016846 | 0.1865787 | 0.0081234 | 0.1381554 | 0.1462788 | 0.9295823 |
| lrf_22wk | 15:110910484:plus | -0.0162969 | 0.1992203 | -0.0063776 | -0.0185724 | -0.02495 | 0.9318165 |
| last_rib | 6:39368241:minus | 0.0174823 | 0.2285796 | 0.0054738 | 0.0868618 | 0.0923356 | 0.9339904 |
| wt_19wk | 1:126167425:minus | -0.0182972 | 0.2358962 | -0.006036 | 0.0065993 | 0.0005632 | 0.9379897 |
| conn_tiss | 16:26512645:minus | -0.0176857 | 0.287292 | -0.003862 | -0.000251 | -0.004113 | 0.9486608 |
| wt_13wk | 1:126167425:minus | 0.0154587 | 0.2464528 | 0.0046351 | -0.0459512 | -0.0413161 | 0.9505577 |
| marb | 1:126167425:minus | 0.0147371 | 0.232828 | 0.0050407 | -0.0251173 | -0.0200765 | 0.953837 |
| L | 6:39368241:minus | -0.0124359 | 0.2145004 | -0.0046694 | 0.0849619 | 0.0802924 | 0.9555233 |
| firm | 15:110910484:plus | -0.0138502 | 0.2668092 | -0.0033419 | 0.0652463 | 0.0619045 | 0.9562916 |
| lma_13wk | 15:110910484:plus | 0.0127701 | 0.2155959 | 0.0046818 | -0.0368351 | -0.0321533 | 0.9581677 |
| lrf_10wk | 16:26512645:minus | 0.0120285 | 0.2349942 | 0.0039431 | 0.150634 | 0.1545772 | 0.9589063 |
| lrf_19wk | 16:26512555:minus | -0.0091436 | 0.1793327 | -0.0047481 | 0.0902383 | 0.0854902 | 0.9596086 |
| tenderness | 1:126167425:minus | 0.0132823 | 0.2489693 | 0.0039214 | -0.0261136 | -0.0221922 | 0.9600551 |
| wt_19wk | 16:26512645:minus | -0.0088705 | 0.2327081 | -0.0030016 | 0.0538658 | 0.0508643 | 0.9702132 |
| wt_16wk | 6:39368241:minus | -0.0086343 | 0.2199931 | -0.0029587 | -0.0682485 | -0.0712071 | 0.9718213 |
| lma_16wk | 16:26512555:minus | 0.0074161 | 0.209505 | 0.0029484 | 0.0505676 | 0.053516 | 0.9731017 |
| bf10_16wk | 16:26512555:minus | 0.005716 | 0.1846633 | 0.002747 | 0.0755197 | 0.0782668 | 0.9751567 |
| wt_16wk | 1:126167425:minus | -0.0061446 | 0.2494703 | -0.0017801 | 0.0408541 | 0.039074 | 0.9826452 |
| bf10_10wk | 15:110910484:plus | 0.0038918 | 0.2092594 | 0.0014812 | -0.0027238 | -0.0012427 | 0.9846833 |
| lma_13wk | 16:26512645:minus | -0.0047349 | 0.235922 | -0.0015644 | 0.0485605 | 0.0469961 | 0.9847688 |
| lma_16wk | 15:110910484:plus | 0.0045142 | 0.224057 | 0.0015134 | -0.0696241 | -0.0681107 | 0.9853733 |
| spareribs | 1:126167425:minus | 0.0101337 | 0.2309478 | 0.0035499 | 0.0160492 | 0.0195991 | 0.9861807 |
| lrf_10wk | 16:26512555:minus | 0.0053121 | 0.2030586 | 0.0022666 | 0.079236 | 0.0815026 | 0.9871669 |
| loin | 6:39368241:minus | -0.0031743 | 0.2258024 | -0.0010487 | -0.056781 | -0.0578296 | 0.9910243 |
| car_length | 6:39368241:minus | 0.0023048 | 0.2052394 | 0.0009366 | -0.0153793 | -0.0144427 | 0.9913493 |
| off_flavor | 16:26512555:minus | 0.0018807 | 0.4125968 | 0.0002467 | -0.0378749 | -0.0376282 | 0.9967745 |
| wt_10wk | 16:26512555:minus | 0.0026681 | 0.2191706 | 0.0009652 | 0.0077715 | 0.0087367 | 0.9970458 |
| lrf_16wk | 15:110910484:plus | 0.001788 | 0.2008969 | 0.0007346 | -0.0522624 | -0.0515279 | 1 |

