Testing pleiotropy vs. separate QTL in multiparental populations

by

Frederick Joseph Boehm, III

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2019

Date of final oral examination: April 22, 2019

The dissertation is approved by the following members of the Final Oral Committee:

Karl W. Broman, Professor, Biostatistics and Medical Informatics

Brian S. Yandell, Professor, Statistics and Horticulture

Jun Zhu, Professor, Statistics and Entomology

Alan D. Attie, Professor, Biochemistry

Qiongshi Lu, Assistant professor, Biostatistics and Medical Informatics

Acknowledgements

I'm forever grateful for my time in the statistics department at the University of Wisconsin-Madison. From my first days as a graduate student, I encountered classmates, faculty, and staff who welcomed and supported me. The rigor and high standards that they encouraged me to pursue have had an enduring impact on my views of scholarly research. I am fortunate to have had continuous financial support for my studies, in the form of assistantships or fellowships.

Professor Yazhen Wang encouraged me to aspire to greater understanding of theoretical statistics. At times, I didn't think that I could master such material, yet he believed in me.

I found receptive collaborators in departments that, historically, didn't emphasize statistics.

I found support for professional development activities through the university's Delta program.

With love and gratitude to my family, friends, and teachers

Contents

1 Introduction **1**

1.1 *Agouti* biology & pleiotropy . . . 2

1.2 Designs for two-parent crosses . . . 2

1.3 QTL mapping in two-parent crosses . . . 6

1.3.1 Overview . . . 6

1.3.2 Marker regression . . . 10

1.3.3 Interval mapping . . . 12

1.3.4 Haley-Knott regression . . . 13

1.4 Testing pleiotropy in two-parent crosses . . . 14

1.5 Designs for multiparental populations and Diversity Outbred mice . . . 18

1.6 QTL mapping in Diversity Outbred mice . . . 20

1.7 Potential benefits of a pleiotropy test in multiparental populations . . . 21

2 Methods development **23**

2.1 Introduction . . . 23

2.2 Methods . . . 24

2.3 Data structures . . . 24

2.3.1 Statistical Models . . . 25

2.3.2 Parameter inference and log likelihood calculation . . . 26

2.3.3 Pleiotropy vs. separate QTL hypothesis testing framework . . . 26

2.3.4 Visualizing profile LOD traces . . . 27

2.3.5 Bootstrap for test statistic calibration . . . 27

2.3.6 Data & Software Availability . . . 28

2.4 Simulation studies . . . 28
2.4.1 Type I error rate analysis . . . 28
2.4.2 Power analysis . . . 29
2.5 Application . . . 31
2.6 Discussion . . . 31

3 Applications **38**
3.1 Expression trait hotspot dissection . . . 38
3.1.1 Introduction . . . 38
3.1.2 Methods . . . 43
3.1.3 Results . . . 45
3.1.4 Discussion . . . 47
3.2 Power analyses . . . 51
3.2.1 Introduction . . . 51
3.2.2 Methods . . . 52
3.2.3 Results . . . 53
3.2.4 Discussion . . . 56
3.3 Microbiome case study . . . 59
3.3.1 Introduction . . . 59
3.3.2 Methods . . . 60
3.3.3 Results . . . 62
3.3.4 Discussion . . . 67

4 Computing vignettes **68**
4.1 Pleiotropy testing . . . 68
4.1.1 Installing `qtl2pleio` . . . 68
4.1.2 Reading data from `qtl2data` repository on github . . . 69
4.1.3 Kinship calculations . . . 70
4.1.4 Statistical model . . . 70
4.1.5 Simulating phenotypes with `qtl2pleio::sim1` . . . 71
4.1.6 Perform two-dimensional scan as first step in pleiotropy vs. separate QTL hypothesis test . . . 74

4.1.7 Session info 79
4.2 Recla analysis 85
4.3 Cluster computing for bootstrap test 85

5 Conclusions **86**
5.1 A multivariate pleiotropy test 86
5.2 X chromosome 87
5.3 Sensitivity analysis for unmeasured confounding in mediation analysis 87

Appendix A Molecular biology primer **97**
A.1 DNA structure 97
A.2 Cell division 97
A.3 Central dogma of molecular biology 97

Appendix B Statistical inference primer **98**
B.1 Overview of statistics 99
B.2 Maximum likelihood methods 99
B.3 Restricted maximum likelihood methods 99
B.4 Genotype inference with expectation-maximization 100
B.5 Genotype inference with hidden Markov models 100

Appendix C Supplementary materials for Chapter 2 **101**

Appendix D Supplementary materials for Chapter 3 **105**

List of Figures

1.1 *Agouti* viable-yellow mutant mouse has yellow hair and is obese in comparison to a wild-type mouse. (https://upload.wikimedia.org/wikipedia/commons/4/4d/Agouti_Mice.jpg) 3
1.2 Simulated hypothetical phenotype data for two inbred lines, F_1 progeny, and a backcross population (Figure 1.4 from K. W. Broman and Sen, 2009) 4
1.3 Breeding scheme for a backcross 5
1.4 Breeding scheme for an intercross 7
1.5 Genetic map for markers from Sugiyama et al. (2001). Figure from K. W. Broman and Sen (2009). 8
1.6 Inferring genotypes at an intermarker position (from K. W. Broman and Sen (2009)) 9
1.7 One pleiotropic QTL 15
1.8 Two separate QTL 16
1.9 Two-dimensional grid of ordered pairs of markers 17
1.10 Schematic for a single funnel in the Collaborative Cross breeding design 19

2.1 Pleiotropy vs. separate QTL power curves for each of four sets of parameter settings. Factors that differ among the four curves are allele effects difference and allele partitioning. Red denotes high allele effects difference, while black is the low allele effects difference. Solid line denotes the even allele partitioning (ABCD:EFGH), while dashed line denotes the uneven allele partitioning (F:ABCDEGH). 30
2.2 Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins near 56 cM and ends about 64 cM. 32

2.3 Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins near 56 cM and ends about 64 cM. 33

2.4 Profile LOD curves for the pleiotropy vs. separate QTL hypothesis test for "percent time in light" and "hot plate latency". Gray trace denotes pleiotropy LOD values. Triangles denote the univariate LOD maxima, while diamonds denote the profile LOD maxima. For "percent time in light", the brown triangle obscures the smaller brown diamond. Likelihood ratio test statistic value corresponds to the height of the blue and brown traces at their maxima. 34

3.1 Biological information is encoded in DNA. This information is passed, via transcription, to sequence-specific RNA molecules. The process of translation transmits the information to sequence-specific proteins. 39

3.2 A DNA variant in the *Dhtkd1* gene affects *Dhtkd1* transcript abundances which, in turn, affect DHTKD1 protein concentrations. 40

3.3 LOD difference proportion vs. pleiotropy test statistic 46

3.4 LOD difference proportion vs. pleiotropy test statistic. 48

3.5 LOD difference proportion vs. pleiotropy test statistic from the per-local gene perspective. . . 49

3.6 Scatter plots for four nonlocal expression traits. Each plot features 13 points, one for each local gene expression trait. The vertical axis denotes LOD difference proportion values, while the horizontal axis corresponds to pleiotropy test statistics. Blue points represent the pairing with local gene expression trait *Hnf4a*. Red points represent the other 12 local gene expression traits. 50

3.7 Pleiotropy LRT vs. chromosomal position plots reveal that higher values of pleiotropy LRT tend to correspond to greater interlocus distance and greater univariate LOD score. 54

3.8 Pleiotropy LRT vs. univariate LOD score plots reveal that greater univariate LOD scores (and greater interlocus distance) tend to correspond to greater pleiotropy LRT values. 57

3.9 Pleiotropy LRT vs. fitted values correlations plots reveal little evidence for a relationship. . . 58

3.10 Cholic acid chemical structure. 60

3.11 Founder allele effects and LOD plots for our two traits over the Chromosome 8 region. Both traits demonstrate QTL peaks between 4 and 8 Mb and similar patterns of founder allele effects. 63

3.12 Scatter plot of plasma cholic acid levels against *Turicibacter* abundance 65

3.13 Profile LODs for *Turicibacter* abundance and plasma cholic acid levels. 66

C.1 Scatter plot of "hot plate latency" against "percent time in light", after applying logarithm transformations and winsorizing both traits. 103

C.2 Genome-wide QTL scan for percent time in light reveals multiple QTL, including one on Chromosome 8. 104

List of Tables

1.1 Probabilities for genotypes at intermediate putative QTL in a backcross, conditional on flanking marker genotypes. 13

2.1 Type I error rates for all runs in our 2^3 experimental design. We set (marginal) genetic variances (*i.e.*, diagonal elements of V_g) to 1 in all runs. V_e was set to the 2 by 2 identity matrix in all runs. We used allele probabilities at a single genetic marker to simulate traits for all eight sets of parameter inputs. In the column "Allele effects partitioning", "ABCD:EFGH" means that lines A–D carry one QTL allele while lines E–H carry the other allele. "F:ABCDEGH" means the QTL has a private allele in strain F. 29

3.1 Local gene annotations for analysis of Chromosome 2 expression trait hotspot. All positions are in units of Mb on Chromosome 2. LOD peak position and LOD peak height refer to those obtained from univariate analyses. "Start" and "end" refer to the local gene's DNA start and end positions, as annotated by Ensembl version 75. 45

3.2 Annotations for four anchor genes. 53

3.3 Founder allele effect estimates at Chromosome 19 QTL peak position. 55

3.4 Genome-wide LOD peaks greater than 5 for plasma cholic acid levels and *Turicibacter* abundance. Both traits map to approximately the same region on Chromosome 8. Chromosome positions are in Mb units. 62

C.1 Eight founder lines and their one-letter abbreviations. 101

C.2 Both "hot plate latency" and "percent time in light" demonstrate multiple QTL peaks with LOD scores above 5. 102

D.1 LOD peak positions and peak heights for 147 expression traits that map to the Chromosome 2 expression trait hotspot. We see that some transcript levels have LOD scores, per our calculations, below the 7.18 genome-wide threshold. We believe that this is due to differences in statistical modeling between our analyses and those of Keller et al. (2018). 105

D.2 Annotations for 76 non-anchor genes on Chromosome 19. 110

Chapter 1

Introduction

Quantitative trait locus (QTL) studies in model organisms like mice can identify genomic regions (termed "QTL") that affect complex traits, such as systolic blood pressure and body weight (Sax, 1923; Lander and Botstein, 1989; K. W. Broman and Sen, 2009; Jansen, 2007). A genome-wide QTL "scan" reveals associations between genotypes and phenotypes by considering each position, one at a time, as a candidate QTL for the trait of interest. A region with strong evidence of association with a complex trait, then, defines a QTL (for that trait). Because nearby markers have correlated genotypes, a QTL in a two-parent cross often spans multiple megabases in length and may contain more than a hundred genes. Identification of the causal gene (for a given complex trait) from among those genes near the QTL is challenging and may require costly and time-consuming experiments. The growing need for greater QTL mapping resolution fueled development over the last two decades of model organism multiparental populations for high-resolution QTL mapping (Koning and McIntyre, 2017; Churchill, Airey, et al., 2004; Svenson et al., 2012; B. E. Huang et al., 2012; Shivakumar et al., 2018; X. Huang et al., 2011; Kover et al., 2009; Tisne et al., 2017; Stanley et al., 2017). With experimentalists now measuring tens of thousands of biomolecular traits in multiparental populations, the systems genetics community needs multivariate statistical tools to fully examine the large volumes of data (Keller et al., 2018; Chick et al., 2016). A test of pleiotropy vs. separate QTL is one multivariate statistical tool that will inform complex trait genetics by enabling researchers to identify the number of unique QTL in a genomic region of interest. In this thesis, I develop a test of pleiotropy vs. separate QTL in multiparental populations. I study its statistical properties and demonstrate its utility by analyzing experimental data.

In the first chapter, I introduce statistical methods in QTL mapping studies and argue for the need to develop a pleiotropy test in multiparental populations. Existing pleiotropy tests for two-parent crosses

don't directly apply to multiparental populations because of the multi-allelic genotype data and complex relatedness patterns. A pleiotropy test for multiparental populations with high-dimensional traits would enable dissection of QTL hotspots (*i.e.*, small genomic regions that affect many traits) and would complement newly developed methods in causal inference for systems genetics.

I begin Chapter 1 by motivating the study of pleiotropy with a case study of the pleiotropic *Agouti* gene. I then consider two-parent crosses, with emphases on experimental design, QTL mapping methods, and pleiotropy testing. I then turn to multiparental populations, where I describe experimental design and QTL mapping methods. I conclude Chapter 1 by reiterating the need to develop a test of pleiotropy vs. separate QTL in multiparental populations.

1.1 *Agouti* biology & pleiotropy

Pleiotropy is the biological phenomenon in which a QTL affects multiple traits. Identifying a gene that affects multiple traits may inform scientific understanding of interactions between biomolecules and ultimately contribute insights that aid development of new therapeutics. For example, mouse studies identified multiple biological roles for the protein product of the *Agouti* gene. Mutations in the *Agouti* gene may lead to both yellow hair (in mice that are typically black) and obesity (Attie, Churchill, and Nadeau, 2017). Subsequent investigations uncovered two related biological roles for the Agouti protein. It antagonizes the action of α-melanocyte-stimulating hormone both to prevent melanocyte-based melanin production and to disrupt melanocortin-4 receptor signaling in the brain (D. Lu et al., 1994; Klebig et al., 1995; Huszar et al., 1997). The former leads to yellow hair, while the latter causes weight gain. Later research identified altered signaling by the melanocortin-4 receptor in the brain as a leading cause of inherited obesity in humans (Farooqi et al., 2003; Vaisse et al., 2000). Therapeutics to mitigate the effects of disruptions in melanocortin receptor signaling are currently being developed (MacNeil et al., 2002; Fani et al., 2014). The case of pleiotropic *Agouti* variants illustrates the potential value of examining pleiotropy in QTL studies. We next consider designs and QTL mapping in two-parent crosses (Sections 1.2 and 1.3).

1.2 Designs for two-parent crosses

Two widely used two-parent crosses are the "backcross" and the "intercross". We first discuss the backcross design (Figure 1.3). A backcross starts with mating between members of two inbred lines. The offspring

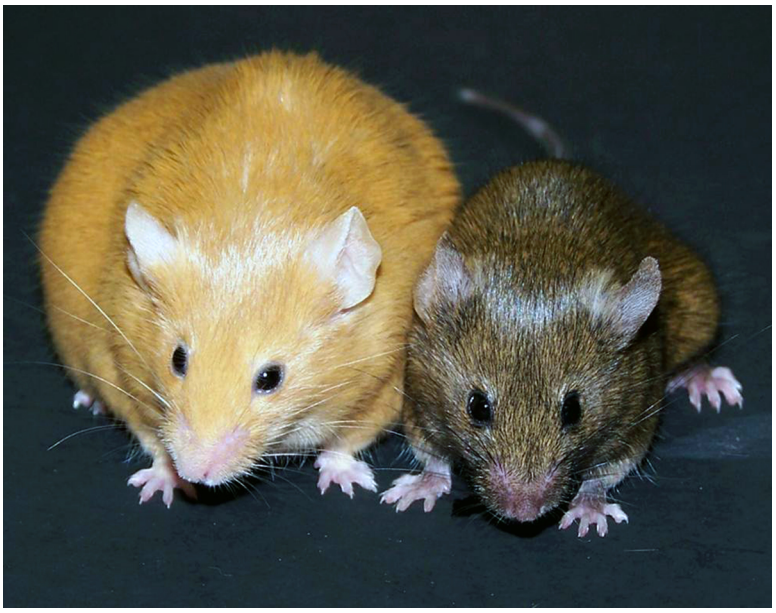


Figure 1.1: *Agouti* viable-yellow mutant mouse has yellow hair and is obese in comparison to a wild-type mouse. (https://upload.wikimedia.org/wikipedia/commons/4/4d/Agouti_Mice.jpg)

of this mating event, termed the "F_1", or "first filial", generation then mate with members of one of the two parent lines. The experiment designers decide which line (A or B) mates with F_1 subjects with. The F_1 subjects mate with the specified founder line (A in Figure 1.3) to create generation "BC" subjects. Generation BC subjects are then genotyped and phenotyped for QTL mapping.

As Figure 1.3 illustrates, all backcross animals (generation BC) have, for any locus, only two possible genotypes: heterozygote (AB) and homozygote (AA). The F_1 generation is heterozygous at all loci (Figure 1.3) because they inherited one chromosome (of each pair) from each parent. While meiotic crossovers may occur in the founder generation, these crossover events are undetectable because the animals are inbred and, thus, have two identical copies of DNA at every locus. This causes the F_1 subjects to be heterozygous at all positions when genotyping with respect to the founder lines. In fact, the F_1 may have identical single nucleotide marker alleles at some positions, but, because one allele is inherited from each parent (and each inbred founder line), we still call them "heterozygous" because, in terms of founder allele genotypes, they have the "AB" heterozygote genotype.

The meiotic crossovers that occur as the F_1 animals produce gametes often are detectable because they result in "BC" generation animals that have individual chromosomes that contain DNA from both founder lines. Due to the limited number of crossovers in a single meiosis, the resulting BC generation subjects have poor QTL mapping resolution. This poor mapping resolution corresponds to the relatively large segments

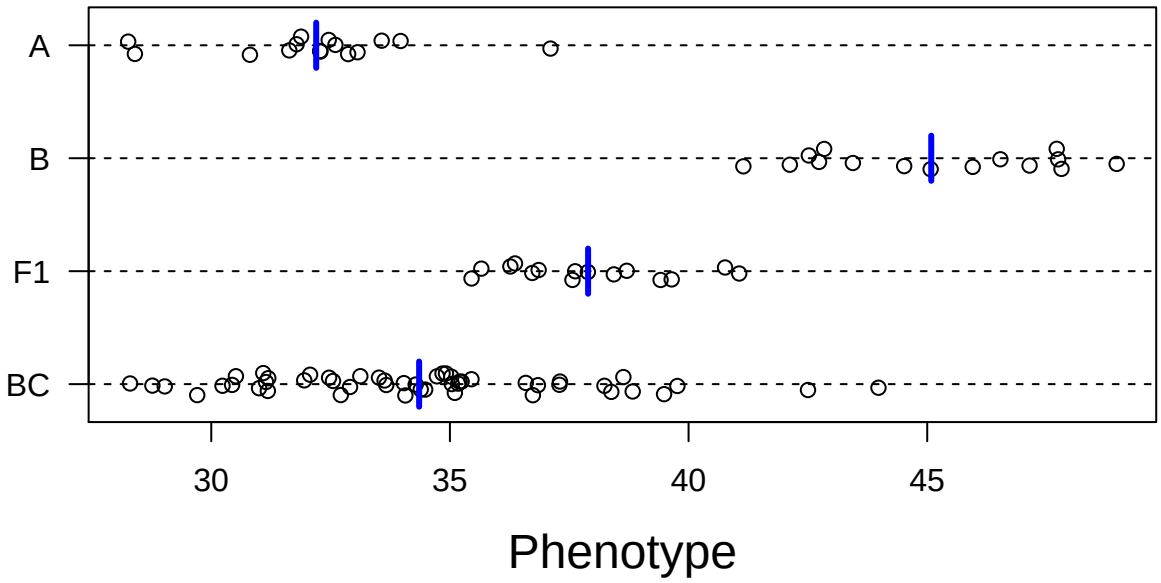


Figure 1.2: Simulated hypothetical phenotype data for two inbred lines, F_1 progeny, and a backcross population (Figure 1.4 from K. W. Broman and Sen, 2009)

of contiguous DNA, on average, that are inherited from, ultimately, the founder lines.

Figure 1.2 illustrates hypothetical values for a single phenotype in two founder lines (A and B), F_1 progeny, and a backcross population. Mean phenotype values for each sample are indicated by blue vertical lines. Each circle corresponds to a single subject. In practice, one might not phenotype F_1 subjects because the QTL mapping is performed in the BC subjects. However, one often measures the most important phenotypes in founder line subjects before the matings. This permits the experimenter to ensure that the two lines differ sufficiently in phenotype values. Notice that the F_1 mean phenotype is intermediate between the two founder line means. The BC mean phenotype is between the F_1 mean and the AA founder line mean. It's also valuable to consider the range in the phenotype values in each of the four samples. The BC subjects' values are spread out much more than those of the three other samples. This corresponds to the observation that BC subjects have either one or two copies of the A allele at each marker (Figure 1.3). The other three samples have an unambiguous number of A alleles at each marker: two for AA founder line, zero for BB founder line, and one for the F_1 subjects. Now that we've discussed two-parent designs, we turn attention to QTL mapping two-parent crosses.

I now consider the intercross design (Figure 1.4). An intercross differs from a backcross design by mating F_1 subjects with each other, rather than with a parent line (as in a backcross). Because F_1 subjects produce gametes with detectable crossovers, the F_2 generation animals have, for every marker, three genotype classes:

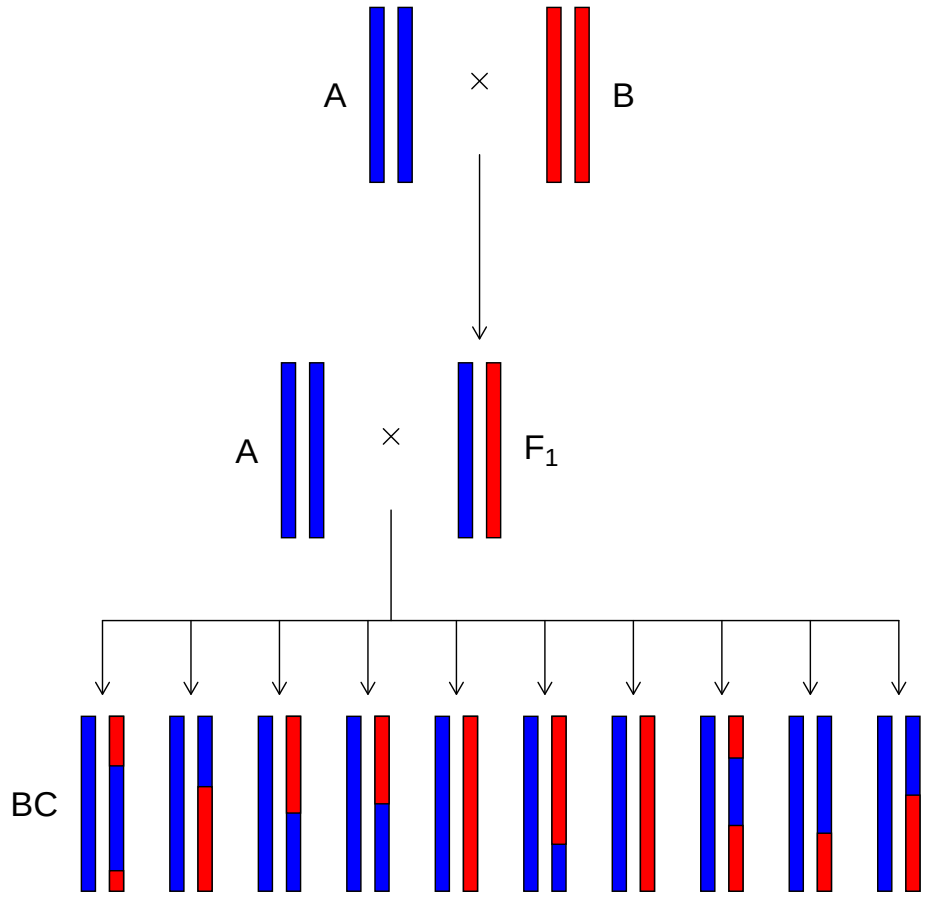


Figure 1.3: Breeding scheme for a backcross. Each pair of autosomes represents a single subject. (From K. W. Broman and Sen (2009))

AA, AB, and BB. Compare this with the backcross, where BC subjects have only two genotype classes (AA and AB in Figure 1.3). I'll discuss the consequences of this distinction on QTL mapping methods below (Section 1.3).