Table B.3 (cont'd)

| trait[a] | editing site | $\hat{\rho}_g$ [b] | $SE\,(\hat{\rho}_g)$ [c] | $\hat{\sigma}_{g_1 g_2}$ [d] | $\hat{\sigma}_{\varepsilon_1 \varepsilon_2}$ [e] | $\hat{\sigma}_{p_1 p_2}$ [f] | $p$-value[g] |
|---|---|---|---|---|---|---|---|
| mtpro | 16:26512555:minus | 0.0015277 | 0.1899246 | 0.0006824 | 0.0161665 | 0.0168489 | 1 |
| overtend | 1:126167425:minus | -0.0024682 | 0.2488897 | -0.0007309 | 0.032218 | 0.0314872 | 1 |

[a] More information about each trait can be found in Velez-Irizarry et al. [106]
[b] Genetic correlation estimate, where $\rho_g = \sigma_{g_1 g_2} / \sqrt{\sigma^2_{g_1} \sigma^2_{g_2}}$
[c] Standard error of genetic correlation estimate
[d] Genomic covariance REML estimate
[e] Residual covariance REML estimate
[f] Phenotypic covariance estimate
[g] $P$-value testing $H_0 : \sigma_{g_1 g_2}$

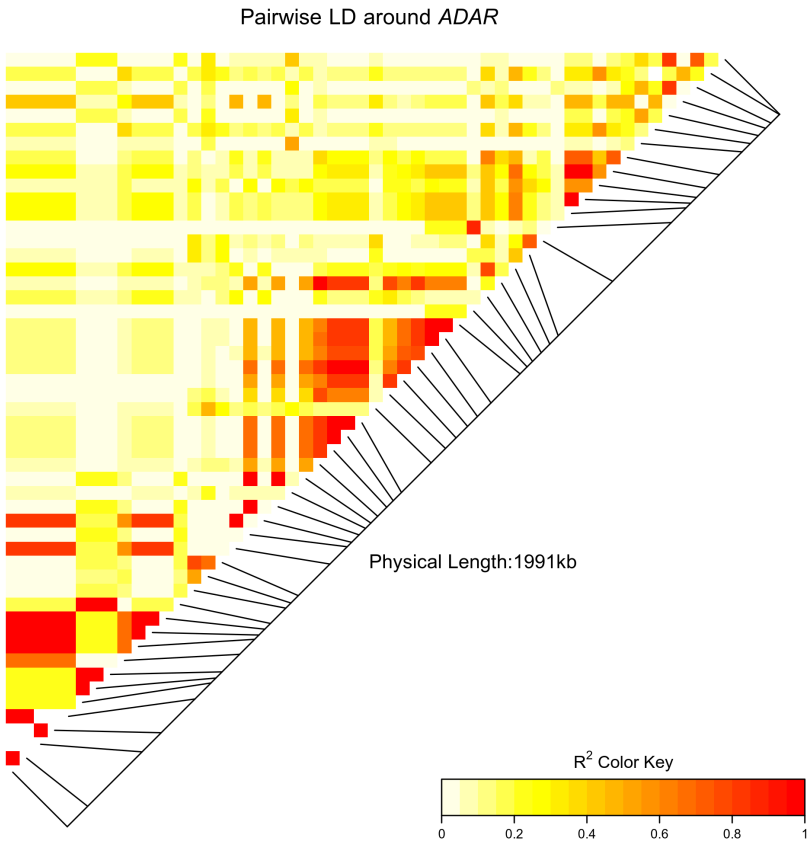


Figure B.1: Pairwise LD plot between SNPs flanking *ADAR*

**BIBLIOGRAPHY**

## REFERENCES


[1] Peter Donnelly. Progress and challenges in genome-wide association studies in humans. *Nature*, 456(7223):728–731, December 2008.

[2] Lucia A. Hindorff, Praveen Sethupathy, Heather A. Junkins, Erin M. Ramos, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106(23):9362–9367, June 2009.

[3] Brendan Maher. Personal genomes: The case of the missing heritability. *Nature*, 456(7218):18–21, November 2008.

[4] Teri A. Manolio, Francis S. Collins, Nancy J. Cox, David B. Goldstein, et al. Finding the missing heritability of complex diseases. *Nature*, 461(7265):747–753, October 2009.

[5] Jian Yang, Beben Benyamin, Brian P. McEvoy, Scott Gordon, et al. Common snps explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7):565–569, 2010.

[6] Gustavo de los Campos, Daniel Sorensen, and Daniel Gianola. Genomic heritability: What is it? *PLoS Genetics*, 11(5):1–21, 2015.

[7] Felix R. Day, Hannes Helgason, Daniel I. Chasman, Lynda M. Rose, et al. Physical and neurobehavioral determinants of reproductive onset and success. *Nature Genetics*, 48(6):617–623, June 2016.