In an intercross, phenotyping and genotyping are typically done in only the F_2 subjects. As with a backcross, intercross planning may involve genotyping and phenotyping of subjects from two (or more) inbred lines. The designers then have the option of choosing as founders two inbred lines that demonstrate a big difference in mean phenotype values (for traits of interest).

A plot of simulated phenotype values, analogous to that of Figure 1.2, for F_2 subjects would tend to reveal an even greater range of measurements than is present in the BC generation. This is due to the three genotype classes (AA, AB, and BB) per marker that are present in samples of F_2 subjects. The extent to which the trait value range is greater in F_2 subjects compared to BC subjects also depends on the true genetic model. If the trait is not a linear function of the number of A alleles at a single marker, then the F_2 trait range may not exceed that in BC subjects. For example, in the extreme case where the B allele dominates the A allele, so that AB and BB genotype classes have the same mean trait value, the range in a sample of F_2 subjects may be similar to that of a sample of BC subjects (where the backcross involves mating F_1 subjects with AA subjects). Now that I've considered two widely used two-parent designs, I discuss QTL mapping methods in both the backcross and intercross.

1.3 QTL mapping in two-parent crosses

1.3.1 Overview

QTL mapping is a systematic, statistical approach for identifying genetic loci where genetic variation affects variation in a measured trait. A variety of statistical methods for QTL mapping exist. I discuss below three QTL mapping methods: marker regression, interval mapping, and Haley-Knott regression. I consider the three methods in analysis of both backcross and intercross designs. Before getting to the details of our three QTL mapping methods, I discuss data inputs that are required for QTL mapping.

Three standard inputs for QTL mapping are:

1. genome-wide marker genotypes for a collection of study subjects
2. a set of trait measurements on the same subjects
3. a genetic map that contains genomic positions for all molecular markers (Figure 1.5)

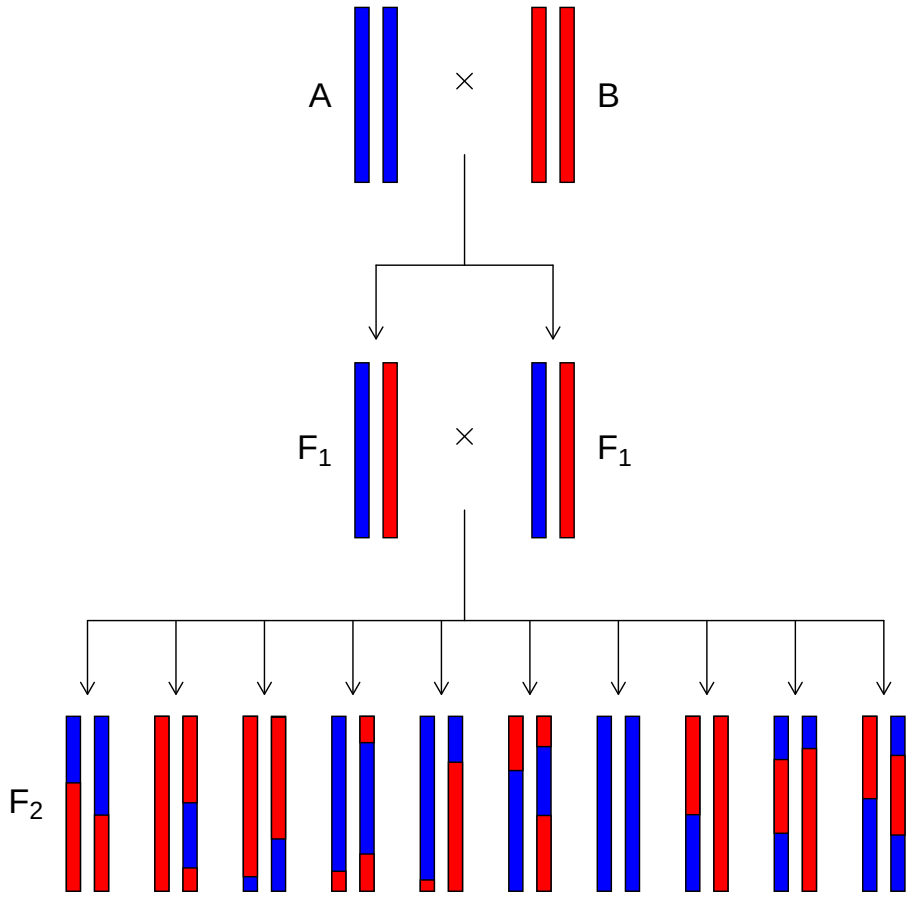


Figure 1.4: Breeding scheme for an intercross. Each pair of autosomes represents a single subject. (From K. W. Broman and Sen (2009))

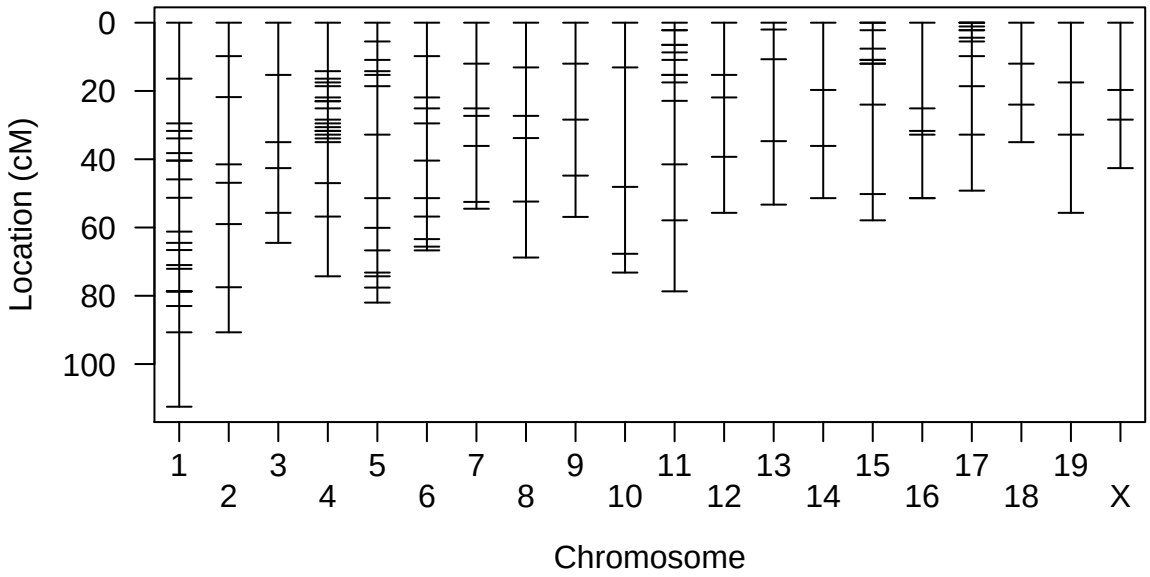


Figure 1.5: Genetic map for markers from Sugiyama et al. (2001). Figure from K. W. Broman and Sen (2009).

A physical map, with marker positions in units of megabases, may also be used. Additional inputs include genotype data (for the same set of markers) from founder lines, *i.e.*, the inbred lines that initiate the mating design. For example, in Figures 1.3 and 1.4 there are two founder lines at the top of the mating design: 1. those with genotypes AA and 2. those with genotypes BB.

Figure 1.5 depicts the positions of markers (horizontal hash marks) on each chromosome (vertical lines). In practice, for QTL mapping, a genetic map object may be stored as a collection of two-column data frames, where each row is a unique marker and the two columns contain marker identities and marker positions. In this structure, each chromosome has its own map object. Marker distribution is not uniform across the genome (Figure 1.5). Some chromosome regions, including parts of Chromosome 4, have dense marker coverage, while others, including Chromosome 2, have relatively sparse marker coverage. Dense marker coverage is preferred in QTL mapping applications because it provides more genetic information and enables more accurate genotype inference between markers. Before the recent development of high-density single nucleotide polymorphism marker arrays, many studies relied on sparse microsatellite and restriction fragment length polymorphism markers.

A univariate QTL scan is a procedure to interrogate the entire genome for genetic variants that affect a single trait of interest. One specifies a statistical model and calculates the likelihood of the model parameters given the observed (trait and marker) data at each genomic position. After obtaining likelihoods, likelihood

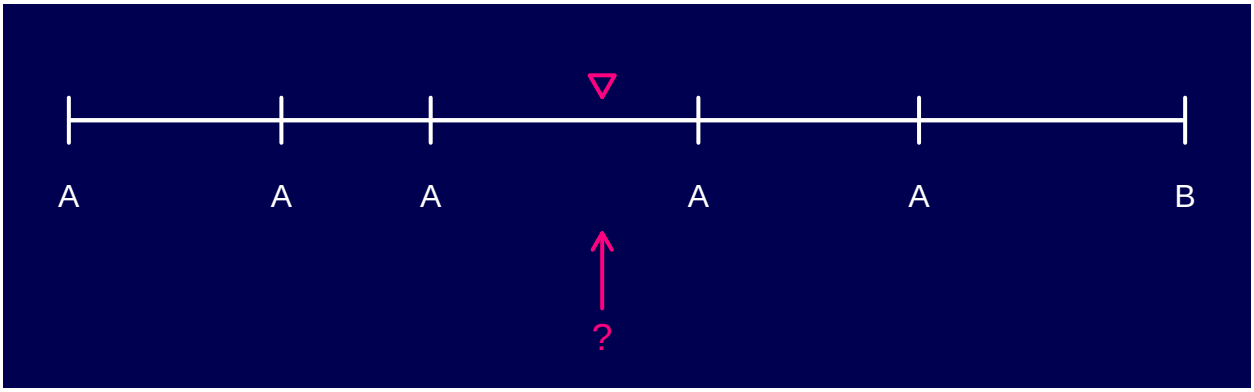


Figure 1.6: Inferring genotypes at an intermarker position (from K. W. Broman and Sen (2009))

ratio test statistics are calculated at every marker. The inputs for these calculations are the likelihoods from the model fits and the likelihood for the null model, which contains no genotype data. The resulting likelihood ratios compare, for every marker, the null hypothesis that there is no QTL (at that marker) against the alternative that there is a QTL (at the specified marker). In other words, one performs a likelihood ratio test for every marker across the genome. In systems genetics studies, the analyst typically visualizes likelihood ratio test statistics as $\log_{10}$ likelihood ratio scores, or "log odds (LOD)" scores. In modern studies, this amounts to thousands of (statistically dependent) hypothesis tests. One then uses a permutation test to determine a genome-wide critical value for the likelihood ratio test statistic (Churchill and Doerge, 1994). Those loci for which the likelihood ratio test statistic is sufficiently large are declared QTL.

In QTL mapping, two major statistical challenges are what K. W. Broman and Sen (2009) call the "missing data" problem and the "model selection" problem. The "missing data" problem arises in QTL studies because genotypes are obtained at only select markers. In this sense, genotypes at positions between markers are "missing" because they aren't explicitly measured (Figure 1.6). Figure 1.6 presents, for one subject, a chromosome with genotyped markers at the tick marks. The triangle represents a position at which we wish to know the allele's identity. Below, we discuss methods for genotype inference at positions between markers.

The previously mentioned statistical methods for QTL mapping, marker regression, interval mapping, and Haley-Knott regression, differ in strategies for addressing the "missing data" problem. I first discuss marker regression, because it is the simplest approach and because the other strategies share some aspects of it.

1.3.2 Marker regression

Marker regression inputs are genome-wide marker genotypes and measured values of a complex trait of interest (from the same subjects). At each marker, one fits a linear model by regressing trait values on marker genotypes (Equation 1.1), sequentially, at each marker. In a backcross, the "B allele count" variable takes values of 0 or 1. In an intercross , the "B allele count" may be 0, 1, or 2. Both backcross and intercross designs permit solution of the regression by a statistical technique called "maximum likelihood estimation". I assume that the random errors for all subjects have the same variance and are statistically independent. Statistical independence means that knowing the value of one subject's random error provides no information about the value of another subject's random error.

$$\text{trait} = (\text{mean for AA genotype}) + (\text{B allele count})(\text{B allele effect}) + (\text{random error}) \tag{1.1}$$

$$(\text{random error}) \sim N(0, \sigma^2), \qquad (\text{random errors}) \text{ independent} \tag{1.2}$$

Equation 1.2 introduces notation that requires explanation. The "$\sim$" symbol designates a probability distribution. The left-hand side "is distributed" the right-hand side. In this case, I describe the distribution of the random errors. Capital "N", in a distribution, abbreviates "Normal". The first number after the "N" is the distribution mean, while the second is the variance. These two numbers, mean and variance, completely specify a normal distribution.

The linear regression equation (Equation 1.1) has three statistical parameters, "mean for AA genotype", "B allele effect", and the variance of the random errors. Notice that the variance of the random errors is not explicitly written in Equation 1.1, while "mean for AA genotype" and "B allele effect" are obviously part of the equation. In this statistical framework, I treat parameters as fixed, but unknown, numbers. I then seek the parameter values that are most compatible with the data, under the assumption that the statistical model is correctly specified. I use the approach called "maximum likelihood estimation". Some details of maximum likelihood estimation are in Appendix B.

To say that a statistical model is correctly specified means, here, that the family, or functional form, of the model is correct. If I say that the data follow a normal distribution, then they do, in fact, follow a normal distribution. I generally will not know the true parameter values, but, in the case of the random errors in Equation 1.2, the model is correctly specified if the random errors really do follow a normal distribution with

some positive variance σ^2.

The maximum likelihood estimates for two parameters, "B allele effect" and the random error variance, are given in Equations 1.3 and 1.4. Note that I've replaced "B allele effect" with the Greek letter β, while σ^2 denotes the random error variance. The "hat" above β indicates that $\hat{\beta}$ is a maximum likelihood estimate of the parameter β (*i.e.*, the unknown "B allele effect").

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{1.3}$$

$$\widehat{\sigma}^2 = \frac{RSS_1}{n} \tag{1.4}$$

The residual sum of squares "RSS_1", which is central to the calculation of the pooled variance estimate $\widehat{\sigma}^2$ in Equation 1.4, is an expression that tells us how far the observed trait measurements are from the predicted values, for a set of specified parameter values. Equation 1.5 defines residual sum of squares for a univariate QTL analysis. I make an additional substitution in Equation 1.5 by replacing "B allele count" with x. In Equation 1.5, I replace "mean for AA genotype" with μ_{AA}. The Greek letter μ is widely used in statistics to designate a mean, while the subscript "AA" tells us the genotype class.

$$RSS_1 = \sum_{i=1}^{n} (y_i - \mu_{AA} - x_i\beta)^2 \tag{1.5}$$

We ultimately want to calculate a likelihood ratio test statistic for the competing statistical hypotheses of Equation 1.6

$$\begin{aligned} H_1 &: \text{A QTL is present at a specified marker} \\ H_0 &: \text{No QTL is present at a specified marker} \end{aligned} \tag{1.6}$$

I wish to calculate the likelihood for both the alternative hypothesis and the null hypothesis. I write the likelihood as a product of per-subject likelihoods in Equation 1.7.

$$L_1(\mu_{AA}, \beta, \sigma^2) = P(\text{data}|\text{QTL at marker}, \mu_{AA}, \beta, \sigma^2) = \prod_{i=1}^{n} \phi(y_i; \mu_{g_i}, \sigma^2) \tag{1.7}$$

I introduce new notation in Equation 1.7. L_1 denotes the likelihood under hypothesis H_1. The likelihood

of the model parameters is equal, by definition, to the joint probability of the data conditional on presence of a QTL and the parameter values. This, by the assumption of statistical independence of subjects, is equal to the product of the n per-subject likelihoods, where $\phi(y; \mu, \sigma^2)$ denotes the density at y for a normal distribution with mean μ and variance σ^2 (Equation 1.8)

$$\phi(y; \mu, \sigma^2 = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{1.8}$$

I also need to calculate the maximum likelihood under the null model constraints, *i.e.*, under the assumption that there is no QTL at the marker, in which case the genotype classes have the same mean trait value.

$$L_0(\mu_0, \sigma_0^2) = P(\text{data}|\mu_0, \sigma_0^2) = \prod_{i=1}^{n} \phi_i(y_i; \mu_0, \sigma_0^2) \tag{1.9}$$

After calculating the likelihood ratio test statistic at every marker, one needs to obtain p-values for every test. Because a researcher performs many tests (*i.e.*, one per marker) in a genome-wide QTL scan, one should perform a permutation test to determine adjusted p-values. In this permutation test, one shuffles or randomizes subjects' trait values to get a "permuted phenotype". One does this many times to obtain, say, 1000 permuted phenotypes. Each permuted phenotype is then analyzed with a univariate QTL scan. The statistician records, for each permuted phenotype, the maximum observed LOD score from the genome-wide scan to obtain an empirical distribution of "max LOD score" statistics. One then determines the 95^{th} percentile of the empirical distribution and uses it as a critical value when assessing statistical significance of observed peaks.

1.3.3 Interval mapping

I highlight elements from the exposition in K. W. Broman and Sen (2009). More details are available in its Chapter 4. Lander and Botstein (1989) describe interval mapping for QTL studies. While marker regression considers only marker positions as candidate QTL, in interval mapping one probabilistically infers genotypes for all genomic positions on a grid along each chromosome. One leverages the entire collection of marker genotypes on each chromosome and their correlations to do this.

To illustrate this approach, consider a backcross sample with two adjacent markers. I suppose that the backcross follows the design in Figure 1.3, with generation BC subjects having either AA or AB genotypes. I wish to know the genotype probabilities at an intermediate point, *i.e.*, the putative QTL, between the

Table 1.1: Probabilities for genotypes at intermediate putative QTL in a backcross, conditional on flanking marker genotypes.

Marker 1 genotype	Marker 2 genotype	P(AA)	P(AB)
AA	AA	$\frac{(1-r_{1Q})(1-r_{Q2})}{1-r_{12}}$	$\frac{r_{1Q}r_{Q2}}{(1-r_{12})}$
AA	AB	$\frac{(1-r_{1Q})r_{Q2}}{r_{12}}$	$\frac{r_{1Q}(1-r_{Q2})}{r_{12}}$
AB	AA	$\frac{r_{1Q}(1-r_{Q2})}{r_{12}}$	$\frac{(1-r_{1Q})r_{Q2}}{r_{12}}$
AB	AB	$\frac{r_{1Q}r_{Q2}}{(1-r_{12})}$	$\frac{(1-r_{1Q})(1-r_{Q2})}{(1-r_{12})}$

two markers. Calling the markers "marker 1" and "marker 2", suppose that the recombination fraction between marker 1 and marker 2 is r_{12}. Assume that all subjects are genotyped at markers 1 and 2. With two additional assumptions, that of no crossover interference and that of no genotyping errors, I can write expressions for genotype probabilites at the putative QTL, conditional on the genotypes at markers 1 and 2 (Table 1.1).

In Table 1.1, I present genotype probabilities for a putative QTL, denoted "Q", between marker 1 and marker 2, in a backcross (Figure 1.3). r_{12} is the recombination probability between marker 1 and marker 2, r_{1Q} is the recombination probability between marker 1 and putative QTL Q, and r_{Q2} is the recombination probability between putative QTL Q and marker 2.

1.3.4 Haley-Knott regression

An innovative and flexible framework for resolving the "missing data" problem involves use of a statistical tool called a hidden Markov model (K. W. Broman and Sen, 2009; K. W. Broman, 2006). In this hidden Markov model, marker genotypes are "observed random variables", while genotypes at intervening bases are "unobserved random variables". Two sets of recursive equations, termed "forward" and "backward" equations, enable efficient calculation of genotype probabilities (conditional on the observed marker genotypes) (Baum et al., 1970). We discuss inference methods for genotype probabilities in Appendix B.

A QTL analysis involves choosing a statistical model to be fitted at every marker. The question of which statistical model to use is the "model selection" problem (K. W. Broman and Sen, 2009). As in any statistical modeling, there is not one best model selection procedure. One needs to decide which main effects and interactions are needed. Also, one often transforms the trait values to achieve approximate normality before performing the QTL scan.

In a two-parent backcross, only two genotypes are possible in generation BC subjects (Figure 1.3). Thus, one may choose a linear model that contains the B allele count (1 or 0 for each subject) at each marker

(Equation 1.1, Figure 1.3).

Assuming that the random errors are normally distributed (and independent with common variance σ^2) for all subjects, we write the likelihood as a function of the parameters conditional on the observed data (Equation 1.10). We provide a primer on likelihood methods in Appendix.

add appendix letter here

$$L(\text{AA mean}, \text{B allele effect}, \text{random error variance}|\text{genotypes}, \text{phenotypes}) = \prod_{i=1}^{n} \phi(y_i; \mu_{AA}, \beta, \sigma^2) \quad (1.10)$$

The symbol $\prod$ denotes a product of indexed terms. The index range for the terms are given above and below the $\prod$ symbol. Here, n is the sample size. In Equation 1.10, the three parameters are "AA mean trait value" (μ_{AA}), "B allele effect" (β), and the random error variance, which we denote σ^2. In other words, I am specifying the mean phenotype values for the three genotype classes. Genotype class AA has mean phenotype value of μ_{AA}, genotype class AB has mean phenotype value of $\mu_{AA} + \beta$ and genotype class BB has mean phenotype value of $\mu_{AA} + 2\beta$. I assume that the error variance, σ^2 is the same for all three genotype classes.