[8] John R. Shaffer, Jinxi Li, Myoung Keun Lee, Jasmien Roosenboom, et al. Multiethnic gwas reveals polygenic architecture of earlobe attachment. *The American Journal of Human Genetics*, 101(6):913–924, December 2017.

[9] Barbara Schormair, Chen Zhao, Steven Bell, Erik Tilch, et al. Identification of novel risk loci for restless legs syndrome in genome-wide association studies in individuals of european ancestry: a meta-analysis. *The Lancet Neurology*, 16(11):898–907, November 2017.

[10] Joëlle A. Pasman, Karin J. H. Verweij, Zachary Gerring, Sven Stringer, et al. Gwas of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability. *Nature Neuroscience*, 21(9):1161–1170, September 2018.

[11] Loic Yengo, Julia Sidorenko, Kathryn E. Kemper, Zhili Zheng, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of european ancestry. *Human Molecular Genetics*, 27(20):3641–3649, October 2018.

[12] M. C. Zillikens, M. Yazdanpanah, L. M. Pardo, F. Rivadeneira, et al. Sex-specific genetic effects influence variation in body composition. *Diabetologia*, 51(12):2233–2241, 2008.

[13] Jian Yang, Andrew Bakshi, Zhihong Zhu, Gibran Hemani, et al. Genome-wide genetic homogeneity between sexes and populations for human height and body mass index. *Human Molecular Genetics*, 24(25):7445–7449, 2015.

[14] Konrad Rawlik, Oriol Canela-Xandri, and Albert Tenesa. Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome Biology*, 17(1):166, December 2016.

[15] Bruce Walsh and Michael Lynch. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc, 1 edition, 1998.

[16] Joshua C. Randall, Thomas W. Winkler, Zoltán Kutalik, Sonja I. Berndt, et al. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genetics*, 9(6):e1003500, June 2013.

[17] Ching-Ti Liu, Karol Estrada, Laura M. Yerges-Armstrong, Najaf Amin, et al. Assessment of gene-by-sex interaction effect on bone mineral density. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*, 27(10):2051–64, October 2012.

[18] Andrew DeWan, Mugen Liu, Stephen Hartman, Samuel Shao-Min Zhang, et al. Htra1 promoter polymorphism in wet age-related macular degeneration. *Science*, 314(5801):989–992, 2006.