We use maximum likelihood methods to estimate parameter values. The maximum likelihood estimates for three parameters are the sample mean of trait values for those with AA genotype, the difference in sample mean between those with AA genotype and those with AB genotype, and the pooled variance estimate, $\widehat{\sigma}^2$ (Equation 1.4)

1.4 Testing pleiotropy in two-parent crosses

In anticipation that multivariate mapping of correlated traits would enhance statistical power to detect QTL and would improve precision of QTL position estimates, both Jiang and Zeng (1995) and Korol, Ronin, and Kirzhner (1995) developed multivariate interval mapping procedures for two-parent crosses. Among the novel methods from Jiang and Zeng (1995) is a test of pleiotropy vs. separate QTL. They explain that such a test is useful when two traits map to a single genomic region. The question then arises "do the two traits associate with the same locus, or do they associate with separate loci"?

If both traits associate with the same locus, then that locus is called "pleiotropic". If the two traits associate with separate loci, then I say that there are separate QTL (one for each trait). Figures 1.7 and 1.8

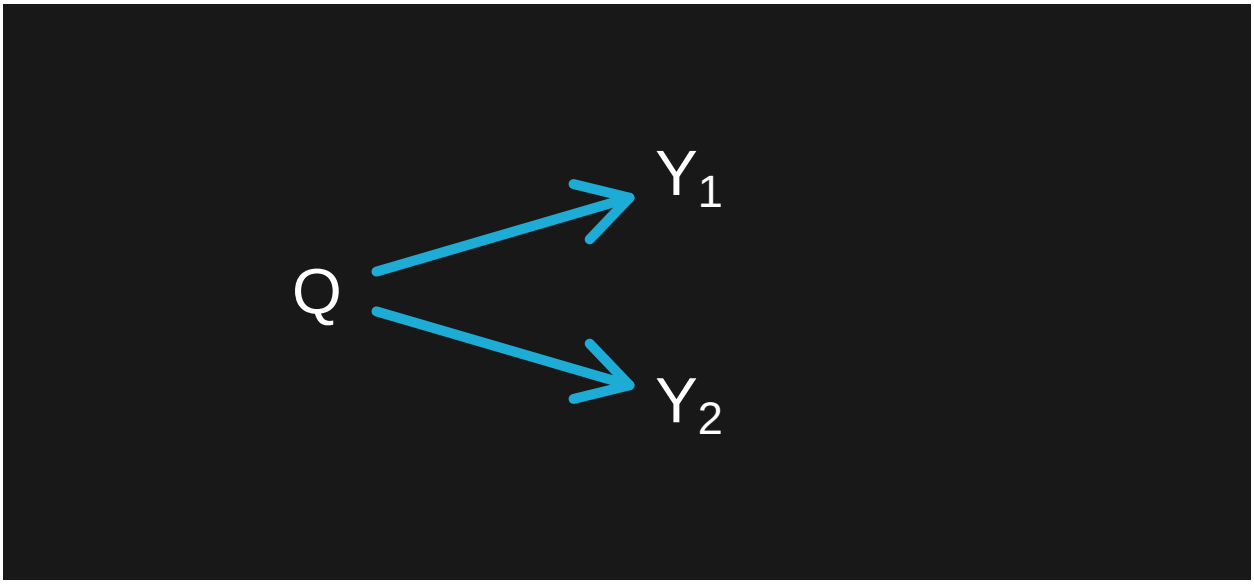


Figure 1.7: One pleiotropic QTL. Figure after code from https://github.com/kbroman/QTLFigs repository.

provide schematics for the two possibilities. In Figure 1.7, a single QTL (denoted by Q) affects both traits. In Figure 1.8 two distinct QTL (Q1 and Q2) affect one trait each. The pink arrows indicate that Q1 and Q2 are on the same chromosome and are correlated by linkage disequilibrium.

In the development of their test of pleiotropy v separate QTL, Jiang and Zeng (1995) model the multivariate trait as the sum of a linear function of the genotypes and a random error term. They assume that the traits matrix is related to the genotype data through Equation 1.11.

$$vec(Y) = Xvec(B) + vec(E) \tag{1.11}$$

In Equation 1.11, Y is a n by 2 matrix of trait values (with each row being one subject and each column being one trait), X is a $2n$ by 4 block-diagonal matrix containing genotype data for two positions, and B is a 4 by 1 matrix of allele effects. The random error is assumed to follow a normal distribution with mean zero and a positive variance. Jiang and Zeng (1995) also incorporate additional terms to model dominant effects and to control for residual genetic variation. To simplify model specification, we forego inclusion of dominant terms and additional markers for controlling residual genetic variation. Extensions of our model and software to accommodate dominance effects are straightforward.

Jiang and Zeng (1995) use a statistical "mixture model" to relate genotypes to phenotypes. They observe that the there are 9 possible ordered pairs of genotypes for two markers in a F_2 population, since each marker

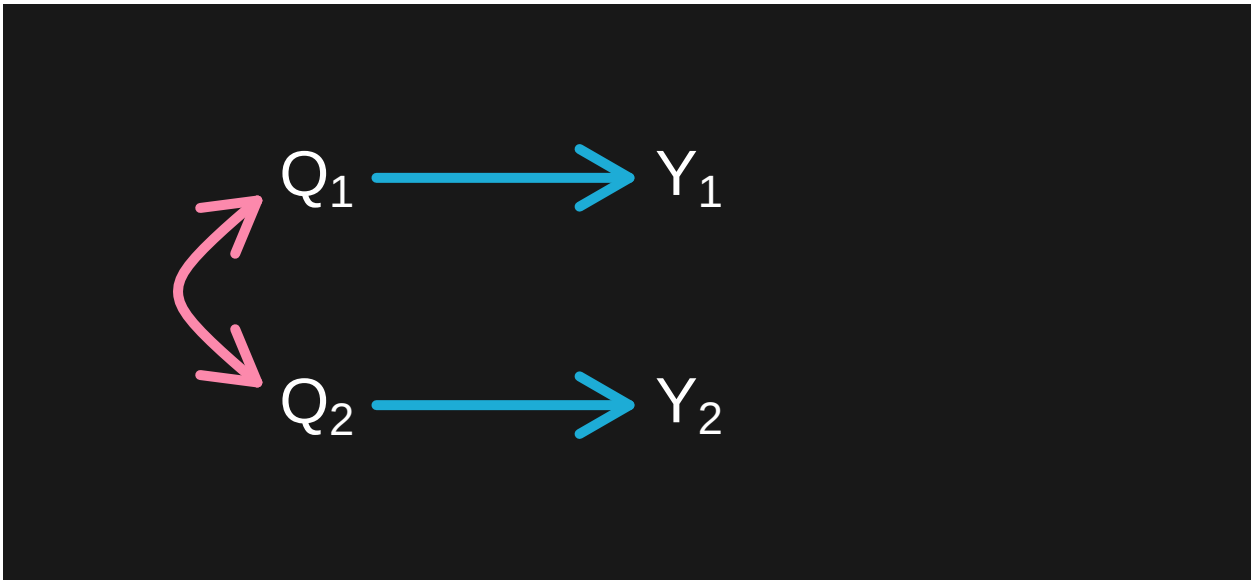


Figure 1.8: Two separate QTL. Figure after code from https://github.com/kbroman/QTLFigs repository.

has one of three genotypes (AA, AB, BB). They thus approach the problem as one with a 9-component mixture model, with one component for each ordered pair of genotypes. They provide the equations needed for an expectation-conditional maximization algorithm, a variant of the expectation-maximization algorithm, to find a local maximum in the likelihood surface (Dempster, Laird, and Rubin, 1977; Meng and Rubin, 1993). They use a chi-square distribution with one degree of freedom as the null distribution of the test statistic.

In statistical terminology, one would say that the parameter space is restricted under the null hypothesis. Here, the parameter space is the collection of ordered pairs of markers in the genomic region of interest. The restriction of that collection under pleiotropy corresponds to limiting consideration to only those ordered pairs that have both traits mapping to a single locus.

We illustrate these ideas in a drawing of a two-dimensional grid (Figure 1.9). Each point in the grid corresponds to an ordered pair of markers. Along the horizontal axis are the markers for the first component of the ordered pair; the second component of the ordered pair is indicated by the vertical axis. The red points correspond to those that are considered under the pleiotropy hypothesis. Under the alternative hypothesis, all grid points are considered.

Jiang and Zeng (1995) then calculate the likelihoods of the models corresponding to all grid points. The likelihood ratio test statistic (for the pleiotropy test) is the ratio of the maximum of the likelihoods under pleiotropy to the maximum of the likelihoods under the separate QTL hypothesis. Jiang and Zeng

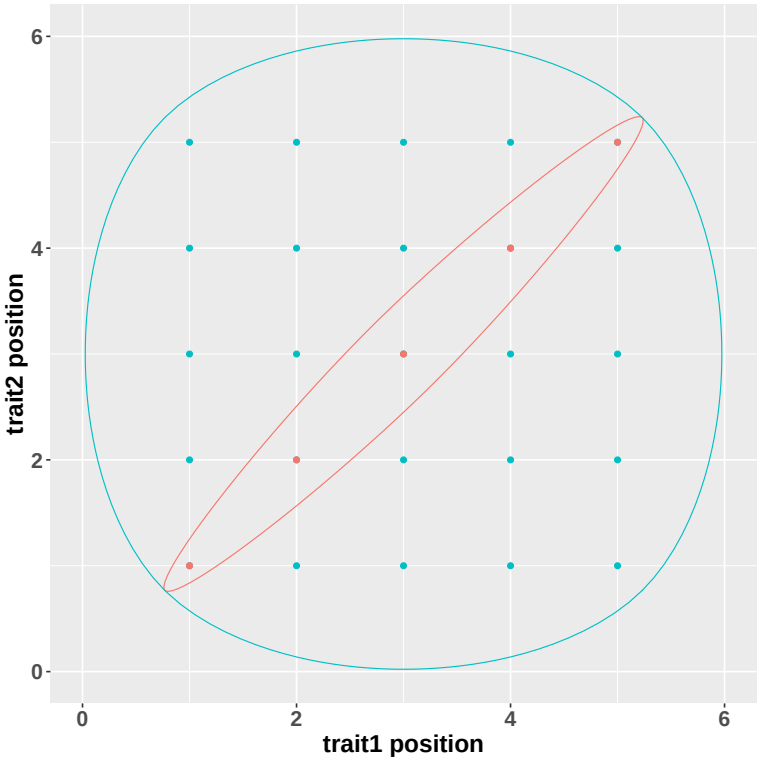


Figure 1.9: Two-dimensional grid of ordered pairs of markers.

(1995) determined p-values for their test statistics by comparing them, after logarithm transformation and multiplication by two, to a chi-squared distribution with 1 degree of freedom.

1.5 Designs for multiparental populations and Diversity Outbred mice

Near the turn of the century, geneticists sought a mammalian gene mapping resource that could be used for study of a wide variety of quantitative traits. The magnitude of such an undertaking required a collaborative, community-supported approach (Koning and McIntyre, 2014). Scientists conceived of the Diversity Outbred mouse population as such a high-resolution gene mapping resource (Churchill, Daniel M Gatti, et al., 2012; Svenson et al., 2012). They elected to seed the population with partially inbred progenitors of the Collaborative Cross lines (Churchill, Airey, et al., 2004). The Collaborative Cross is a collection of dozens of inbred lines that result from shuffling of the eight founder line genomes and multiple generations of inbreeding. The Collaborative Cross mating design started with mice from A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ, WSB/EiJ. Three of these lines, CAST/EiJ, PWK/PhJ, and WSB/EiJ, are wild-derived. Together, the three wild-derived lines contribute a high proportion of the genetic variants in the Diversity Outbred mice.

The designers of the Collaborative Cross used a "multi-funnel" mating scheme to generate mice with DNA from all eight founder lines over the course of 3 generations (Churchill, Airey, et al., 2004). The term "funnel" refers to the design for the first 3 mating generations in which the DNA from eight founder lines "funnels" into animals that have DNA from all eight founders. For example, in one funnel, mating pairs are: A x B, C x D, E x F, and G x H in the first generation (Figure 1.10) (where "x" denotes mating). AB offspring then mate with CD offspring and EF offspring mate with GH mice. Finally, the ABCD mice mate with the EFGH mice to create a generation of mice that contain genetic material from all eight inbred founder lines. Subsequent generations of inbreeding resulted in multiple inbred lines for the Collaborative Cross.

The designers of the Diversity Outbred population started with 167 breeding pairs from Collaborative Cross generations F4 to F12 (Svenson et al., 2012). They since have maintained the Diversity Outbred mouse population with about 175 mating pairs. They produce three or four generations per year. Offspring from each mating are either used to create the next generation or shared with researchers around the world (Svenson et al., 2012; Chesler et al., 2016).

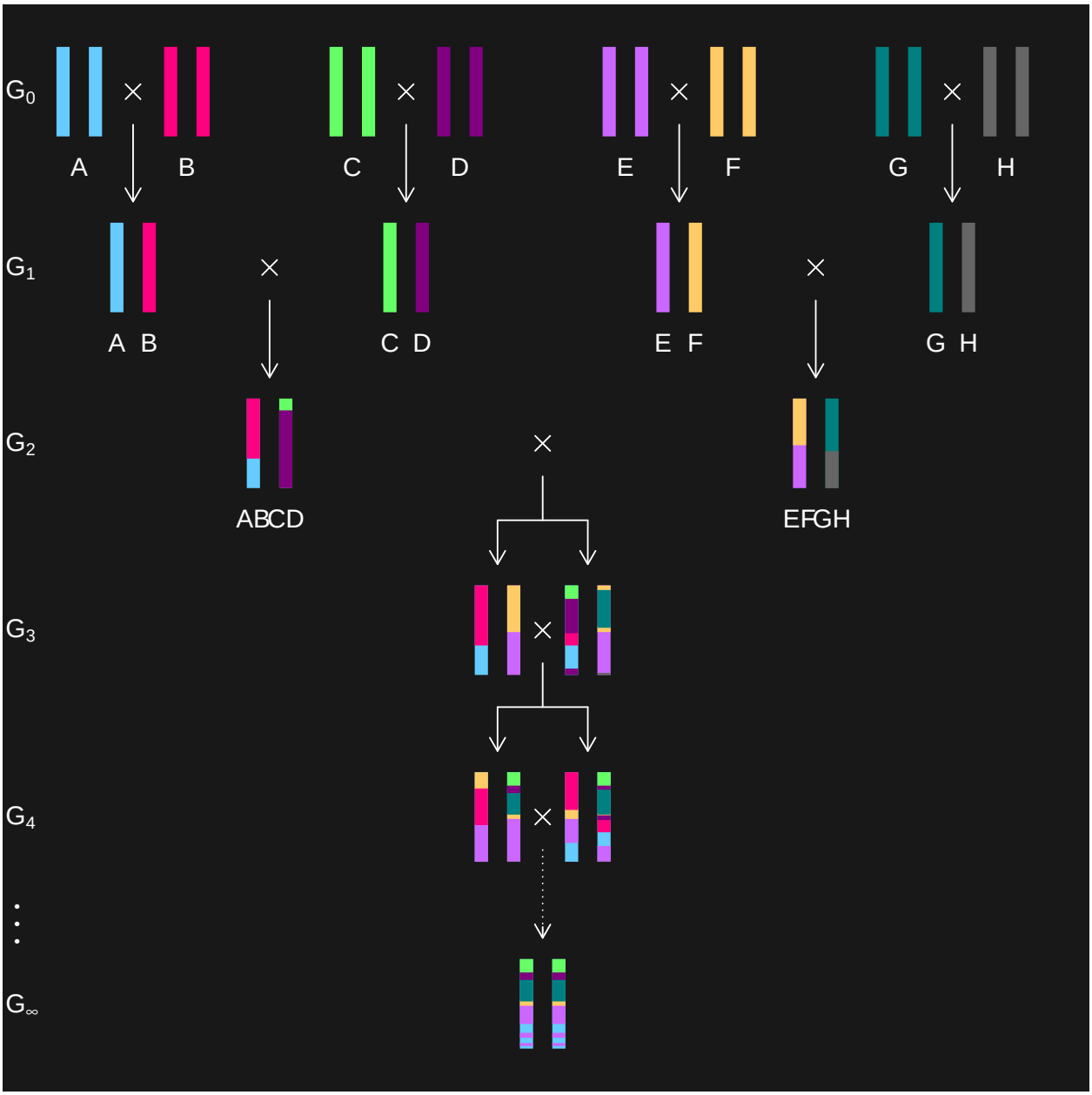


Figure 1.10: Schematic for a single funnel in the Collaborative Cross breeding design.

Due to the breeding scheme, each Diversity Outbred mouse is a highly heterozygous and essentially unique mosaic of founders' DNA With each new generation, the Diversity Outbred mice accumulate meiotic recombinations. This is because each mouse inherits DNA from its parents, and each meiosis provides opportunities for recombinations.

add figure for one mouse's DNA?? like figure 5A in Svenson et al. 2012

need to explain this better. How does Andrew explain it in his thesis?

The Diversity Outbred mice enable high-resolution QTL mapping (Daniel M Gatti et al., 2014; K. W. Broman, D. M. Gatti, et al., 2019). The cumulative effect of meiotic recombinations over many generations of outbreeding is to create mice whose DNA mosaics contain smaller and smaller contiguous pieces from each founder line. QTL mapping resolution improves as the total number of (detectable) meiotic recombinations increases. Thus, Diversity Outbred mice from later generations have greater mapping resolution.

1.6 QTL mapping in Diversity Outbred mice

QTL mapping in Diversity Outbred mice, like that in mice from two-parent crosses, is a multi-step procedure: 1. data acquisition, 2. inference of missing genotypes, and 3. modeling phenotypes as a function of genotypes. Data acquisition involves measurement of phenotypes and, at specified genetic markers, termed "single nucleotide polymorphism" or SNP markers, measurement of two-allele genotypes. Often, the SNP marker genotypes are obtained by use of a microarray, such as the GigaMUGA SNP microarray (Morgan et al., 2015).

The next step, missing genotypes inference, is needed because of the "missing data problem" (K. W. Broman and Sen, 2009). It takes as input the two-allele genotypes at the measured SNP markers. An expectation-maximization algorithm (Dempster, Laird, and Rubin, 1977) for a hidden Markov model, developed by K. W. Broman (2012b) and K. W. Broman (2012a) and implemented in the `qtl2` R package (K. W. Broman, D. M. Gatti, et al., 2019), outputs 36-state genotype probabilities for all nuclear autosomal markers and pseudomarkers. Pseudomarkers, as we use the term, are arbitrary nucleotide bases at which the researcher wants 36-state genotype probabilities. Finally, we collapse the 36-state genotype probabilities to eight founder allele dosages at each marker. This last step is optional, but often helpful, because the simplified models require specification of fewer parameters. We then treat the founder allele dosages as known quantities in subsequent steps.

After inferring founder allele dosages, we address the second major statistical challenge of QTL mapping: the "model selection" problem (K. W. Broman and Sen, 2009). Linear regression is widely used because of

its ease of implementation, computational speed, and interpretability of results. A genetically "additive" linear model, in which we assume a linear relationship between a trait's values and each founder allele's dosage, is a popular default model. (Daniel M Gatti et al., 2014; K. W. Broman, D. M. Gatti, et al., 2019). The R package `qtl2` provides a straightforward user interface to include covariates and interaction terms in the linear model (K. W. Broman, D. M. Gatti, et al., 2019).

Relatedness is one issue that arises in the Diversity Outbred mice that is not a major concern in two-parent crosses. Due to the complexity of the breeding design, Diversity Outbred mice have complicated pairwise relationships. A given study cohort may include pairs of first cousins, parent-offspring pairs, grandparent-grandchild pairs, second cousin pairs, and others. Contrasting this with the two-parent backcross (Figure 1.3) and intercross (Figure 1.4) designs, we see in the backcross that average relatedness between any two members of the BC generation is equal to the same, shared value. One measure of relatedness is the probability of identity by state, *i.e.*, the probability that a randomly chosen allele from one subject is identical to a randomly chosen allele of the second subject. Measured in terms of the probability that a randomly chosen allele from one subject is identical to a randomly chosen allele of a second subject, we note that this probability of identity by state is $\frac{5}{8}$ in the backcross. F_2 subject pairs in an intercross, likewise, have a single, shared average relatedness, with an identity by state probability of $\frac{1}{2}$.

Because relatedness can confound genotype-phenotype associations (J. Yang et al., 2014), researchers have developed methods to account for relatedness in their statistical models. One popular approach involves use of a polygenic random effect in the statistical model (Kang, Zaitlen, et al., 2008). The addition of a random effect to a model with fixed (*i.e.*, nonrandom) effects results in what statisticians call a "linear mixed effects model".

1.7 Potential benefits of a pleiotropy test in multiparental populations

The complex traits genetics community needs a pleiotropy test for multiparental populations. Having such a test, when experimentalists are measuring tens of thousands of traits on subjects from multiparental populations, would enable our community to address new and interesting biological questions. For example, we can dissect an expression trait QTL hotspot by asking how many separate QTL are present within the hotspot. Additionally, we can inform experimental validation studies by addressing whether a pair of putatively related traits share a pleiotropic QTL.

I present in Chapter 2 a pleiotropy test for multiparental populations. I discuss in detail our statistical methods before studying the test's statistical properties, including type I error rate and power. I then apply the test to data from Diversity Outbred mice. In Chapter 3, I study the test through three vignettes. I first examine its contributions relative to mediation analysis in the dissection of expression QTL hotspots. I then examine statistical power to distinguish local expression trait QTLs. I conclude Chapter 3 by applying the pleiotropy test to two microbiome-related traits, where we obtain a result consistent with a pleiotropic QTL that affects both traits. Chapter 4 consists of a user guide to the R software package `qtl2pleio`, which is available on Github (https://github.com/fboehm/qtl2pleio, https://fboehm.us/static/software/qtl2pleio/). The Github repository also features integration with the Binder software project. Thus, a user may explore the code in an interactive R session.