[19] Kristin L. Ayers and Heather J. Cordell. Snp selection in genome-wide and candidate gene studies via penalized logistic regression. *Genetic Epidemiology*, 34(8):879–891, December 2010.

[20] Yongtao Guan and Matthew Stephens. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Annals of Applied Statistics*, 5(3):1780–1815, 2011.

[21] Hui Yi, Patrick Breheny, Netsanet Imam, Yongmei Liu, et al. Penalized multimarker vs. single-marker regression methods for genome-wide association studies of quantitative traits. *Genetics*, 199(1):205–222, January 2015.

[22] Jeremy A. Sabourin, William Valdar, and Andrew B. Nobel. A permutation approach for selecting the penalty parameter in penalized model selection. *Biometrics*, 71(4):1185–1194, December 2015.

[23] T. H. E. Meuwissen, B. J. Hayes, and M. E. Goddard. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829, 2001.

[24] David Habier, Rohan L. Fernando, Kadir Kizilkaya, and Dorian J. Garrick. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12, 2011.

[25] James G. Scott and James O. Berger. Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38(5):2587–2619, October 2010.

[26] Rohan Fernando, Ali Toosi, Anna Wolc, Dorian Garrick, et al. Application of whole-genome prediction methods for genome-wide association studies: A bayesian approach. *Journal of Agricultural, Biological, and Environmental Statistics*, 22(2):172–193, 2017.

[27] Peter M. Visscher, Matthew A. Brown, Mark I. McCarthy, and Jian Yang. Five years of gwas discovery. *The American Journal of Human Genetics*, 90(1):7–24, January 2012.

[28] Dan L. Nicolae, Eric Gamazon, Wei Zhang, Shiwei Duan, et al. Trait-associated snps are more likely to be eqtls: Annotation to enhance discovery from gwas. *PLoS Genetics*, 6(4):e1000888, April 2010.

[29] Yang I. Li, Bryce van de Geijn, Anil Raj, David A. Knowles, et al. Rna splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, 2016.

[30] Rob Benne, Janny Van Den Burg, Just P.J. Brakenhoff, Paul Sloof, et al. Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the dna. *Cell*, 46(6):819–826, September 1986.

[31] Brenda L. Bass and Harold Weintraub. An unwinding activity that covalently modifies its double-stranded rna substrate. *Cell*, 55(6):1089–1098, December 1988.

[32] Alekos Athanasiadis, Alexander Rich, and Stefan Maas. Widespread a-to-i rna editing of alu-containing mrnas in the human transcriptome. *PLoS Biology*, 2(12):e391, November 2004.

[33] Miyoko Higuchi, Stefan Maas, Frank N. Single, Jochen Hartner, et al. Point mutation in an ampa receptor gene rescues lethality in mice deficient in the rna-editing enzyme adar2. *Nature*, 406(6791):78–81, 2000.

[34] Yukio Kawahara, Adda Grimberg, Sarah Teegarden, Cedric Mombereau, et al. Dysregulated editing of serotonin 2c receptor mrnas results in energy dissipation and loss of fat mass. *J Neurosci*, 28(48):12834–12844, 2008.

[35] Meng How Tan, Qin Li, Raghuvaran Shanmugam, Robert Piskol, et al. Dynamic landscape and regulation of rna editing in mammals. *Nature*, 550(7675):249–254, 2017.

[36] Zishuai Wang, Xikang Feng, Zhonglin Tang, and Shuai Cheng Li. Genome-wide investigation and functional analysis of sus scrofa rna editing sites across eleven tissues. *Genes*, 10(5):327, April 2019.

[37] Yuebo Zhang, Longchao Zhang, Jingwei Yue, Xia Wei, et al. Genome-wide identification of rna editing in seven porcine tissues by matched dna and rna high-throughput sequencing. *Journal of Animal Science and Biotechnology*, 10(1):1–14, 2019.

[38] Lauren A. Weiss, Lin Pan, Mark Abney, and Carole Ober. The sex-specific genetic architecture of quantitative traits in humans. *Nature Genetics*, 38(2):218–222, 2006.

[39] Thomas J. Hoffmann, Georg B. Ehret, Priyanka Nandakumar, Dilrini Ranatunga, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nature Genetics*, 49(1):54–64, 2017.

[40] Gustavo de los Campos, Daniel Gianola, and David B. Allison. Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nature Reviews Genetics*, 11(12):880–886, December 2010.

[41] Gustavo de los Campos, Ana I. Vazquez, Rohan Fernando, Yann C. Klimentidis, et al. Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genetics*, 9(7), 2013.

[42] Louis Lello, Steven G. Avery, Laurent Tellier, Ana I. Vazquez, et al. Accurate genomic prediction of human height. *Genetics*, 210(2):477–497, October 2018.

[43] Nengjun Yi, Varghese George, and David B. Allison. Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics*, 164(3):1129 – 1138, July 2003.

[44] Jian Zeng, Marcin Pszczola, Anna Wolc, Tomasz Strabel, et al. Genomic breeding value prediction and qtl mapping of qtlmas2011 data using bayesian and gblup methods. *BMC Proceedings*, 6(SUPPL. 2):S13, 2012.

[45] Bjarni J. Vilhjálmsson, Jian Yang, Hilary K. Finucane, Alexander Gusev, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *American Journal of Human Genetics*, 97(4):576–592, 2015.

[46] Gustavo de los Campos, Yogasudha Veturi, Ana I. Vazquez, Christina Lehermeier, et al. Incorporating genetic heterogeneity in whole-genome regressions using interactions. *Journal of Agricultural, Biological, and Environmental Statistics*, 20(4):467–490, 2015.

[47] Yogasudha Veturi, Gustavo de los Campos, Nengjun Yi, Wen Huang, et al. Modeling heterogeneity in the genetic architecture of ethnically diverse groups using random effect interaction models. *Genetics*, 211(4):1395–1407, 2019.

[48] Iris M. Heid, Anne U. Jackson, Joshua C. Randall, Thomas W. Winkler, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genetics*, 42(11):949–960, 2010.

[49] Thomas W. Winkler, Anne E. Justice, Mariaelisa Graff, Llilda Barata, et al. The influence of age and sex on genetic associations with adult body size and shape: A large-scale genome-wide interaction study. *PLOS Genetics*, 11(10):e1005378, October 2015.

[50] Dmitry Shungin, Thomas W. Winkler, Damien C. Croteau-Chonka, Teresa Ferreira, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, 518(7538):187–196, 2015.

[51] David Karasik and S. L. Ferrari. Contribution of gender-specific genetic factors to osteoporosis risk. *Annals of Human Genetics*, 72(5):696–714, 2008.

[52] Paulino Pérez and Gustavo De Los Campos. Genome-wide regression and prediction with the bglr statistical package. *Genetics*, 198(2):483–495, 2014.

[53] Serena Sanna, Anne U. Jackson, Ramaiah Nagaraja, Cristen J. Willer, et al. Common variants in the gdf5-uqcc region are associated with variation in human height. *Nature Genetics*, 40(2):198–203, 2008.

[54] María Correa-Rodríguez, Jacqueline Schmidt Rio-Valle, and Blanca Rueda-Medina. Akap11 gene polymorphism is associated with bone mass measured by quantitative ultrasound in young adults. *International Journal of Medical Sciences*, 15(10):999–1004, 2018.

[55] Dong-Li Zhu, Xiao-Feng Chen, Wei-Xin Hu, Shan-Shan Dong, et al. Multiple functional variants at 13q14 risk locus for osteoporosis regulate ranklexpression through long-range super-enhancer. *Journal of Bone and Mineral Research*, 33(7), 2018.