Chapter 2

Methods development

2.1 Introduction

Complex trait studies in multiparental populations present new challenges in statistical methods and data analysis. Among these is the development of strategies for multivariate trait analysis. The joint analysis of two or more traits allows one to address additional questions, such as whether two traits share a single pleiotropic locus.

Previous research addressed the question of pleiotropy vs. separate QTL in two-parent crosses. Jiang and Zeng (1995) developed a likelihood ratio test for pleiotropy vs. separate QTL for a pair of traits. Their approach assumed that each trait was affected by a single QTL. Under the null hypothesis, the two traits were affected by a common QTL, and under the alternative hypothesis the two traits were affected by distinct QTL. Knott and Haley (2000) used linear regression to develop a fast approximation to the test of Jiang and Zeng (1995), while Tian et al. (2016) used the methods from Knott and Haley (2000) to dissect QTL hotspots in a F_2 population.

Multiparental populations, such as the Diversity Outbred (DO) mouse population (Churchill, Daniel M Gatti, et al., 2012), enable high-precision mapping of complex traits (Koning and McIntyre, 2014). The DO mouse population began with progenitors of the Collaborative Cross (CC) mice (Churchill, Airey, et al., 2004) Each DO mouse is a highly heterozygous genetic mosaic of alleles from the eight CC founder lines. Random matings among non-siblings have maintained the DO population for more than 23 generations (Chesler et al., 2016).

Several limitations of previous pleiotropy vs. separate QTL tests prevent their direct application in

multiparental populations. First, multiparental populations can have complex patterns of relatedness among subjects, and failure to account for these patterns of relatedness may lead to spurious results (J. Yang et al., 2014). Second, previous tests allowed for only two founder lines (Jiang and Zeng, 1995). Finally, Jiang and Zeng (1995) assumed that the null distribution of the test statistic follows a chi-square distribution.

We developed a pleiotropy vs. separate QTL test for two traits in multiparental populations. Our test builds on research that Jiang and Zeng (1995), Knott and Haley (2000), Tian et al. (2016), and Zhou and Stephens (2014) initiated. Our innovations include the accommodation of k founder alleles per locus (compared to the traditional two founder alleles per locus) and the incorporation of multivariate polygenic random effects to account for relatedness. Furthermore, we implemented a parametric bootstrap to calibrate test statistic values (Efron, 1979; Tian et al., 2016).

Below, we describe our likelihood ratio test for pleiotropy vs. separate QTL. In simulation studies, we find that it is slightly conservative, and that it has power to detect two separate loci when the univariate LOD peaks are strong. We further illustrate our approach with an application to data on a pair of behavior traits in a population of 261 DO mice (Logan et al., 2013; Recla et al., 2014). We find modest evidence for distinct QTL in a 2.5-cM region on mouse Chromosome 8.

2.2 Methods

Our strategy involves first identifying two traits that map to a common genomic region. We then perform a two-dimensional, two-QTL scan over the genomic region, with each trait affected by one QTL of varying position. We identify the QTL position that maximizes the likelihood under pleiotropy (that is, along the diagonal where the two QTL are at a common location), and the ordered pair of positions that maximizes the likelihood under the model where the two QTL are allowed to be distinct. The logarithm of the ratio of the two likelihoods is our test statistic. We calibrate this test statistic with a parametric bootstrap.

2.3 Data structures

The data consist of three objects. The first is an n by k by m array of allele probabilities for n subjects with k alleles and m marker positions on a single chromosome [derived from the observed SNP genotype data by a hidden Markov model; see K. W. Broman, D. M. Gatti, et al. (2019)]. The second object is an n by 2 matrix of phenotype values. Each column is a phenotype and each row is a subject. The third object is an

n by c matrix of covariates, where each row is a subject and each column is a covariate.

One additional object is the genotype-derived kinship matrix, which is used in the linear mixed model to account for population structure. We are focusing on a defined genomic interval, and we prefer to use a kinship matrix derived by the "leave one chromosome out" (LOCO) method (J. Yang et al., 2014), in which the kinship matrix is derived from the genotypes for all chromosomes except the chromosome under test.

2.3.1 Statistical Models

Focusing on a pair of traits and a particular genomic region of interest, the next step is a two-dimensional, two-QTL scan (Jiang and Zeng, 1995). We consider two QTL with each affecting a different trait, and consider all possible pairs of locations for the two QTL. For each pair of positions, we fit the multivariate linear mixed effects model defined in Equation 2.1. Note that we have assumed an additive genetic model throughout our analyses, but extensions to design matrices that include dominance are straightforward.

$$vec(Y) = Xvec(B) + vec(G) + vec(E) \tag{2.1}$$

where Y is the n by 2 matrix of phenotypes values; X is a $2n$ by $2(k+c)$ matrix that contains the k allele probabilities for the two QTL positions and the c covariates in diagonal blocks; B is a $(k+c)$ by 2 matrix of allele effects and covariate effects; G is a n by 2 matrix of random effects; and E is a n by 2 matrix of random errors. n is the number of mice. The 'vec' operator stacks columns from a matrix into a single vector. For example, a 2 by 2 matrix inputted to 'vec' results in a vector with length 4. Its first two entries are the matrix's first column, while the third and fourth entries are the matrix's second column.

We also impose distributional assumptions on G and E:

$$G \sim MN_{nx2}(0, K, V_g) \tag{2.2}$$

and

$$E \sim MN_{nx2}(0, I, V_e) \tag{2.3}$$

where $MN_{nx2}(0, V_r, V_c)$ denotes the matrix-variate (n by 2) normal distribution with mean being the n by 2 matrix with all zero entries and row covariance V_r and column covariance V_c. We assume that G and E are independent.

2.3.2 Parameter inference and log likelihood calculation

Inference for parameters in multivariate linear mixed effects models is notoriously difficult and can be computationally intense (K. Meyer, 1989; K. Meyer, 1991). Thus, we estimate V_g and V_e under the null hypothesis of no QTL, and then take them as fixed and known in our two-dimensional, two-QTL genome scan. We use restricted maximum likelihood methods to fit the model:

$$vec(Y) = X_0 vec(B) + vec(G) + vec(E) \tag{2.4}$$

where X_0 is a $2n$ by $2(c+1)$ matrix whose first column of each diagonal block in X_0 has all entries equal to one (for an intercept); the remaining columns are the covariates.

We draw on our R implementation (Boehm, 2018a) of the GEMMA algorithm for fitting a multivariate linear mixed effects model with expectation-maximization (Zhou and Stephens, 2014). We use restricted maximum likelihood fits for the variance components V_g and V_e in subsequent calculations of the generalized least squares solution $\hat{B}$.

$$\hat{B} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} vec(Y) \tag{2.5}$$

where

$$\hat{\Sigma} = \hat{V}_g \otimes K + \hat{V}_e \otimes I_n \tag{2.6}$$

where $\otimes$ denotes the Kronecker product, K is the kinship matrix, and I_n is a n by n identity matrix. We then calculate the log likelihood for a normal distribution with mean $X vec(\hat{B})$ and covariance $\hat{\Sigma}$ that depends on our estimates of V_g and V_e (Equation 2.6).

2.3.3 Pleiotropy vs. separate QTL hypothesis testing framework

Our test applies to two traits considered simultaneously. Below, λ_1 and λ_2 denote putative locus positions for traits one and two. We quantitatively state the competing hypotheses for our test as:

$$\begin{aligned} H_0 &: \lambda_1 = \lambda_2 \\ H_A &: \lambda_1 \neq \lambda_2 \end{aligned} \tag{2.7}$$

Our likelihood ratio test statistic is:

$$\text{LOD} = \log_{10}\left[\frac{\max_{\lambda_1,\lambda_2} L(B,\Sigma,\lambda_1,\lambda_2)}{\max_{\lambda} L(B,\Sigma,\lambda,\lambda)}\right] \tag{2.8}$$

where L is the likelihood for fixed QTL positions, maximized over all other parameters.

2.3.4 Visualizing profile LOD traces

The output of the above analysis is a two-dimensional $\log_{10}$ likelihood surface. To visualize these results, we followed an innovation of Zeng et al. (2000) and Tian et al. (2016), and plot three traces: the results along the diagonal (corresponding to the null hypothesis of pleiotropy), and then the profiles derived by fixing one QTL's position and maximizing over the other QTL's position.

We define the LOD score for our test:

$$\text{LOD}(\lambda_1,\lambda_2) = ll_{10}(\lambda_1,\lambda_2) - \max ll_{10}(\lambda,\lambda) \tag{2.9}$$

where ll_{10} denotes $\log_{10}$ likelihood.

We follow Zeng et al. (2000) and Tian et al. (2016) in defining profile LOD by the equation

$$\text{profile LOD}_1(\lambda_1) = \max_{\lambda_2} \text{LOD}(\lambda_1,\lambda_2) \tag{2.10}$$

We define profile $\text{LOD}_2(\lambda_2)$ analogously. The maximum value for the profile LOD_1 profile LOD_2 traces are the same and are non-negative, and give the overall LOD test statistic.

We construct the pleiotropy trace by calculating the log-likelihoods for the pleiotropic models at every position.

$$LOD_p(\lambda) = ll_{10}(\lambda,\lambda) - \max ll_{10}(\lambda,\lambda) \tag{2.11}$$

By definition, the maximum value for this pleiotropy trace is zero.

2.3.5 Bootstrap for test statistic calibration

We use a parametric bootstrap to calibrate our test statistic (Efron, 1979). While Jiang and Zeng (1995) used quantiles of a chi-squared distribution to determine p-values, this does not account for the two-dimensional

search over QTL positions. We follow the approach of Tian et al. (2016), and identify the maximum likelihood estimate of the QTL position under the null hypothesis of pleiotropy. We then use the inferred model parameters under that model and with the QTL at that position to simulate bootstrap data sets according to the model in equations 2.1–2.3. For each of b bootstrap data sets, we perform a two-dimensional QTL scan (over the genomic region of interest) and derive the test statistic value. We treat these b test statistics as the empirical null distribution, and calculate a p-value as the proportion of the b bootstrap test statistics that equal or exceed the observed one, with the original data, $p = \#\{i : \text{LOD}_i^* \geq \text{LOD}\}/b$ where LOD_i^* denotes the LOD score for the ith bootstrap replicate and LOD is the observed test statistic.

2.3.6 Data & Software Availability

Our methods have been implemented in an R package, `qtl2pleio`, available at GitHub:

https://github.com/fboehm/qtl2pleio

Custom R code for our analyses and simulations are at GitHub:

https://github.com/fboehm/qt2pleio-manuscript-clean

The data from Recla et al. (2014) and Logan et al. (2013) are available at the Mouse Phenome Database:

https://phenome.jax.org/projects/Chesler4 and https://phenome.jax.org/projects/Recla1.

They are also available in R/qtl2 format at https://github.com/rqtl/qtl2data.

2.4 Simulation studies

We performed two types of simulation studies, one for type I error rate assessment and one to characterize the power to detect separate QTL. To simulate traits, we specified X, B, V_g, K, and V_e matrices (Equations 2.1–2.3). For both we used the allele probabilities from a single genomic region derived empirically from data for a set of 479 Diversity Outbred mice from Keller et al. (2018).

2.4.1 Type I error rate analysis

To quantify type I error rate (*i.e.*, false positive rate), we simulated 400 pairs of traits for each of eight sets of parameter inputs (Table 2.1). We used a 2^3 factorial experimental design with three factors: allele effects difference, allele effects partitioning, and genetic correlation, *i.e.*, the off-diagonal entry in the 2 by 2 matrix V_g.

Table 2.1: Type I error rates for all runs in our 2^3 experimental design. We set (marginal) genetic variances (*i.e.*, diagonal elements of V_g) to 1 in all runs. V_e was set to the 2 by 2 identity matrix in all runs. We used allele probabilities at a single genetic marker to simulate traits for all eight sets of parameter inputs. In the column "Allele effects partitioning", "ABCD:EFGH" means that lines A–D carry one QTL allele while lines E–H carry the other allele. "F:ABCDEGH" means the QTL has a private allele in strain F.

Run	Δ(Allele effects)	Allele effects partitioning	Genetic correlation	Type I error rate
1	6	ABCD:EFGH	0	0.032
2	6	ABCD:EFGH	0.6	0.035
3	6	F:ABCDEGH	0	0.040
4	6	F:ABCDEGH	0.6	0.045
5	12	ABCD:EFGH	0	0.038
6	12	ABCD:EFGH	0.6	0.042
7	12	F:ABCDEGH	0	0.025
8	12	F:ABCDEGH	0.6	0.025

We chose two strong allele effects difference values, 6 and 12. These ensured that the univariate phenotypes mapped with high LOD scores to the region of interest. For the allele partitioning factor, we used either equally frequent QTL alleles, or a private allele in the CAST strain (F). For the residual genetic correlation (the off-diagonal entry in V_g), we considered the values 0 and 0.6. The marginal genetic variances (*i.e.*, the diagonal entries in V_g) for each trait were always set to one.

We performed 400 simulation replicates per set of parameter inputs, and each used $b = 400$ bootstrap samples. For each bootstrap sample, we calculated the test statistic (Equation 2.8). We then compared the test statistic from the simulated trait against the empirical distribution of its 400 bootstrap test statistics. When the simulated trait's test statistic exceeded the 0.95 quantile of the empirical distribution of bootstrap test statistics, we rejected the null hypothesis. We observed that the test is slightly conservative over our range of parameter selections (Table 2.1), with estimated type I error rates < 0.05.

2.4.2 Power analysis

We also investigated the power to detect the presence of two distinct QTL. We used a $2 \times 2 \times 5$ experimental design, where our three factors were allele effects difference, allele effects partitioning, and inter-locus distance. The two levels of allele effects difference were 1 and 2. The two levels of allele effects partitioning were as in the type I error rate studies, ABCD:EFGH and F:ABCDEGH (Table C.1). The five levels of interlocus distance were 0, 0.5, 1, 2, and 3 cM. V_g and V_e were both set to the 2 by 2 identity matrix in all power study simulations.

We simulated 400 pairs of traits per set of parameter inputs. For each simulation replicate, we calculated

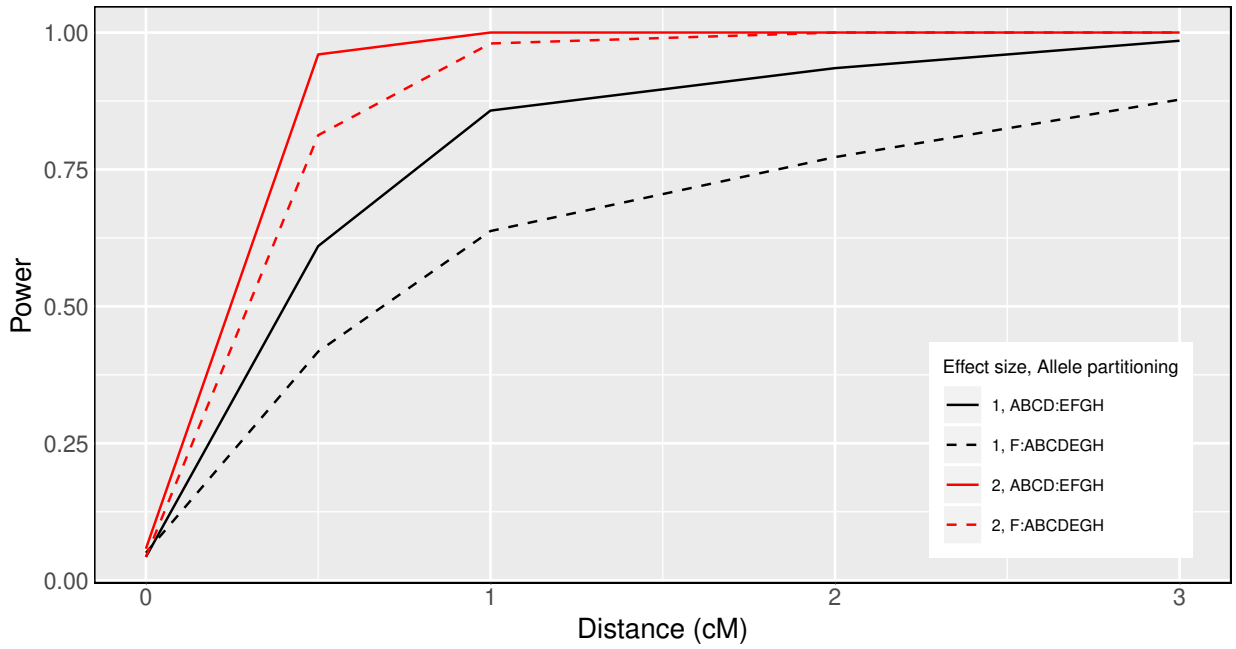


Figure 2.1: Pleiotropy vs. separate QTL power curves for each of four sets of parameter settings. Factors that differ among the four curves are allele effects difference and allele partitioning. Red denotes high allele effects difference, while black is the low allele effects difference. Solid line denotes the even allele partitioning (ABCD:EFGH), while dashed line denotes the uneven allele partitioning (F:ABCDEGH).

the likelihood ratio test statistic. We then applied our parametric bootstrap to calibrate the test statistics. For each simulation replicate, we used $b = 400$ bootstrap samples. Because the bootstrap test statistics within a single set of parameter inputs followed approximately the same distribution, we pooled the $400 * 400 = 160,000$ bootstrap samples per set of parameter inputs and compared each test statistic to the empirical distribution derived from the 160,000 bootstrap samples. However, for parameter inputs with interlocus distance equal to zero, we didn't pool the 160,000 bootstrap samples; instead, we proceeded by calculating power (*i.e.*, type I error rate, in this case), as we did in the type I error rate study above.

We present our power study results in Figure 2.1. Power increases as interlocus distance increases. The top two curves correspond to the case where the QTL effects are largest. For each value for the QTL effect, power is greater when the QTL alleles are equally frequent, and smaller when a QTL allele is private to one strain. One can have high power to detect that the two traits have distinct QTL when they are separated by > 1 cM and when the QTL have large effect.

2.5 Application

To illustrate our methods, we applied our test to data from Logan et al. (2013) and Recla et al. (2014), on 261 DO mice measured for a set of behavioral phenotypes. Recla et al. (2014) identified *Hydin* as the gene that underlies a QTL on Chromosome 8 at 57 cM for the "hot plate latency" phenotype (a measure of pain tolerance). The phenotype "percent time in light" in a light-dark box (a measure of anxiety) was measured on the same set of mice (Logan et al., 2013) and also shows a QTL near this location, which led us to ask whether the same locus affects both traits. The two traits show a correlation of -0.15 (Figure C.1).

QTL analysis with the LOCO method, and using sex as an additive covariate, showed multiple suggestive QTL for each phenotype (Figure C.2; Table C.2). For our investigation of pleiotropy, we focused on the interval 53–64 cM on Chromosome 8. The univariate QTL results for this region are shown in Figure 2.2.

The estimated QTL allele effects for the two traits are quite different (Figure 2.3). With the QTL placed at 55 cM, for "percent time in light", the WSB and PWK alleles are associated with large phenotypes and NOD with low phenotypes. For "hot plate latency", on the other hand, CAST and NZO show low phenotypes and NOD and PWK are near the center.

In applying our test for pleiotropy, we performed a two-dimensional, two-QTL scan for the pair of phenotypes. With these results, we created a profile LOD plot (Figure 2.4). The profile LOD for "percent time in light" (in brown) peaks near 55 cM, as was seen in the univariate analysis. The profile LOD for "hot plate latency" (in blue) peaks near 57 cM, also similar to the univariate analysis. The pleiotropy trace (in gray) peaks near 55 cM.

The likelihood ratio test statistic for the test of pleiotropy was 1.2. Based on a parametric bootstrap with 1,000 bootstrap replicates, the estimated p-value was 0.11, indicating weak evidence for distinct QTL for the two traits.

2.6 Discussion

We developed a test of pleiotropy vs. separate QTL for multiparental populations, extending the work of Jiang and Zeng (1995) for multiple alleles and with a linear mixed model to account for population structure (Kang, Sul, et al., 2010; J. Yang et al., 2014). Our simulation studies indicate that the test has power to detect presence of separate loci, especially when univariate trait associations are strong (Figure 2.1). Type I error rates indicate that our test is slightly conservative (Table 2.1).

In the application of our method to two behavioral phenotypes in a study of 261 Diversity Outbred mice

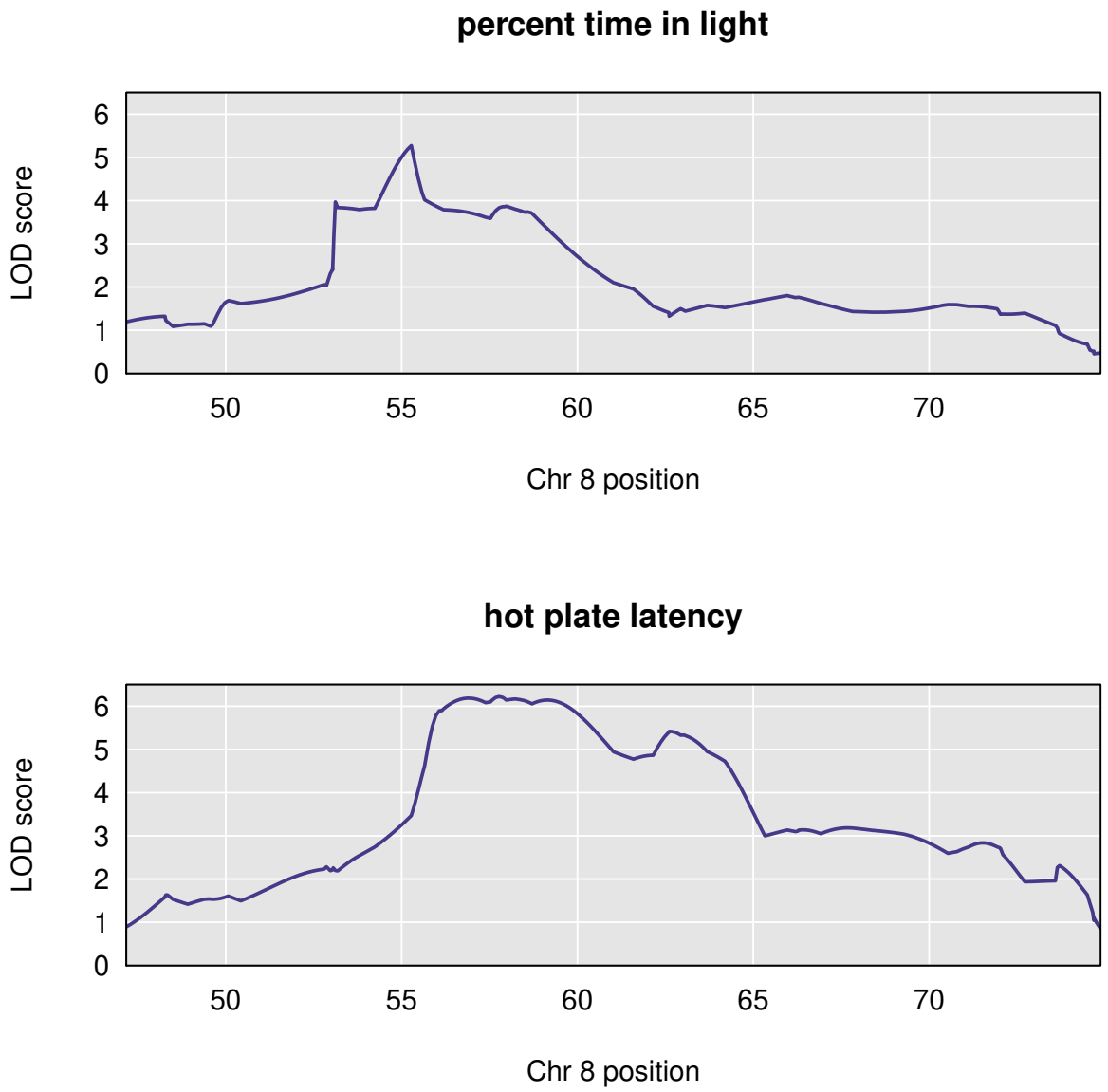


Figure 2.2: Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins near 56 cM and ends about 64 cM.

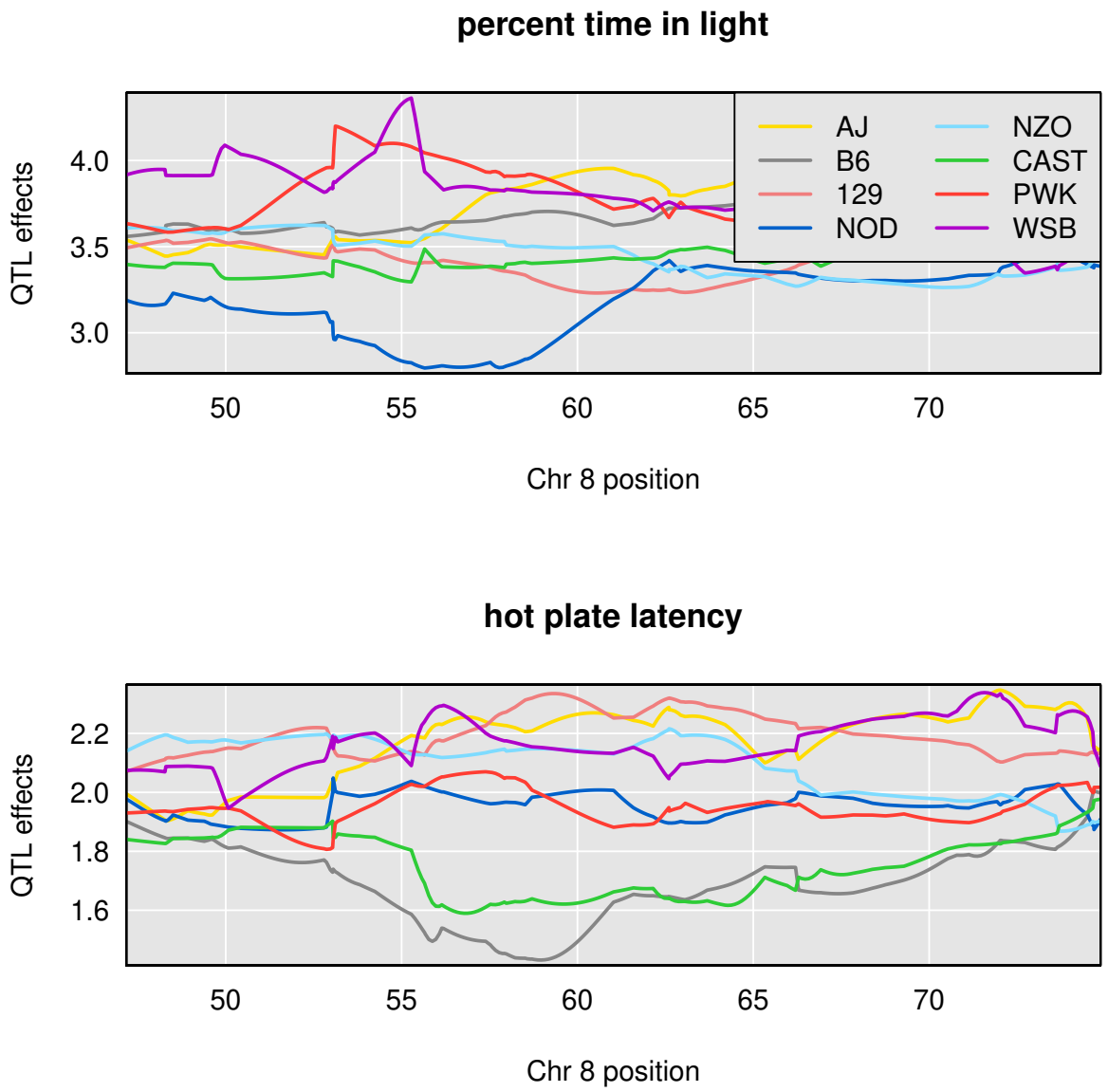


Figure 2.3: Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins near 56 cM and ends about 64 cM.

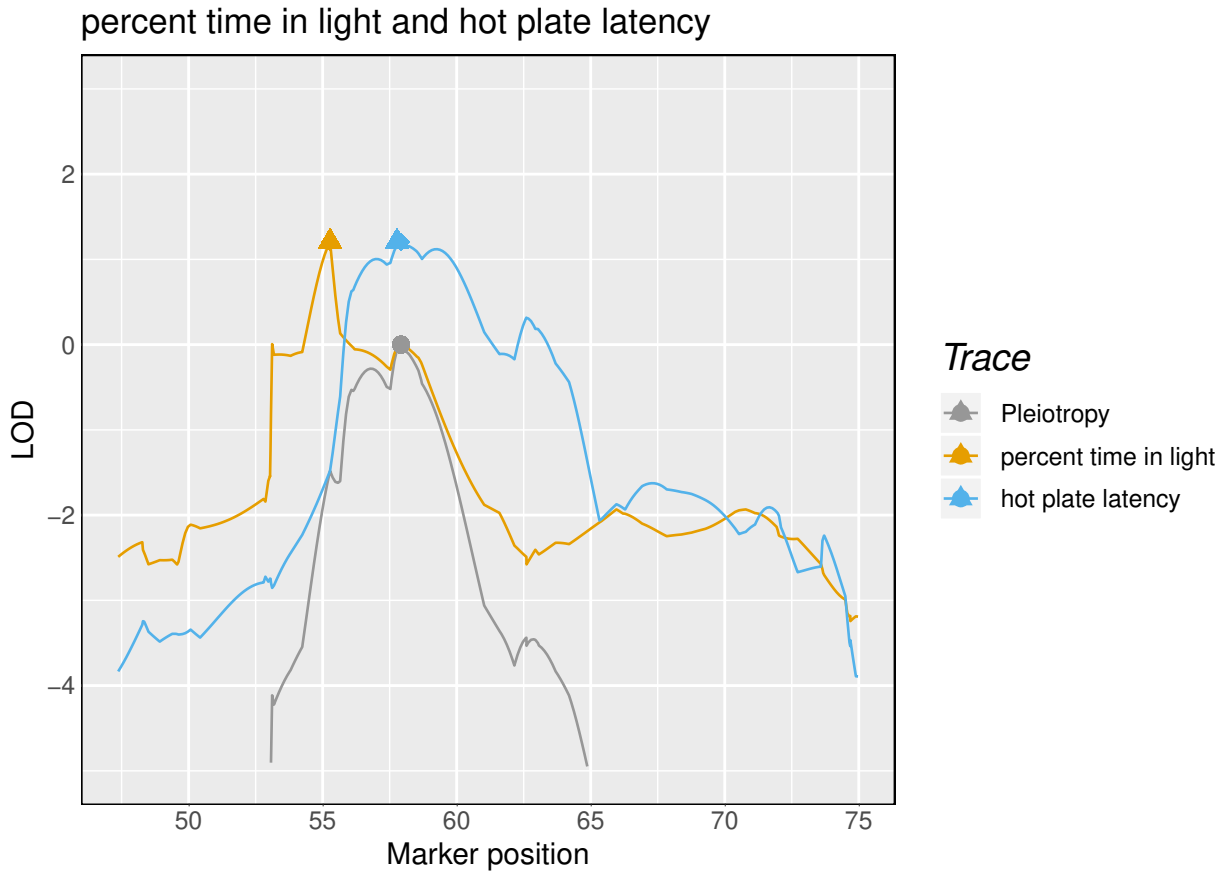


Figure 2.4: Profile LOD curves for the pleiotropy vs. separate QTL hypothesis test for "percent time in light" and "hot plate latency". Gray trace denotes pleiotropy LOD values. Triangles denote the univariate LOD maxima, while diamonds denote the profile LOD maxima. For "percent time in light", the brown triangle obscures the smaller brown diamond. Likelihood ratio test statistic value corresponds to the height of the blue and brown traces at their maxima.

(Recla et al., 2014; Logan et al., 2013), we obtained weak evidence (p=0.11) for the presence of two distinct QTL, with one QTL (which contained the *Hydin* gene) affecting only "hot plate latency" and a second QTL affecting "percent time in light" (Figure 2.4).

Founder allele effects plots provide further evidence for the presence of two distinct loci. As Macdonald and Long (2007) and King et al. (2012) have demonstrated in their analyses of multiparental *Drosophila* populations, a biallelic pleiotropic QTL would result in allele effects plots that have similar patterns. While we don't know that "percent time in light" and "hot plate latency" arise from biallelic QTL, the dramatic differences that we observe in allele effects patterns further support the argument for two distinct loci.

We have implemented our methods in an R package `qtl2pleio`, but analyses can be computationally intensive and time consuming. `qtl2pleio` is written mostly in R, and so we could likely obtain improved computational speed by porting parts of the calculations to a compiled language such as C or C++. To accelerate our multi-dimensional QTL scans, we have integrated C++ code into `qtl2pleio`, using the Rcpp package (Eddelbuettel et al., 2011).

Another computational bottleneck is the estimation of the variance components V_g and V_e. To accelerate this procedure, especially for the joint analysis of more than two traits, we will consider other strategies for variance component estimation, including that described by H. V. Meyer et al. (2018). H. V. Meyer et al. (2018), in joint analysis of dozens of traits, implement a bootstrap strategy to estimate variance components for lower-dimensional phenotypes before combining bootstrap estimates into valid covariance matrices for the full multivariate phenotype. Such an approach may ease some of the computational burdens that we encountered.

We view tests of pleiotropy as complementary to mediation tests and related methods that have become popular for inferring biomolecular causal relationships (Chick et al., 2016; Schadt et al., 2005; Baron and Kenny, 1986). A mediation test proceeds by including a putative mediator as a covariate in the regression analysis of phenotype and QTL genotype; a substantial reduction in the association between genotype and phenotype corresponds to evidence of mediation.

Mediation analyses and our pleiotropy test ask distinct, but related, questions. Mediation analysis seeks to establish causal relationships among traits, including molecular traits, or dependent biological and behavioral processes. Pleiotropy tests examine whether two traits share a single source of genetic variation, which may act in parallel or in a causal network. Pleiotropy is required for causal relations among traits. In many cases, the pleiotropy hypothesis is the only reasonable one.

Schadt et al. (2005) argued that both pleiotropy tests and causal inference methods may contribute to

gene network reconstruction. They developed a model selection strategy, based on the Akaike Information Criterion (Akaike, 1974), to determine which causal model is most compatible with the observed data. Schadt et al. (2005) extended the methods of Jiang and Zeng (1995) to consider more complicated alternative hypotheses, such as the possibility of two QTL, one of which associates with both traits, and one of which associates with only one trait. As envisioned by Schadt et al. (2005), we foresee complementary roles emerging for our pleiotropy test and mediation tests in the dissection of complex trait genetic architecture.

CAPE (Combinatorial Analysis of Pleiotropy and Epistasis) is a strategy for identifying higher-order relationships among traits and marker genotypes (Tyler, W. Lu, et al., 2013). Tyler, Ji, et al. (2017) used CAPE to identify epistatic gene networks in Diversity Outbred mice. Tyler, Donahue, et al. (2016) found evidence for weak epistasis in a large intercross population.

CAPE uses linear models that are distinct from those in our pleiotropy test. A CAPE starts with founder allele dosages at all markers and a collection of two or more traits (Tyler, Ji, et al., 2017). After eigendecomposition to get two or more eigentraits, univariate QTL scans are performed and founder allele effects are estimated at all markers (or a subset of all markers). Next, one identifies markers with sufficiently strong effects of at least one founder allele. Resulting (eigentrait, marker, founder allele) triples are then subjected to a second model fitting. This second round of modeling involves two (eigentrait, marker, founder allele) triples that share an eigentrait. The shared (univariate) eigentrait is modeled as a linear function of the two founder allele dosages and their interaction.

In comparing CAPE and our pleiotropy test, it's important to recognize that the two methods ask different questions. CAPE enables assessment of interactions among specific founder allele dosages at two (possibly identical) markers while examining one eigentrait at a time. Our pleiotropy test, on the other hand, jointly models two phenotypes and quantifies the evidence against the pleiotropy hypothesis by performing a two-dimensional QTL scan over a genomic region. One limitation of CAPE is its inability to account for population structure. Incorporation of a polygenic random effect into CAPE's linear models may improve its performance in multiparental populations. CAPE's methods also highlight a current limitation of our pleiotropy test. While our multivariate linear models accommodate interactions between founder allele dosages at two markers, we haven't yet incorporated this functionality into our software. It is one direction for future research.

Technological advances in mass spectrometry and RNA sequencing have enabled the acquisition of high-dimensional biomolecular phenotypes (Ozsolak and Milos, 2011; X. Han, K. Yang, and Gross, 2012). Multiparental populations in *Arabidopsis*, maize, wheat, oil palm, rice, *Drosophila*, yeast, and other organisms

enable high-precision QTL mapping (Yu et al., 2008; Tisne et al., 2017; Stanley et al., 2017; Raghavan et al., 2017; Mackay et al., 2012; Kover et al., 2009; Cubillos et al., 2013). The need to analyze high-dimensional phenotypes in multiparental populations compels the scientific community to develop tools to study genotype-phenotype relationships and complex trait architecture. Our test, and its future extensions, will contribute to these ongoing efforts.

Acknowledgments

The authors thank Lindsay Traeger, Julia Kemis, and Rene Welch for valuable suggestions to improve the manuscript. This work was supported in part by National Institutes of Health grant R01GM070683 (to K.W.B.). The research made use of compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences at UW-Madison, which is supported by the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

Chapter 3

Applications

3.1 Expression trait hotspot dissection

3.1.1 Introduction

A central goal of systems genetics studies is to identify causal relationships between biomolecules. Recent work by Chick et al. (2016), which builds on research from Baron and Kenny (1986), has popularized linear regression-based methods, such as mediation analysis, for causal inference in genetics. Because of the great successes of mediation analysis in systems genetics (Chick et al., 2016; Keller et al., 2018), we need to clarify a role for our pleiotropy test. We argue below that our pleiotropy test complements mediation analysis in two ways. First, our test limits the set of candidate mediators by ruling out traits that don't share a single pleiotropic QTL. This is reasonable because it's unlikely that one trait mediates a relationship between a genetic variant and a second trait unless the two traits share a pleiotropic QTL. Second, when regression-based mediation analysis fails to identify a mediator, our pleiotropy test still provides information on the number of QTL, which may aid biological understanding and inform subsequent studies. Below, we first review the prerequisite molecular biology, including the "central dogma", before discussing in some detail regression-based mediation analysis in the systems genetics context.

F. H. Crick (1958) articulated a pathway for transmission of biological information that is now known as the central dogma of molecular biology (Figure 3.1) (F. Crick, 1970). In it, he argued that information encoded in DNA sequence is transmitted via transcription to RNA molecules, which, in turn, transfer the information to proteins via translation. The process of transcription uses DNA as a template for creating a

Figure 3.1: Biological information is encoded in DNA. This information is passed, via transcription, to sequence-specific RNA molecules. The process of translation transmits the information to sequence-specific proteins.

RNA molecule that conveys the information encoded in DNA. In this sense, every gene leads to a unique RNA molecule. Translation is the molecular biology process by which a RNA molecule's information is transferred to a protein. As in transcription, the nucleic acid (DNA in transcription and RNA in translation) serves as a template for synthesis of the product (RNA in transcription and protein in translation). Thus, RNA molecules from distinct genes lead to different proteins.

This sequence of information transfer, from DNA to RNA to protein, provides a natural setting by which to examine mediation analysis. If a DNA variant affects protein concentrations only through its gene's RNA transcripts, then conditioning on RNA transcript levels would greatly reduce the strength of association between DNA variant and protein concentration. Before we continue our discussion, we define key terms and discuss an example below.

We continue by stating what it means for one trait to mediate a relationship between a DNA variant and another trait. To clarify our discussion, we refer to an example from Chick et al. (2016) (Figure 3.2). Chick et al. (2016), in studying livers of 192 Diversity Outbred mice, found evidence that *Dhtkd1* transcript levels associated with a Chromosome 2 marker near the *Dhtkd1* gene. They also found that the same marker affected DHTKD1 protein concentrations. (Note that DHTKD1 protein is the product of translation of *Dhtkd1* transcripts.) As anticipated, mediation analysis, in which the DHTKD1 protein concentrations are regressed on founder allele dosages (at the Chromosome 2 marker) demonstrated that *Dhtkd1* transcript levels act as a mediator between DHTKD1 protein concentrations and founder allele dosages. In fact, the extent of the reduction in association strength indicates that the primary pathway by which the genetic marker affects DHTKD1 protein concentrations is through *Dhtkd1* transcript levels.

In our studies below, we follow Keller et al. (2018) by generalizing this setting to the case where a DNA variant affects a local transcript level, which then affects a nonlocal transcript level. We term a transcript "local" to a marker when its gene is near that marker. We use a threshold of no more than 2 Mb to restrict

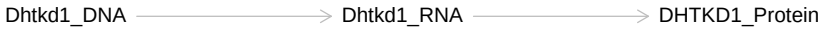


Figure 3.2: A DNA variant in the *Dhtkd1* gene affects *Dhtkd1* transcript abundances which, in turn, affect DHTKD1 protein concentrations.

the number of local transcripts for a given marker. A nonlocal transcript, then, is either one that arises from a gene on another chromosome or from a distant gene on the same chromosome as the marker.

It is highly plausible that concentration variations in one transcript may affect abundances of a second transcript. For example, the first transcript may encode a transcription factor protein. In this case, the transcription factor protein may influence expression patterns of the second transcript (and perhaps other transcripts, too).

To determine whether a local transcript level mediates the relationship between a nonlocal transcript level and a DNA variant, we perform a series of regression analyses, which we detail below (Frame 3.1). In brief, we regress the nonlocal transcript levels on founder allele dosages at the DNA variant, with and without conditioning on the candidate mediator (the local transcript levels). If the LOD score diminishes sufficiently upon conditioning on a candidate, then we declare the candidate a mediator.

The rationale behind this strategy follows. If the DNA variant affects the nonlocal transcript levels solely by way of local transcript levels, then conditioning on the local transcript levels would nullify the relationship between DNA variant and the nonlocal transcript levels. At the other extreme, if the DNA variant affects nonlocal transcript levels solely through mechanisms that don't involve the local transcript levels, then conditioning on local transcript levels would not affect the association between the DNA variant and nonlocal transcript levels.

We now consider the procedures needed for a mediation analysis in systems genetics. In the four linear regression models (Frame 3.1), X is a n by 8 matrix of founder allele dosages at a single marker, B is a 8 by 1 matrix of founder allele effects, E is a n by 1 matrix of random errors, b is an number, 1 is a n by 1 matrix with all entries set to 1, Y is a n by 1 matrix of phenotype values (for a single trait), and M is a n by 1 matrix of values for a putative mediator. C is a matrix of covariate effects, and W is a matrix of covariates. We denote the coefficient of the mediator by β.

Frame 3.1: Four regressions for a single mediation analysis

$$Y = b1 + WC + E \tag{3.1}$$

$$Y = XB + WC + E \tag{3.2}$$

$$Y = b1 + WC + M\beta + E \tag{3.3}$$

$$Y = XB + WC + M\beta + E \tag{3.4}$$

We assume that the vector E is (multivariate) normally distributed with zero vector as mean and covariance matrix $\Sigma = \sigma^2 I_n$, where I_n is the n by n identity matrix.

In the above models with normally distributed random errors, the log-likelihoods are easily calculated. For example, in Equation 3.1, the vector Y follows a multivariate normal distribution with mean $(b1 + WC)$ and covariance $\Sigma = \sigma^2 I$. Thus, we can write the likelihood for Model 3.1 as:

$$L(b, C, \sigma^2 | Y, W) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}(Y - b1 - WC)^T \Sigma^{-1} (Y - b1 - WC)\right) \tag{3.5}$$

We thus have the following equation (3.6 for the log-likelihood for Model 3.1:

$$\log L(b, C, \sigma^2 | Y, W) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}(Y - b1 - WC)^T \Sigma^{-1} (Y - b1 - WC) \tag{3.6}$$

Chick et al. (2016) calculated the $\log_{10}$ likelihoods for all four models before determining two LOD scores (Equations 3.7 and 3.8).

$$LOD_1 = log_{10}(\text{Model 3.2 likelihood}) - log_{10}(\text{Model 3.1 likelihood}) \tag{3.7}$$

$$LOD_2 = log_{10}(\text{Model 3.4 likelihood}) - log_{10}(\text{Model 3.3 likelihood}) \tag{3.8}$$

And, finally, Chick et al. (2016) calculated the LOD difference statistic (Equation 3.9).

$$\text{LOD difference} = LOD_1 - LOD_2 \tag{3.9}$$

LOD difference values need not be positive. For example, in the setting where the putative mediator is

Frame 3.2: Four assumptions for causal inference

1. No unmeasured confounding of the DNA variant-nonlocal transcript levels relationship
2. No unmeasured confounding of the local transcript levels-nonlocal transcript levels relationship
3. No unmeasured confounding of the DNA variant-local transcript levels relationship
4. No local transcript levels-nonlocal transcript levels confounder that is affected by the DNA variant

not a true mediator, LOD_1 and LOD_2 values may lead to a negative LOD difference statistic.