[56] Benjamin H. Mullin, John P. Walsh, Hou Feng Zheng, Suzanne J. Brown, et al. Genome-wide association study using family-based cohorts identifies the wls and ccdc170/esr1 loci as associated with bone mineral density. *BMC Genomics*, 17(1):1–11, 2016.

[57] Benjamin H. Mullin, Jing Hua Zhao, Suzanne J. Brown, John R.B. Perry, et al. Genome-wide association study meta-analysis for quantitative ultrasound parameters of bone identifies five novel loci for broadband ultrasound attenuation. *Human molecular genetics*, 26(14):2791–2802, 2017.

[58] Adam E. Locke, Bratati Kahali, Sonja I. Berndt, Anne E. Justice, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538):197–206, 2015.

[59] Jong Weon Choi and Soo Hwan Pai. Associations between abo blood groups and osteoporosis in postmenopausal women. *Annals of Clinical and Laboratory Science*, 34(2):150–153, 2004.

[60] B. B. Lu and K. H. Li. Association between abo blood groups and osteoporosis severity in chinese adults aged 50 years and over. *Journal of International Medical Research*, 39(3):929–933, 2011.

[61] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, et al. Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, 12(3):e1001779, 2015.

[62] Scott A. Funkhouser, Juan P. Steibel, Ronald O. Bates, Nancy E. Raney, Darius Schenk, et al. Evidence for transcriptome-wide rna editing among sus scrofa pre-1 sine elements. *BMC Genomics*, 18:360, 2017.

[63] Matthew Blow, P. Andrew Futreal, Richard Wooster, and Michael R. Stratton. A survey of rna editing in human brain. *Genome Research*, 14(12):2379–2387, November 2004.

[64] Erez Y. Levanon, Eli Eisenberg, Rodrigo Yelin, Sergey Nemzer, et al. Systematic identification of abundant a-to-i editing sites in the human transcriptome. *Nature Biotechnology*, 22(8):1001–1005, 2004.

[65] Eli Eisenberg, Sergey Nemzer, Yaron Kinar, Rotem Sorek, et al. Is abundant a-to-i rna editing primate-specific? *Trends in Genetics*, 21(2):77–81, 2005.

[66] Yossef Neeman, Erez Y. Levanon, Michael F. Jantsch, and Eli Eisenberg. Rna editing level in the mouse is determined by the genomic repeat repertoire. *RNA*, 12(10):1802–9, 2006.

[67] Lily Bazak, Ami Haviv, Michal Barak, Jasmine Jacob-Hirsch, et al. A-to-i rna editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Research*, 24(3):365–376, March 2014.

[68] Jia Yu Chen, Zhiyu Peng, Rongli Zhang, Xin Zhuang Yang, et al. Rna editome in rhesus macaque shaped by purifying selection. *PLoS Genetics*, 10(4):8–12, 2014.

[69] Eric S. Lander, Lauren M. Linton, Bruce Birren, Chad Nusbaum, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.

[70] Lily Bazak, Erez Y. Levanon, and Eli Eisenberg. Genome-wide analysis of alu editability. *Nucleic Acids Research*, 42(11):6876–6884, June 2014.

[71] Nikita S. Vassetzky and Dmitri A. Kramerov. Sinebase: A database and tool for sine analysis. *Nucleic Acids Research*, 41(D1):83–89, 2013.

[72] Galit Lev-Maor, Rotem Sorek, Erez Y. Levanon, Nurit Paz, et al. Rna-editing-mediated exon evolution. *Genome Biology*, 8(2):R29, 2007.

[73] A. D. J. Scadden and Christopher W. J. Smith. Rnai is antagonized by a–>i hyper-editing. *EMBO reports*, 2(12):1107–1111, December 2001.

[74] Claudia L. Kleinman, Véronique Adoue, and Jacek Majewski. Rna editing of protein sequences: A rare event in human transcriptomes. *RNA*, 18(9):1586–1596, September 2012.