In our analyses below, we consider the LOD difference proportion (Equation 3.10).

$$\text{LOD difference proportion} = \frac{(LOD_1 - LOD_2)}{LOD_1} \tag{3.10}$$

In other words, we consider what proportion of the association strength, on the LOD scale, is diminished by conditioning on a putative mediator. This statistic differs from the LOD difference statistic in that it scales the LOD difference by LOD_1 value. We do this in efforts to accommodate the diversity of LODs in our data. For example, a LOD difference statistic of 5 may be relevant when a trait has a LOD_1 of 10, but unimportant if the trait's LOD_1 is 100. Now that we've named our summary statistics for a mediation analysis in systems genetics, we turn attention to two statistical challenges in this area of research, 1. confounding and 2. significance thresholds.

Our first statistical challenge is due to confounding. To claim that the LOD difference statistic reflects a causal relationship, four assumptions about confounding are needed (Frame 3.2) (VanderWeele, 2015), yet it is often difficult or impossible to recognize unmeasured confounders. In studies of Diversity Outbred mice, relatedness is a possible confounder, yet the linear models (Equations 3.1, 3.2, 3.3, 3.4 in Frame 3.1) fail to account for the complex relatedness patterns among Diversity Outbred mice. In fact, the only covariates in our models are for wave membership and sex. Other sources of confounding, such as batch effects in the phenotyping, may be unmeasured.

In efforts to quantify the potential impact of unmeasured confounding, scientists have developed a suite of sensitivity analysis tools for use in regression-based mediation analysis. While a discussion of sensitivity analysis is beyond the scope of this thesis, it may be useful in future systems genetics studies to assess robustness of mediation analysis results in the presence of unmeasured confounders. VanderWeele (2015) discusses sensitivity analysis in the context of epidemiological studies.

Besides accounting for confounding, a second statistical challenge in mediation analysis is assessing signif-

icance of the LOD difference statistic. Chick et al. (2016) used individual transcript levels as sham mediators and tabulated their LOD difference statistics. They then compared the observed LOD difference statistics for putative mediators to the empirical distribution of LOD difference statistics obtained from the collection of sham mediators. Keller et al. (2018), on the other hand, in their study of pancreatic islet cell biology, declared mediators those local transcripts that diminished the LOD score of nonlocal transcripts by at least 1.5. While significance threshold determination remains an active area of research, we proceed below by examining all 147 nonlocal transcript levels that Keller et al. (2018) identified as mapping to the Chromosome 2 hotspot.

3.1.2 Methods

We examined the potential that the two methods, 1. pleiotropy vs. separate QTL testing and 2. mediation analysis, play complementary roles in efforts to dissect gene expression trait hotspots. We use the term "hotspot" to refer to a contiguous genomic region, typically no more than 5 Mb in length, that affects many expression traits. After describing our data below, we detail our statistical analyses involving 13 local gene expression traits and 147 nonlocal gene expression traits, all of which map to a 4-Mb hotspot on Chromosome 2.

We analyzed data from 378 Diversity Outbred mice (Keller et al., 2018). Keller et al. (2018) genotyped tail biopsies with the GigaMUGA microarray (Morgan et al., 2015). They also used RNA sequencing to measure genome-wide pancreatic islet cell gene expression for each mouse at the time of sacrifice (Keller et al., 2018).

We examined the Chromosome 2 pancreatic islet cell expression trait hotspot that Keller et al. (2018) identified. Keller et al. (2018) found that 147 nonlocal traits map to the 4-Mb region centered at 165.5 Mb on Chromosome 2. The 147 nonlocal traits all exceeded the genome-wide LOD significance threshold, 7.18 (Keller et al., 2018). With regression-based mediation analyses, they identified transcript levels of local gene *Hnf4a* as a mediator of 88 of these 147 nonlocal traits.

We designed a study to examine the possible roles for mediation analysis and pleiotropy testing. Because Keller et al. (2018) reported that some nonlocal traits that map to the Chromosome 2 hotspot did not demonstrate evidence of mediation by *Hnf4a* expression levels, we elected to study a collection of local gene expression traits, rather than *Hnf4a* alone. This strategy enabled us to ask whether one of twelve other local traits mediates those nonlocal hotspot traits that are not mediated by *Hnf4a*. Our set of local gene expression traits includes *Hnf4a* and 12 other local genes (Table 3.1). Our 13 local genes are the only genes

Frame 3.3: Local gene inclusion criteria

1. QTL peak with LOD > 40
2. QTL peak position within 2 Mb of hotspot center (165.5 Mb)
3. Gene midpoint within 2 Mb of hotspot center (165.5 Mb)

that met three criteria (Frame 3.3). The 147 nonlocal traits that we studied all had LOD peak heights above 7.18 and QTL positions within 2 Mb of the center of the hotspot (at 165.5 Mb).

We now describe our statistical analyses. After univariate QTL mapping to identify expression traits that map to the Chromosome 2 hotspot, we performed both bivariate QTL scans and mediation analyses of all 13 * 147 = 1911 pairs involving one local expression trait and one nonlocal expression trait.

Our bivariate QTL analyses involved the same 13 local expression traits and 147 nonlocal expression traits. We described above (Frame 3.3) the criteria for choosing these expression traits.

We performed a series of two-dimensional QTL scans in which we paired each local gene's transcript levels with each nonlocal gene's transcript levels, for a total of 13 x 147 = 1,911 two-dimensional scans. Each scan examined the same set of 180 markers, which spanned the interval from 163.1 Mb to 167.8 Mb and included univariate peaks for all 13 + 147 = 160 expression traits. We performed these analyses with the R package `qtl2pleio` (Boehm, 2018b).

For each bivariate QTL scan, we fitted a collection of bivariate models for all 180 * 180 = 32,400 ordered pairs of markers. For each ordered pair of markers, we fitted a bivariate linear mixed effects model using the methods of Chapter 2.

We performed mediation analyses for all 1,911 local-nonlocal trait pairs in which we probed the extent to which each nonlocal trait's association strength diminished upon conditioning on transcript levels of a putative mediator.

Each of the 13 local expression traits, considered one at a time, served as putative mediators. We thus fitted the four linear regression models that we describe above (Equations 3.1, 3.2, 3.3, 3.4).

One question that needs clarification is the choice of genetic marker for each mediation analysis. We elected to use the founder allele dosages at the marker that demonstrated the univariate LOD peak for the putative mediator. Alternative analyses, in which one uses the founder allele dosages at which the nonlocal trait has its univariate peak, are also possible.

To visualize the summary statistics for our two methods, we plotted, for each local gene, a scatterplot of

LOD difference proportion values against pleiotropy test statistics.

Gene	Start	End	QTL peak position	LOD
Pkig	163.66	163.73	163.52	51.68
Serinc3	163.62	163.65	163.58	126.93
Hnf4a	163.51	163.57	164.02	48.98
Stk4	164.07	164.16	164.03	60.39
Pabpc1l	164.03	164.05	164.03	52.50
Slpi	164.35	164.39	164.61	40.50
Neurl2	164.83	164.83	164.64	64.58
Cdh22	165.11	165.23	165.05	53.84
2810408M09Rik	165.49	165.49	165.57	67.34
Eya2	165.60	165.77	165.72	98.89
Prex1	166.57	166.71	166.75	46.91
Ptgis	167.19	167.24	167.27	56.25
Gm14291	167.20	167.20	167.27	73.72

Table 3.1: Local gene annotations for analysis of Chromosome 2 expression trait hotspot. All positions are in units of Mb on Chromosome 2. LOD peak position and LOD peak height refer to those obtained from univariate analyses. "Start" and "end" refer to the local gene's DNA start and end positions, as annotated by Ensembl version 75.

We also examined the 1,911 pairs from the per-nonlocal gene perspective. For each of the 147 nonlocal genes, we plotted LOD difference proportion against pleiotropy test statistic values. We present below (Figure 3.6) examples that illustrate some of the observed patterns between pleiotropy test statistics and LOD difference proportion values.

3.1.3 Results

Below, we visualize results of our pleiotropy tests and mediation analyses. After examining a plot involving all tested pairs, we take two distinct perspectives. In these, we visualize results from the perspective of every local gene expression trait before taking the complementary perspective with visualizations from the perspective of every nonlocal gene expression trait. We present below our scatter plot for all 147 x 13 = 1,911 pairs of traits (Figure 3.3). Each pair contains one local expression trait and one nonlocal expression trait. Each point in the figure represents a single pair. We see that points with high values of LOD difference proportion tend to have small values of pleiotropy test statistic, and those points with high values of the pleiotropy test statistic tend to have small values of LOD difference proportion.

Some points demonstrate large values of pleiotropy test statistic (*eg.*, > 10), yet still have sizeable LOD difference proportion statistics (Figure 3.3. One possible explanation for such points is that the univariate LOD (LOD_1) value is small, so that even a modest LOD difference gives rise to a sizeable LOD difference

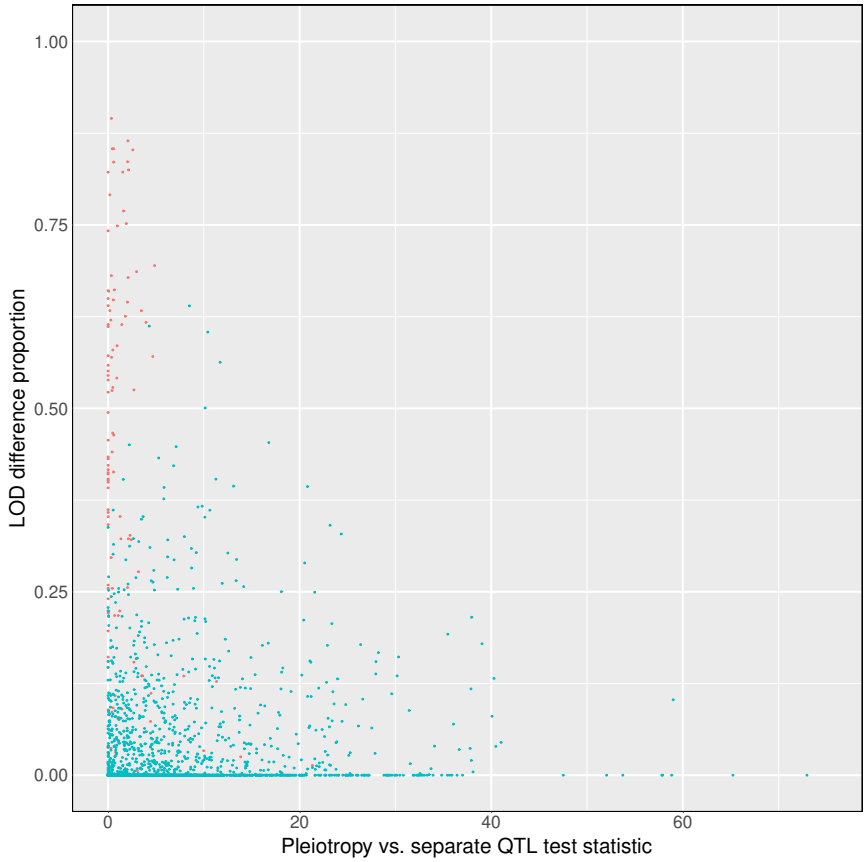


Figure 3.3: LOD difference proportion against pleiotropy vs. separate QTL test statistics for 1911 pairs of traits. Each point represents a single pair. Pairs that involve local expression trait Hnf4a are colored blue, while all others are colored red. Note the high prevalence of blue points in the upper left quadrant of the figure. These points, with low values of the pleiotropy vs. separate QTL test statistic and high values of LOD difference proportion, are consistent with *Hnf4a* transcript levels mediating the effect of *Hnf4a* genetic variants affecting nonlocal transcript abundances.

proportion value. A second possible explanation is that the phenotype variances for both phenotypes in the pair are large enough to skew the null distribution of the test statistic towards larger values. To distinguish between these two, we would examine the LOD difference statistics and obtain bootstrap p-values.

In Figure 3.3, we colored blue the points that involve *Hnf4a*; points with other local genes are red. The most striking feature of the coloring is that many blue points have small values of the pleiotropy test statistics and very high values of LOD difference proportion.

To more thoroughly examine the pleiotropy test and mediation analysis relationships across the 13 local genes, we created 13 plots of LOD difference proportion against pleiotropy test statistic (Figures 3.4 and 3.5). They reveal common patterns. First, we see no points in the upper right quadrant of each plot (Figure 3.5).

This tells us that those nonlocal genes with high values of pleiotropy test statistic (when paired with the specified local gene) have low values of LOD difference proportion. Similarly, those nonlocal genes with high values of LOD difference proportion tend to have small values of the pleiotropy test statistic. Finally, some trait pairs demonstrate low values of both the LOD difference proportion and pleiotropy test statistic. This observation suggests that, for a given local expression trait, the nonlocal trait is not mediated by the local expression trait yet it shares a pleiotropic locus. Low power to resolve multiple nearby QTL may give rise to such findings.

In comparing the *Hnf4a* plot (Figure 3.4) with the other 12 plots (Figure 3.5), we see that none of the 12 plots in Figure 3.5 closely resembles Figure 3.4. *Serinc3*, *Stk4*, *Neurl2*, and *Cdh22* are closest in appearance to the plot of *Hnf4a*. However, each of *Serinc3*, *Stk4*, *Neurl2*, and *Cdh22* has very few points with LOD difference proportion above 0.5, while *Hnf4a* has many points with LOD difference proportion above 0.5.

Now that we've examined our results from the perspective of every local gene expression trait, we turn attention to the nonlocal gene expression trait perspective. Scatter plots of LOD difference proportion values against pleiotropy vs. separate QTL test statistics for each of the 147 nonlocal expression traits demonstrated multiple patterns. Eighty-nine nonlocal genes' plots showed *Hnf4a* to have the greatest value, among the 13 putative mediators, of the LOD difference proportion statistic. In many cases, *Hnf4a*'s LOD difference proportion statistic was at least twice that of any of the other 12 local gene expression levels.

3.1.4 Discussion

Our pairwise analyses with both mediation analyses and pleiotropy tests provide additional evidence for the importance of *Hnf4a* in the biology of the Chromosome 2 hotspot in pancreatic islet cells. Our analyses, and, specifically, the test of pleiotropy vs. separate QTL, may be more useful when studying nonlocal traits that map to a hotspot yet don't show strong evidence of mediation by local expression traits. In such a setting, the pleiotropy test can, at least, provide some information about the genetic architecture at the hotspot. Specifically, our pleiotropy test may inform inferences about the number of underlying QTL in a given expression trait hotspot. Additionally, our test may limit the number of expression traits that are potential intermediates between a QTL and a specified nonlocal expression trait. This relies on the assumption that a causally intermediate (local) expression trait and a target (nonlocal) expression trait presumably share a QTL. On the other hand, mediation analyses, when they provide evidence for mediation of a nonlocal trait by a local expression trait, are possibly more valuable than the test of pleiotropy vs. separate QTL, since the mediation analyses identify precisely the intermediate expression trait.

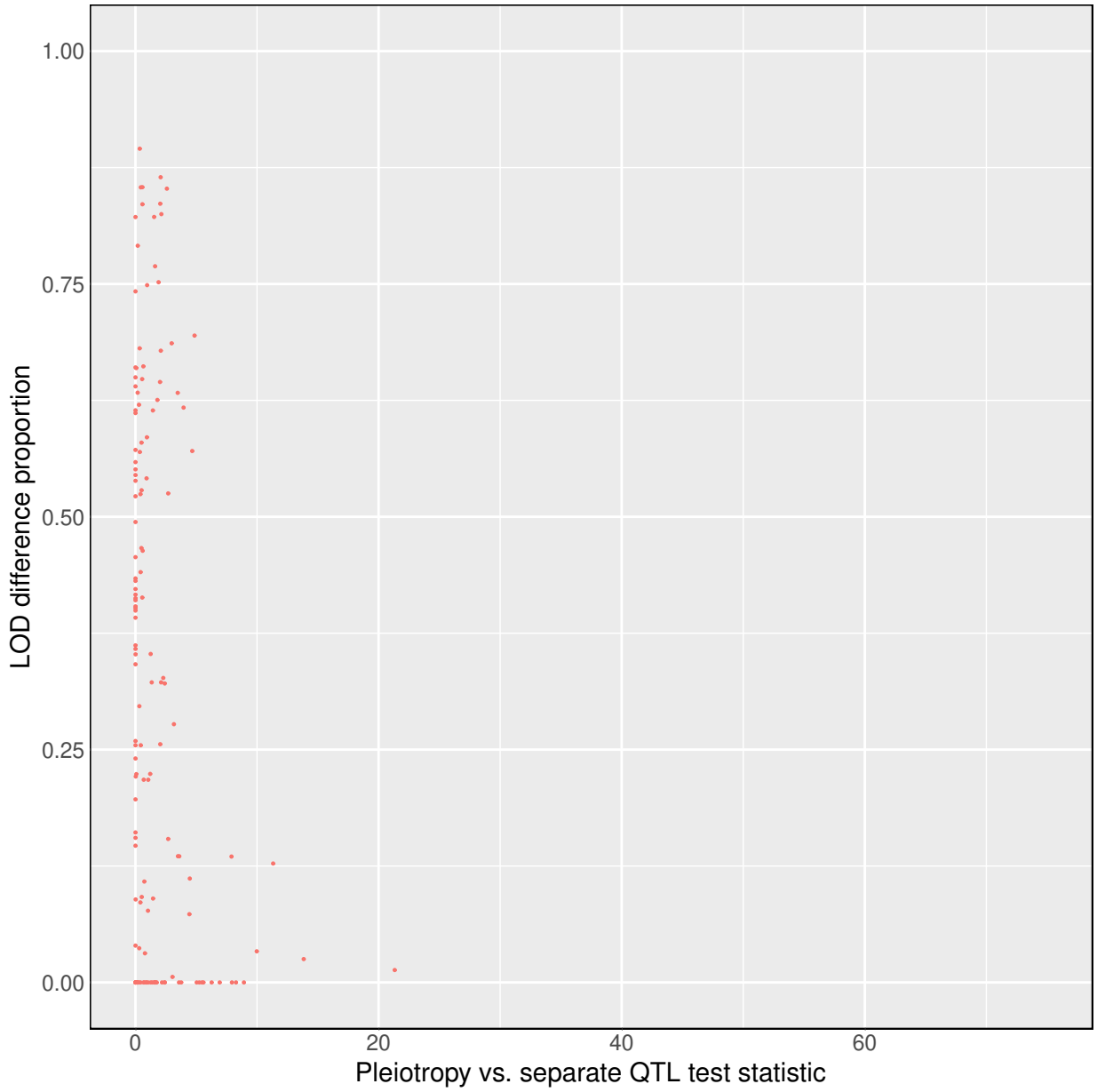


Figure 3.4: Scatter plot of LOD difference proportion against pleiotropy vs. separate QTL test statistic for 147 pairs of traits. Each pair includes *Hnf4a* and one of the nonlocal gene expression traits that map to the Chromosome 2 hotspot.

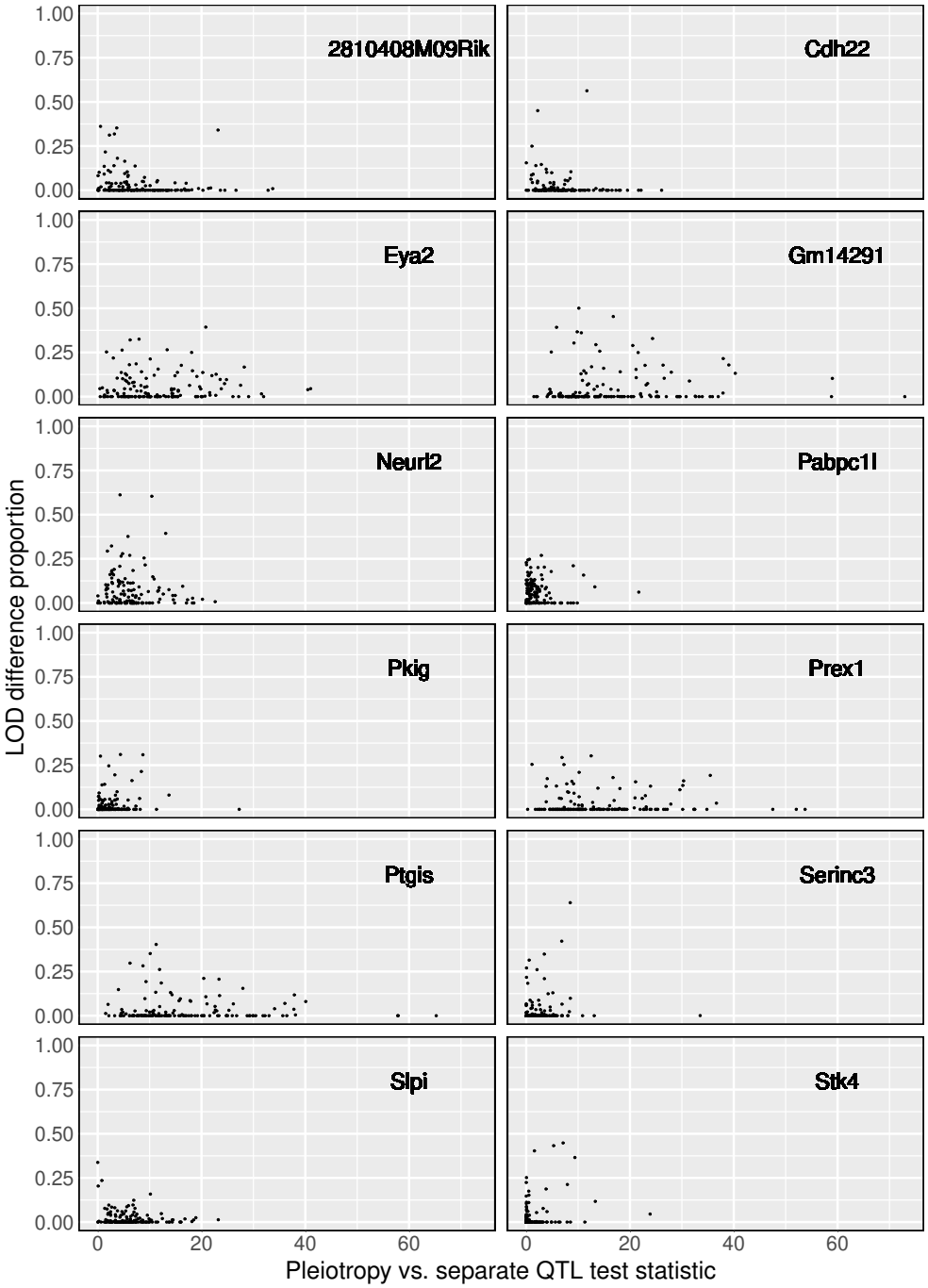
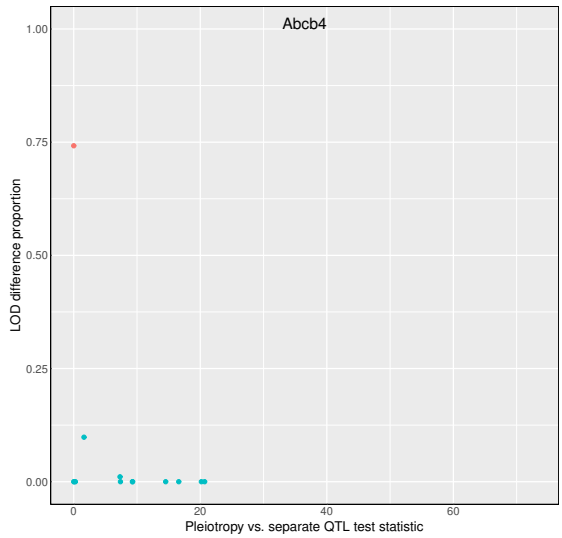
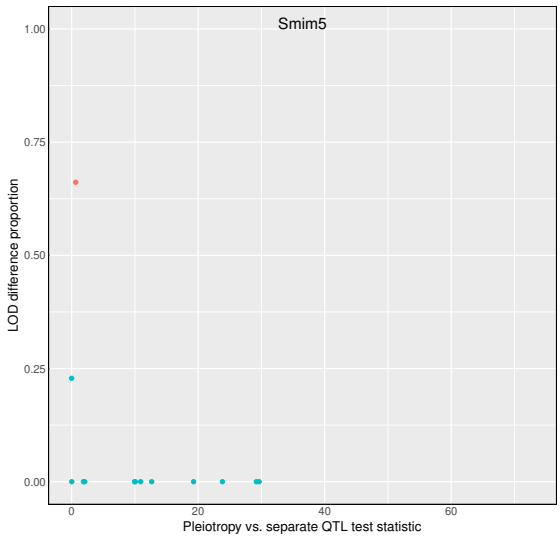


Figure 3.5: Scatter plots of LOD difference proportion against pleiotropy vs. separate QTL test statistic with 147 pairs of traits per panel. Each panel includes a local gene expression trait (per the label in the upper right quadrant) and one of the 147 nonlocal gene expression traits that map to the Chromosome 2 hotspot.

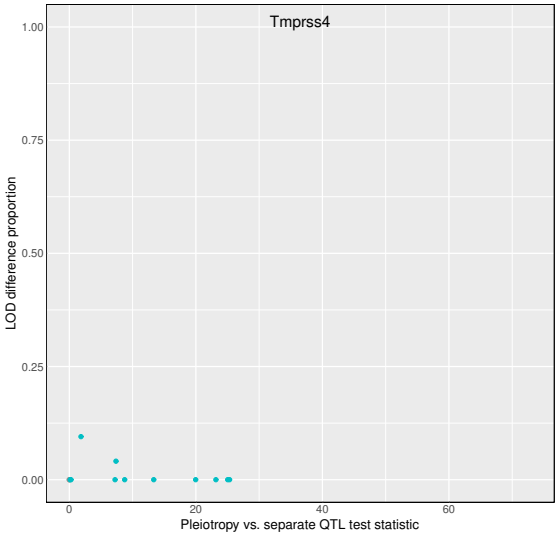
Figure 3.6: Scatter plots for four nonlocal expression traits. Each plot features 13 points, one for each local gene expression trait. The vertical axis denotes LOD difference proportion values, while the horizontal axis corresponds to pleiotropy test statistics. Blue points represent the pairing with local gene expression trait *Hnf4a*. Red points represent the other 12 local gene expression traits.



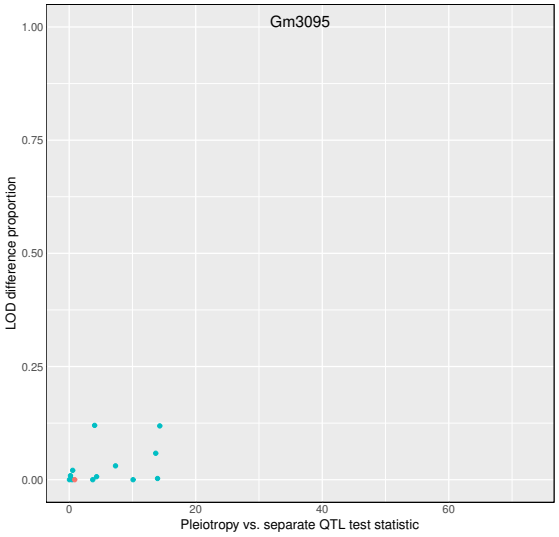
(a) *Hnf4a* transcript levels mediate *Abcb4* transcript levels. The high LOD difference proportion and the very small pleiotropy test statistic together provide evidence that *Hnf4a* mediates *Abcb4* transcript levels.



(b) *Hnf4a* transcript levels mediate *Smim5* transcript levels. The high LOD difference proportion and the very small pleiotropy test statistic together provide evidence that *Hnf4a* mediates *Smim5* transcript levels.



(c) *Tmprss4* transcript levels are not mediated by any of the 13 local gene expression traits. However, several local gene expression traits arise from separate QTL, as evidenced by their large (greater than 5) values of the pleiotropy test statistic.



(d) *Gm3095* transcript levels are not mediated by any of the 13 local genes. The tests of pleiotropy demonstrate evidence of separate QTL for some local expression traits when paired with *Gm3095*.

We recommend using both tests of pleiotropy vs separate QTL and mediation analyses when dissecting an expression trait QTL hotspot. In practice, mediation analyses, which involve a collection of univariate regressions at a small set of pre-specified markers, are less computationally intense than tests of pleiotropy vs. separate QTL, since the latter requires a two-dimensional QTL scan over the region of interest. This two-dimensional QTL scan often involves fitting more than ten thousand bivariate regression models. In light of the computational cost of testing pleiotropy vs separate QTL, future researchers may wish to use it as a follow-up to mediation analyses when examining expression trait hotspots.

Future research may investigate the use of polygenic random effects in the statistical models for mediation analysis. Additional methodological questions include approaches for declaring significant a mediation LOD difference proportion and consideration of other possible measures and scales of extent of mediation. Additionally, future researchers may wish to consider biological models that contain two mediators.

The social and health sciences have witnessed much methods research in mediation analysis. The field of statistical genetics has not fully adopted these strategies yet, but, given the nature of current and future data, many opportunities exist for translation of approaches from causal inference in epidemiology to systems genetics. For example, VanderWeele (2015) contains detailed discussions of many methods issues that arose in mediation analyses in epidemiology studies.

3.2 Power analyses

3.2.1 Introduction

The goal of this section is to characterize the statistical power of our pleiotropy test under a variety of conditions by studying a real data set. We examine pancreatic islet expression traits from the Keller et al. (2018) data. As in chapters 2 and 3A, we test only two traits at a time. Because we've chosen local expression traits in our analysis, we both know where each trait's true QTL location (approximately), and we anticipate that each trait has a unique QTL that is distinct from QTL for other local expression traits. This design thus provides opportunities to study statistical power for our test.

We anticipate that inter-locus distance, univariate QTL strength, and correlation of founder allele effects patterns are three factors that contribute to power for our test. Specifically, we expect that greater inter-locus distance, greater univariate LOD scores, and less similar founder allele effects patterns correspond to greater statistical power to detect two separate QTL.

We use pancreatic islet gene expression traits from a publicly available data set, which Keller et al.

(2018) first collected, analyzed, and shared. We examine a collection of 80 local traits on Chromosome 19 and perform our test for pleiotropy on pairs of traits. We also examine pairwise relationships among gene expression traits to characterize the impacts of univariate LOD score, inter-locus distance, and similarity of founder allele effects patterns on pleiotropy test statistics.

3.2.2 Methods

We analyzed data from 378 Diversity Outbred mice (Keller et al., 2018). Keller et al. (2018) genotyped tail biopsies with the GigaMUGA microarray (Morgan et al., 2016). They used RNA sequencing to measure genome-wide pancreatic islet cell gene expression for each mouse at the time of sacrifice (Keller et al., 2018). They shared these data, together with inferred founder allele probabilities, on the Data Dryad site (https://datadryad.org/resource/doi:10.5061/dryad.pj105). We performed analyses with the R statistical computing environment (R Core Team, 2018) and the packages `qtl2` (K. W. Broman, D. M. Gatti, et al., 2019) and `qtl2pleio` (Boehm, 2018b).

We study below 80 Chromosome 19 local expression QTL and their corresponding transcript levels. We define a local expression QTL to be an expression QTL that is on the same chromosome as the gene itself. For example, the *Asah2* gene is located on Chromosome 19 and its transcript levels have an expression QTL on Chromosome 19 (Table 3.2). Thus, we term the Chromosome 19 *Asah2* expression QTL a local expression QTL.

We choose to focus on local expression QTL, while ignoring nonlocal expression QTL, because we know, approximately, the true locations for local expression QTL. That is, a local expression QTL is near the corresponding gene position. Additionally, we expect that a given local expression QTL affects only one local expression trait. In our example above, we expect that the *Asah2* expression QTL is near the *Asah2* gene position and that no other local expression traits map to it.

Our design involves selection of a set of "anchor" expression traits. Gene *Asah2* is located near the center of Chromosome 19 and has a very strong local expression QTL (Table 3.2). We chose it as our first "anchor" gene expression trait. To diversify our collection of anchor genes, we chose three additional expression traits with local expression QTL. These three are *Lipo1*, *Lipo2*, and *4933413C19Rik* (Table 3.2). Together, the four anchor genes represent a variety of strong local expression trait LOD scores (from 60 to 101) and demonstrate modest variability in their founder allele effects (Table 3.3). All four anchor genes are located near the middle of Chromosome 19 (Table 3.2).

We identified a set of 76 non-anchor local expression traits that map to the 20-Mb region centered on

the peak for *Asah2*, at 32.1 Mb. Each trait among the 76 maps to Chromosome 19 with a univariate LOD score of at least 10 (Table D.2).

Table 3.2: Annotations for four anchor genes.

Gene	Start	End	QTL peak position	LOD
Asah2	31.98	32.06	32.14	101.20
Lipo1	33.52	33.76	33.67	85.46
Lipo2	33.72	33.76	33.02	77.21
4933413C19Rik	28.58	28.58	28.78	60.41

We estimated founder allele and covariate effects for every trait on Chromosome 19. For each of the 80 expression traits, we calculated fitted values for each subject with these estimated founder allele and covariate effects. We then calculated correlations between fitted values for pairs of traits. Our motivation for working with the fitted values vectors (instead of the estimated founder allele effects vectors) is that the fitted values approximately weight the allele effects by allele frequency. We anticipated that more similar two traits' founder allele effects would correspond, on average, to smaller pleiotropy test statistics. We base this expectation on findings from Macdonald and Long (2007) and King et al. (2012), who found that two traits that associate with a single pleiotropic QTL tended to have similar founder allele effects patterns for biallelic markers.

We performed two-dimensional QTL scans for $4 * 76 + \binom{4}{2} = 310$ pairs. Each pair included one of the four anchor gene expression traits and either one of 76 non-anchor gene expression traits or one of the remaining three anchor gene expression traits. Our two-dimensional QTL scan encompassed a 1000 by 1000 marker grid from 18.1 Mb to 42.5 Mb on Chromosome 19. Each scan involved fitting 1000 x 1000 = 1,000,000 models via generalized least squares. For a given ordered pair of markers, we used the bivariate linear mixed effects model and methods of Chapter 2. These methods are implemented in the R package `qtl2pleio` (Boehm, 2018b).

3.2.3 Results

All four anchor traits demonstrate strong PWK ("G") allele effects. *Lipo2* and *Asah2* have similar patterns among allele effects (at their respective QTL peaks) (Table 3.3).

Each anchor gene has its own panel in Figure 3.7. Along the horizontal axis is Chromosome 19 position. The vertical axis is for pleiotropy test statistics. Each point corresponds to a local gene expression trait (paired with the appropriate anchor gene expression trait). Point color corresponds to the local gene's

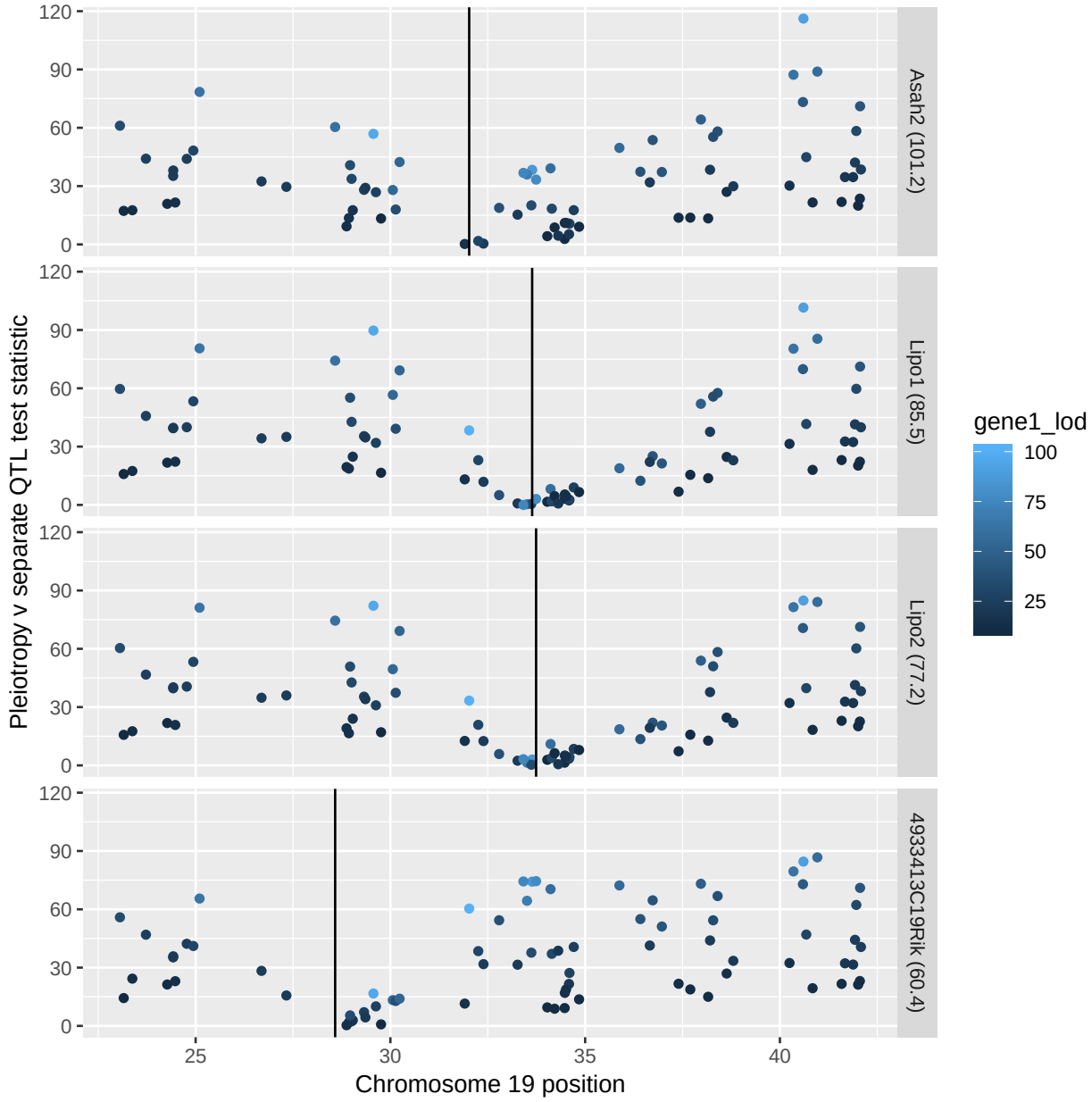


Figure 3.7: Each anchor gene has its own panel. Along the horizontal axis is Chromosome 19 position. The vertical axis is for pleiotropy test statistic value. Each point corresponds to a local gene expression trait (paired with the anchor gene). Point color corresponds to the local gene's univariate LOD score, with lighter shades of blue denoting greater values of univariate LOD score. Vertical black bar denotes the anchor gene's position on Chromosome 19. All four panels reveal that points further from the anchor gene tend to show greater test statistic values. Additionally, the *Lipo1* and *Lipo2* panels offer an opportunity to compare the impact of anchor gene univariate LOD score on pleiotropy test statistic values.

Table 3.3: Founder allele effect estimates at Chromosome 19 QTL peak position.

Gene	Founder allele	Effect	Standard error
Asah2	A	-0.96	0.17
	B	1.01	0.19
	C	0.14	0.17
	D	-1.16	0.17
	E	1.05	0.16
	F	-0.61	0.20
	G	1.81	0.16
	H	-0.18	0.18
Lipo1	A	0.29	0.18
	B	0.13	0.21
	C	0.28	0.20
	D	0.23	0.19
	E	-0.17	0.18
	F	-0.28	0.21
	G	2.55	0.19
	H	-0.72	0.19
Lipo2	A	-0.10	0.18
	B	-0.28	0.23
	C	0.00	0.20
	D	0.01	0.18
	E	-0.77	0.17
	F	-0.89	0.22
	G	2.65	0.18
	H	-0.70	0.20
4933413C19Rik	A	0.29	0.23
	B	0.76	0.24
	C	0.81	0.21
	D	0.49	0.24
	E	0.67	0.20
	F	-0.53	0.22
	G	-1.65	0.18
	H	0.67	0.21

univariate LOD score, with lighter shades of blue denoting greater values of univariate LOD score. Vertical black bar denotes the anchor gene's position on Chromosome 19. All four panels reveal that points further from the anchor gene tend to show greater test statistic values. Additionally, because of their nearly identical positions, the *Lipo1* and *Lipo2* panels offer an opportunity to compare the impact of anchor gene univariate LOD score on pleiotropy test statistics. The difference in univariate LOD scores is modest, 85.5 for *Lipo1* vs. 77.2 for *Lipo2*. The effect of univariate LOD scores is most apparent when examining local expression traits that have strong univariate LOD scores. For example, near 32 Mb is a point with light blue color. The light blue indicates that it has a high univariate LOD score. *Lipo2*, which has the lower univariate LOD score, shows a pleiotropy test statistic just over 30 when paired with this local expression trait. *Lipo1*

demonstrates a greater test statistic value, about 40, when paired with the same local expression trait.

Similarly, a local expression trait near 29 Mb has a strong univariate LOD score (near 100). We see that its pleiotropy test statistic values in the *Lipo1* and *Lipo2* panels reflect the difference in univariate LOD for the two anchor gene expression traits. The pleiotropy test statistic value is greater when paired with *Lipo1*, the anchor gene expression trait with greater univariate LOD (compared to *Lipo2*).

In comparing the plot for *Asah2* with those of *Lipo1*, *Lipo2*, and *4933413C19Rik*, we also see that *Asah2*, with the largest (101.2) of the four univariate LOD scores, demonstrates the steepest ascent of points as interlocus distance increases.

Analyses for all four anchor gene expression traits demonstrate that greater univariate LOD scores tend to correspond to greater values of the pleiotropy test statistic (Figure 3.8). The coloring of points reflects interlocus distance from the anchor gene. All four panels reveal a common pattern in the coloring of points. For a given univariate LOD value, those genes with greater interlocus distance tend to have greater values of the pleiotropy test statistic. Anchor gene 4933413C19Rik, due to its position on Chromosome 19, has multiple light blue points. Nearly all of these light blue points have large pleiotropy test statistics relative to their univariate LOD scores.

Figure 3.9 features four panels, one for each anchor gene. Each point corresponds to a pairing between the specified anchor and one of the 79 other gene expression traits. None of the four panels reveals a strong relationship between fitted values correlation and pleiotropy test statistics. There may be weak evidence for a downward trend in pleiotropy test statistics at larger values of fitted values correlation, as demonstrated by *Asah2*, *Lipo1*, and *Lipo2*.

3.2.4 Discussion

Our goal for this study was to characterize the impacts of univariate LOD score, inter-locus distance, and founder allele effects pattern similarities on pleiotropy test statistic values. Our study design, in which we examined 310 pairs of local gene expression traits on Chromosome 19, allowed us to interrogate both the effects of univariate association strength and the effects of inter-locus distance. We found that stronger univariate associations and greater inter-locus distances correspond to greater pleiotropy test statistic values (Figures 3.7 and 3.8). We expected these trends based on our simulation studies in Chapter 2.