[75] Gokul Ramaswami, Wei Lin, Robert Piskol, Meng How Tan, et al. Accurate identification of human alu and non-alu rna editing sites. *Nature Methods*, 9(6):579–581, June 2012.

[76] Wei Lin, Robert Piskol, Meng How Tan, and Jin Billy Li. Comment on "widespread rna and dna sequence differences in the human transcriptome". *Science*, 335(6074):1302, March 2012.

[77] Heng Li. A statistical framework for snp calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21):2987–2993, 2011.

[78] Sarah N. Deffit and Heather A. Hundley. To edit or not to edit: Regulation of adar editing specificity and efficiency. *Wiley Interdisciplinary Reviews: RNA*, 7(1):113–127, 2016.

[79] Joseph K. Pickrell, Yoav Gilad, and Jonathan K. Pritchard. Comment on "widespread rna and dna sequence differences in the human transcriptome". *Science*, 335(6074):1302, 2012.

[80] Heng Li, Jue Ruan, and Richard Durbin. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, 18(11):1851–1858, November 2008.

[81] Xueda Hu, Shengqing Wan, Ying Ou, Boping Zhou, et al. Rna over-editing of blcap contributes to hepatocarcinogenesis identified by whole-genome and transcriptome sequencing. *Cancer Letters*, 357(2):510–519, 2015.

[82] Chammiran Daniel, Gilad Silberberg, Mikaela Behm, and Marie Öhman. Alu elements shape the primate transcriptome by cis-regulation of rna editing. *Genome Biology*, 15(2):R28, 2014.

[83] A.F.A. Smit, R. Hubley, and P. Green. Repeatmasker open-4.0, 2013.

[84] J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, et al. Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, 110(1-4):462–467, 2005.

[85] D. B. Edwards, C. W. Ernst, R. J. Tempelman, G. J M Rosa, et al. Quantitative trait loci mapping in an f2 duroc x pietrain resource population: I. growth traits. *Journal of Animal Science*, 86(2):241–253, 2008.

[86] S Andrews. Fastqc: a quality control tool for high throughput sequence data, 2010.

[87] Linnéa Smeds and Axel Künstner. Condetri - a content dependent read trimmer for illumina data. *PLoS ONE*, 6(10):e26314, October 2011.

[88] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357–9, 2012.

[89] Cole Trapnell, Lior Pachter, and Steven L. Salzberg. Tophat: Discovering splice junctions with rna-seq. *Bioinformatics*, 25(9):1105–1111, 2009.

[90] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[91] Rachel B. Brem, Gaël Yvert, Rebecca Clinton, and Leonid Kruglyak. Genetic dissection of transcriptional regulation in budding yeast. *Science*, 296(5568):752–755, April 2002.

[92] Ritsert C. Jansen and J. P. Nap. Genetical genomics: the added value from segregation. *Trends in genetics : TIG*, 17(7):388–91, July 2001.

[93] K. G. Ardlie, D. S. Deluca, A. V. Segre, T. J. Sullivan, et al. The genotype-tissue expression (gtex) pilot analysis: Multitissue gene regulation in humans. *Science*, 348(6235):648–660, May 2015.

[94] Alexis Battle, Sara Mostafavi, X. Zhu, J. B. Potash, et al. Characterizing the genetic basis of transcriptome diversity through rna-sequencing of 922 individuals. *Genome Research*, 24(1):14–24, January 2014.

[95] Gokul Ramaswami, Patricia Deng, Rui Zhang, Mary Anna Carbone, et al. Genetic mapping uncovers cis-regulatory landscape of rna editing. *Nature Communications*, 6(1):8194, November 2015.

[96] Yerbol Z. Kurmangaliyev, Sammi Ali, and Sergey V. Nuzhdin. Genetic determinants of rna editing levels of adar targets in drosophila melanogaster. *Genes|Genomes|Genetics*, 6(2):391–396, February 2016.

[97] Tongjun Gu, Daniel M. Gatti, Anuj Srivastava, Elizabeth M. Snyder, et al. Genetic architectures of quantitative variation in rna editing pathways. *Genetics*, 202(2):787–798, 2016.

[98] Eddie Park, Jiguang Guo, Shihao Shen, Levon Demirdjian, et al. Population and allelic variation of a-to-i rna editing in human transcriptomes. *Genome Biology*, 18(1):1–15, 2017.

[99] Thomas J. Lopdell, Victoria Hawkins, Christine Couldrey, Kathryn Tiplady, et al. Widespread cis -regulation of rna editing in a large mammal. *RNA*, 25(3):319–335, March 2019.

[100] Wouter van Rheenen, Wouter J. Peyrot, Andrew J. Schork, S. Hong Lee, et al. Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews Genetics*, 2019.

[101] Petr Danecek, Christoffer Nellåker, Rebecca E. McIntyre, Jorge E. Buendia-Buendia, et al. High levels of rna-editing site conservation amongst 15 laboratory mouse strains. *Genome Biology*, 13(4), 2012.

[102] J.-H. Shin, S. Blay, B. McNeney, and J. Graham. Ldheatmap: An r function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Soft*, 16:Code Snippet 3, 2006.

[103] Elizabeth A. Hackler, David C. Airey, Caitlin C. Shannon, Monsheel S. Sodhi, et al. 5-ht2c receptor rna editing in the amygdala of c57bl/6j, dba/2j, and balb/cj mice. *Neuroscience Research*, 55(1):96–104, 2006.

[104] E. Lander and N. Schork. Genetic dissection of complex traits. *Science*, 265(5181):2037–2048, September 1994.

[105] Natalia S. Forneris, Andres Legarra, Zulma G. Vitezica, Shogo Tsuruta, et al. Quality control of genotypes using heritability estimates of gene content at the marker. *Genetics*, 199(3):675–681, 2015.

[106] Deborah Velez-Irizarry, Sebastian Casiro, Kaitlyn R. Daza, Ronald O. Bates, et al. Genetic control of longissimus dorsi muscle gene expression variation and joint analysis with phenotypic quantitative trait loci in pigs. *BMC Genomics*, 20(1):3, December 2019.

[107] S. Casiró, D. Velez-Irizarry, C. W. Ernst, N. E. Raney, et al. Genome-wide association study in an f2 duroc x pietrain resource population for economically important meat quality and carcass traits. *Journal of Animal Science*, 95(2):545–558, 2017.

[108] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, August 2014.

[109] Jose L. Gualdrón Duarte, Rodolfo J.C. Cantet, Ronald O. Bates, Catherine W. Ernst, et al. Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinformatics*, 15(1):246, 2014.

[110] Y. L. Bernal Rubio, J. L. Gualdrón Duarte, R. O. Bates, C. W. Ernst, et al. Implementing meta-analysis from genome-wide association studies for pork quality traits1. *Journal of Animal Science*, 93(12):5607–5617, December 2015.

[111] Min Kang Hyun, Noah A. Zaitlen, Claire M. Wade, Andrew Kirby, et al. Efficient control of population structure in model organism association mapping. *Genetics*, 178(3):1709–1723, 2008.

[112] Steven G. Self and Kung-yee Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

[113] Troy M. Fischer, Arthur R. Gilmour, and Julius Werf. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information reml. *Genetics Selection Evolution*, 36(3):363, 2004.

[114] Tomaz Berisa and Joseph K. Pickrell. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, 32(2):283–285, January 2016.

[115] Chen Yao, Roby Joehanes, Andrew D. Johnson, Tianxiao Huan, et al. Sex- and age-interacting eqtls in human complex diseases. *Human Molecular Genetics*, 23(7):1947–1956, April 2014.