Figure 3.9 revealed no strong relationship between fitted values correlations and pleiotropy test statistics. However, close examination of Figure 3.9 reveals the possibility that there is an interaction between 1) fitted values correlations and 2) univariate LOD scores. In every panel, those expression traits with stronger

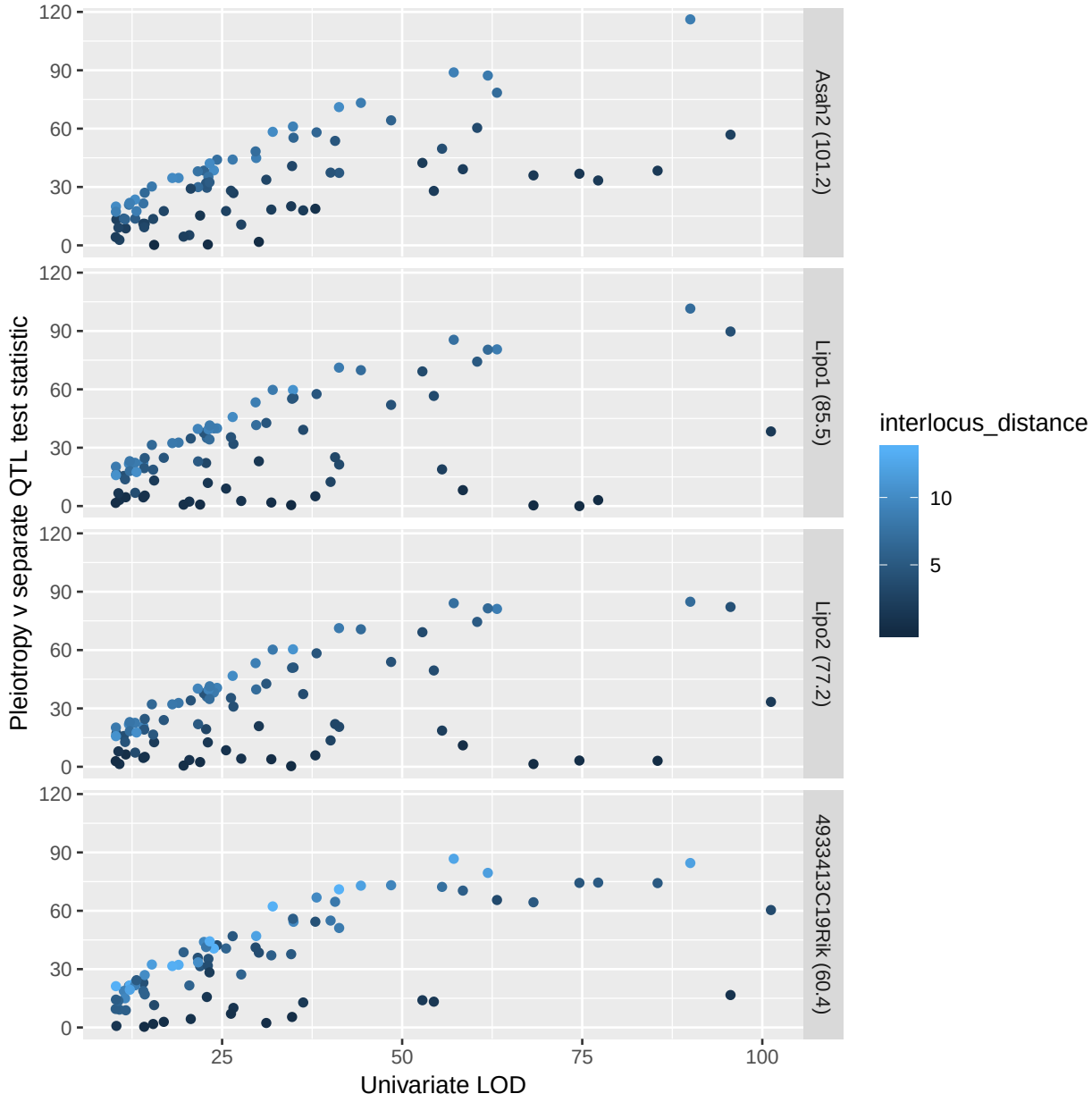


Figure 3.8: Vertical axis denotes pleiotropy test statistic value, while horizontal axis denotes univariate LOD score. Each point corresponds to a single gene expression trait. Panels correspond to the anchor gene expression trait. The pleiotropy test statistics correspond to analyses involving a single gene expression trait and the specified anchor gene expression trait.

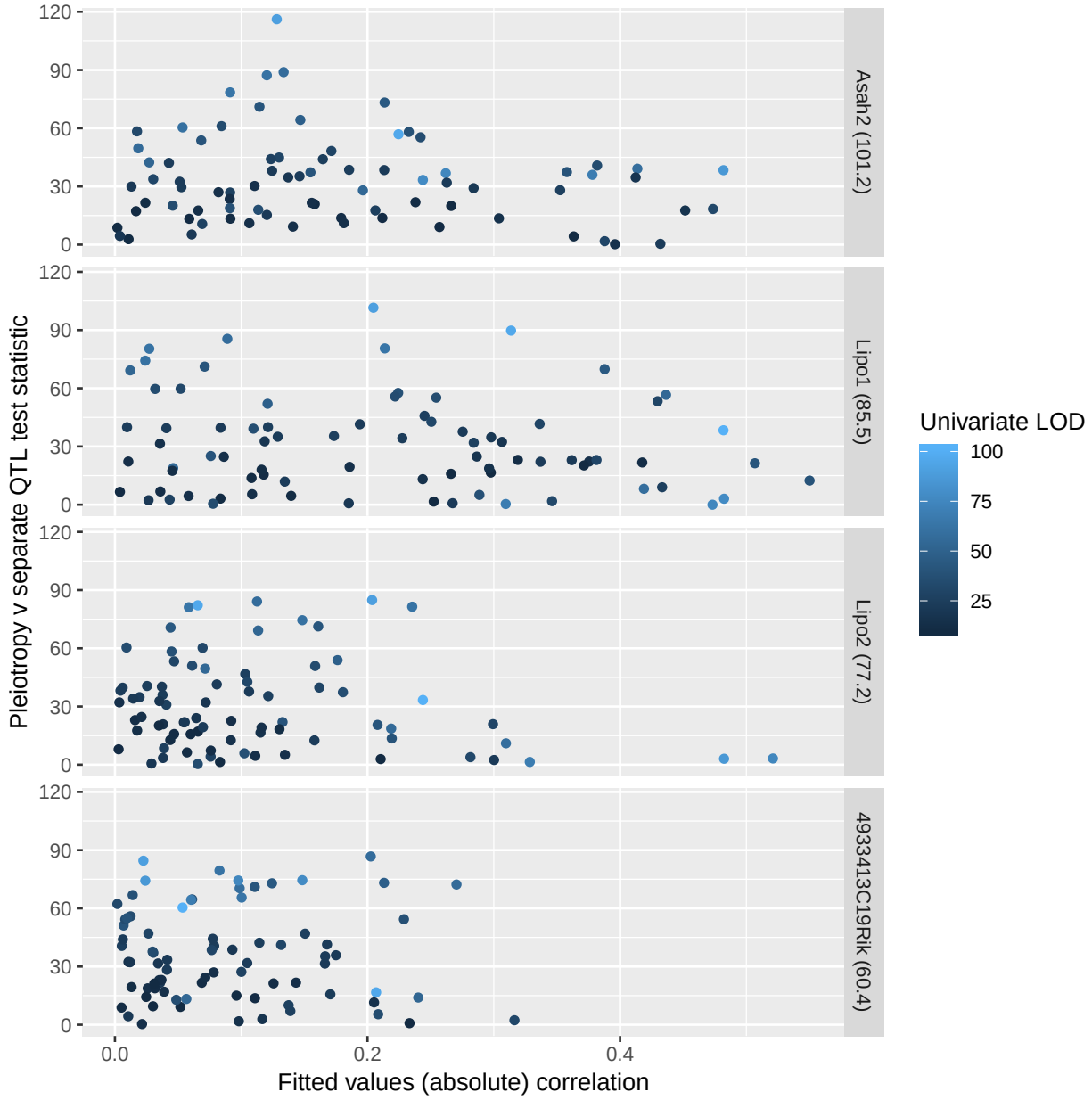


Figure 3.9: Vertical axis denotes the pleiotropy test statistic value, and horizontal axis indicates absolute value of the correlation between vectors of fitted values. Each point corresponds to a pairing between the specified anchor expression trait and one of the 79 other expression traits.

univariate associations tend to have steeper slopes between the conditional mean pleiotropy test statistic values and fitted values correlations. The plots weakly suggest that, at greater univariate LOD values, there is a greater (negative) relationship between fitted values correlation and pleiotropy test statistic value.

We anticipated that more similar founder allele effects patterns would correspond to smaller values for the pleiotropy test statistic, when holding other factors constant. As we stated above, Macdonald and Long (2007) and King et al. (2012) argued that, for biallelic markers, two pleiotropic traits should have similar founder allele effects patterns. In our setting, it's unclear whether the markers are biallelic in the collection of eight founder lines.

We've demonstrated strong evidence in support of the roles of 1) univariate QTL LOD scores and 2) interlocus distances impacting pleiotropy test statistic values. Greater univariate QTL scores and greater interlocus distance lead to greater pleiotropy test statistics. Future research may clarify the impoact of founder allele effects patterns on pleiotropy test statistics. The fact that all four anchor traits had strong PWK effects (Table 3.3) limited our ability to fully define the impact of allele effects patterns on our test statistics.

Throughout this study, we elected to use test statistic values rather than p-values as our measure of evidence supporting the separate QTL hypothesis. The primary reason for doing this is to avoid the computationally costly bootstrap sampling and two-dimensional QTL scans that we would need to get bootstrap p-values.

We share our analysis `R` code (R Core Team, 2018) as a `git` repository at this URL: https://github.com/fboehm/keller-2018-chr19-power.

3.3 Microbiome case study

3.3.1 Introduction

Recent technological innovations have fueled exploration of ecological relationships between gut microbiota and their hosts. Advances in mass spectrometry experimental methods have enabled high-throughput quantification of lipid levels and protein concentrations. These developments, when coupled with experiments to quantify gut microbiota, have the potential to uncover new microbiome-host interactions. Such discoveries would lead to a more nuanced understanding of organismal biology and health implications of the gut microbiome.

Previous research demonstrated that hosts and gut microbiota communicate with each other via small

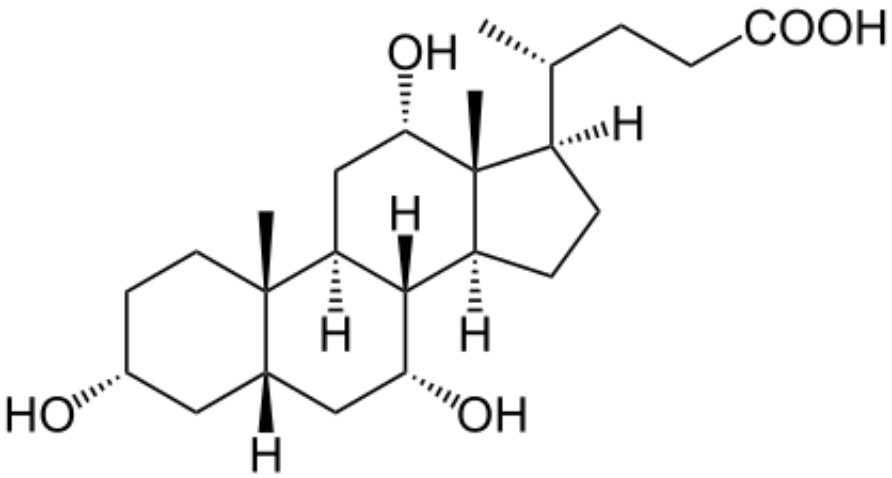


Figure 3.10: Cholic acid chemical structure. From Wikipedia, https://upload.wikimedia.org/wikipedia/commons/2/21/Chols%C3%A4ure.svg.

molecule metabolites, including bile acids. The liver uses cholesterol to synthesize an array of bile acids. These bile acids are secreted into the gut, where bacteria chemically transform some bile acids. One function of bile acids is to aid digestion by emulsification of dietary fats. While a portion of bile acids are lost in the feces, much of the secreted bile acids is reabsorbed in the distal gut. Finally, reabsorbed bile acids are transported via the circulatory system back to the liver.

Below, we use our pleiotropy test to identify a pleiotropic QTL that affects both the abundance level of a group of bacteria in the distal gut and plasma cholic acid levels in the host. While many questions remain after our investigation, our identification of a single pleiotropic locus that affects these two phenotypes is an important preliminary step for further investigations.

3.3.2 Methods

As we describe in Kemis et al. (submitted), we analyzed data from 384 Diversity Outbred mice. Keller et al. (2018) analyzed data from many of these same mice. Specifically, we examined two microbiome-related phenotypes, 1. plasma cholic acid and 2. *Turicibacter* abundance in the distal gut. Both traits map to Chromosome 8 (4.3 Mb and 5.7 Mb, respectively) in univariate QTL scans (Table 3.4). Our methods proceed in three ordered steps. First, we collect the biological samples. Then, we prepare and process the data. Finally, we perform our pleiotropy test.

Our sample collection involves three steps, all of which are done at the time of sacrifice (at age 22 weeks). We obtained fecal samples from all mice immediately after a four-hour fast. We also obtained blood plasma

from every mouse at the time of sacrifice. Tail clippings provided DNA for host genotyping.

Data preparation and processing involves two parallel sets of procedures. In one set of procedures, we obtain genotype probabilities for all markers for all mice. In the other set of procedures, we obtain phenotypes. Obtaining phenotypes in this study involves processing both fecal samples and blood samples.

As the first ordered step in inferring genotype probabilities, we extracted DNA from tail clippings and subjected it to SNP genotyping with the GigaMUGA microarray (Morgan et al., 2015). We inferred 36-state genotype probabilities from SNP genotype calls for every (autosomal) marker and every mouse. We used a hidden Markov model strategy developed by K. W. Broman (2012a) and K. W. Broman (2012b) and implemented in the R package `qtl2` (K. W. Broman, D. M. Gatti, et al., 2019). We treated these inferred genotype probabilities as known quantities in the analyses below. Lastly, we calculated founder allele dosages by summing the appropriate genotype probabilities.

Phenotype processing proceeds in parallel by separately processing fecal samples and blood samples. Processing of fecal samples to obtain microbial taxa counts involves multiple experimental and computational steps. We extracted microbiome DNA from fecal samples and subjected it to 16S ribosomal RNA gene sequencing to infer abundances of microbial taxa. Demultiplexed, paired-end FASTQ files resulted from the sequencing. We used the QIIME2 (version 2018.4) software package (Bolyen et al., 2018) for quality control and processing of sequence data. We applied the DADA2 software package (Callahan et al., 2016) to denoise sequencing reads and to identify de novo sub-operational taxonomic units. We aligned sequence variants with the software package mafft (Katoh and Standley, 2013). After FastTree-based phylogeny reconstruction (Price, Dehal, and Arkin, 2010), we assigned taxonomic classifications with classify-sklearn against the Greengenes OTUs reference sequences (DeSantis et al., 2006). We normalized sequencing data with cumulative sum scaling with MetagenomeSeq (Paulson, Pop, and Bravo, 2013). We limited study of microbiota-derived traits to those that we detected in at least 20% of subjects. The *Turicibacter* abundance trait is one element of the resulting core measurable microbiota.

Contemporaneously with the fecal sample processing, we subjected plasma samples to multiple experimental and computational steps to obtain traits for QTL mapping. After removing soluble proteins from each plasma sample, we analyzed each sample by mass spectrometry to measure abundances of pre-specified bile acids. Finally, we processed and normalized plasma bile acid measurements to obtain QTL analysis traits.

We then had the needed inputs for QTL mapping. We performed univariate QTL analyses for our microbiome-related traits. For each marker and each univariate phenotype we fitted a linear mixed effects

model (with a polygenic random effect), as implemented in the R package `qtl2` (K. W. Broman, D. M. Gatti, et al., 2019). We treated founder allele effects as additive and neglected possible interactions. We incorporated four covariates into the model: sex and three binary indicators for wave number.

We then calculated LOD values comparing the univariate models' log-likelihoods at each marker to the log-likelihood of the model without founder allele dosages. In summarizing LOD peak results, we identified a small region on Chromosome 8 that contains LOD peaks for both *Turicibacter* abundance and plasma cholic acid levels.

We use the R package `qtl2` to estimate founder allele effects at Chromosome 8 markers with the univariate linear mixed effects model.

We then had the inputs needed for our pleiotropy test: founder allele probabilities for each mouse at every marker and two traits that map to a single genomic region. We performed a test of pleiotropy vs. separate QTL for our two traits, *Turicibacter* abundance and plasma cholic acid levels, using the methods of Chapter 2. To determine statistical significance, we performed a parametric bootstrap analysis to acquire 1000 bootstrap samples. With the collection of 1000 bootstrap test statistic values, we calculated a bootstrap p-value as the proportion of the 1000 test statistics that were at least as large as the true test statistic. Calculation of a bootstrap p-value is the last step in our hypothesis test for pleiotropy vs. separate QTL.

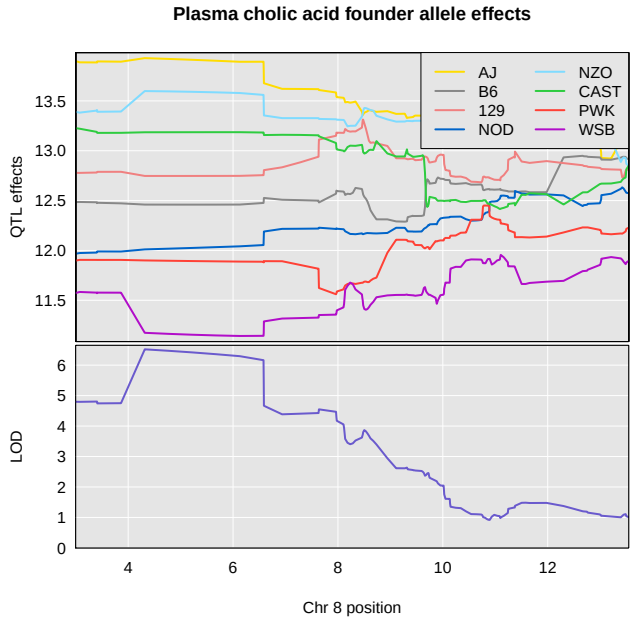
3.3.3 Results

phenotype	chromosome	position	LOD
plasma cholic acid	1	91.61	5.37
	3	40.53	5.80
	7	122.19	6.83
	8	4.32	6.52
	12	16.60	5.24
Turicibacter abundance	2	17.20	5.40
	8	5.68	7.22
	12	76.34	5.17

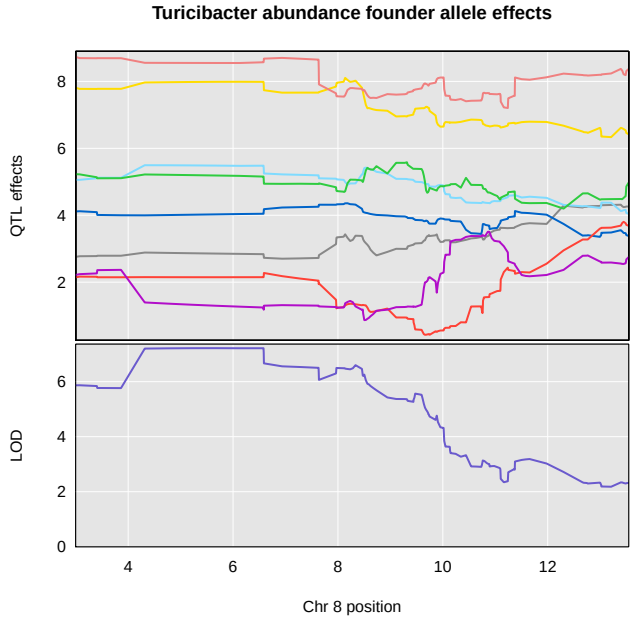
Table 3.4: Genome-wide LOD peaks greater than 5 for plasma cholic acid levels and *Turicibacter* abundance. Both traits map to approximately the same region on Chromosome 8. Chromosome positions are in Mb units.

Both plasma cholic acid and *Turicibacter* abundance demonstrate multiple univariate QTL across the genome. Each trait also maps to a 1.4-Mb region on Chromosome 8. Because these two traits map to a single genomic region, we elected to study them further.

Univariate LOD plots for the two traits reveal broad LOD peaks on Chromosome 8 (Figure 3.11). Plasma



(a) Plasma cholic acid founder allele effects over the scan interval.



(b) *Turicibacter* abundance founder allele effects over the scan interval.

Figure 3.11: Founder allele effects and LOD plots for our two traits over the Chromosome 8 region. Both traits demonstrate QTL peaks between 4 and 8 Mb and similar patterns of founder allele effects.

cholic acid's peak starts about 4 Mb and extends to approximately 6.5 Mb. *Turicibacter* abundance has an even broader peak, starting about 4 Mb and extending to 7.5 Mb. Although the two traits' peak position point estimates differ by 1.4 Mb, the univariate LOD plots demonstrate highly similar patterns for the two traits.

It is unclear from the univariate LOD plots whether there are two nearby QTLs that both affect *Turicibacter* abundance (Figure 3.11). Consistent with the possibility of two peaks is the observation that the *Turicibacter* peak extends further to the right than does the plasma cholic acid peak.

Founder allele effects plots for the two traits demonstrate similarities in the relative magnitudes of the eight effects. Specifically, the plasma cholic acid founder allele effects, at 5Mb, has AJ, NZO, CAST, and 129 above average, while WSB is well below average. Similarly, *Turicibacter* abundance has founder allele effects at 5Mb with AJ and 129 above average and WSB below average.

To explore relationships between the two traits, we created a scatter plot. The scatter plot of plasma cholic acid levels and *Turicibacter* abundance levels demonstrates modest correlation (Pearson correlation coefficient = 0.5) (Figure 3.12). *Turicibacter* was unobserved in fecal samples from 192 of 384 mice. After our data processing steps (described above), these 192 mice all have the smallest possible value of the *Turicibacter* abundance trait. Because these 192 mice possess a range of plasma cholic acid values, they appear as a vertical "line" on the left-hand side of the scatter plot.

Once we examined the univariate QTL scan results, we had the needed inputs for a bivariate scan; namely, two traits that map to a single genomic region and genotype probabilities for all mice and all markers. We performed a bivariate QTL scan over a 180-marker region on Chromosome 8. The profile LOD plot over our two-dimensional scan region reveals broad peaks for both traits' profile LODs (Figure 3.13). Plasma cholic acid has a peak that begins about 4 Mb and extends to approximately 6.5 Mb. *Turicibacter* abundance has a broader peak, with endpoints near 4 Mb and 7.5 Mb. Here, but not in the univariate LOD plots, one sees what appears to be a second peak in the *Turicibacter* abundance phenotype that begins near 7.5 Mb and ends near 8.7 Mb.

We used the R package `qtl2pleio` (Boehm, 2018b) to calculate the likelihood ratio test statistic, $\Lambda = 0.45$. We determined the bootstrap p-value to be 0.531 with 1000 bootstrap samples. Thus, we failed to reject the null hypothesis of pleiotropy.

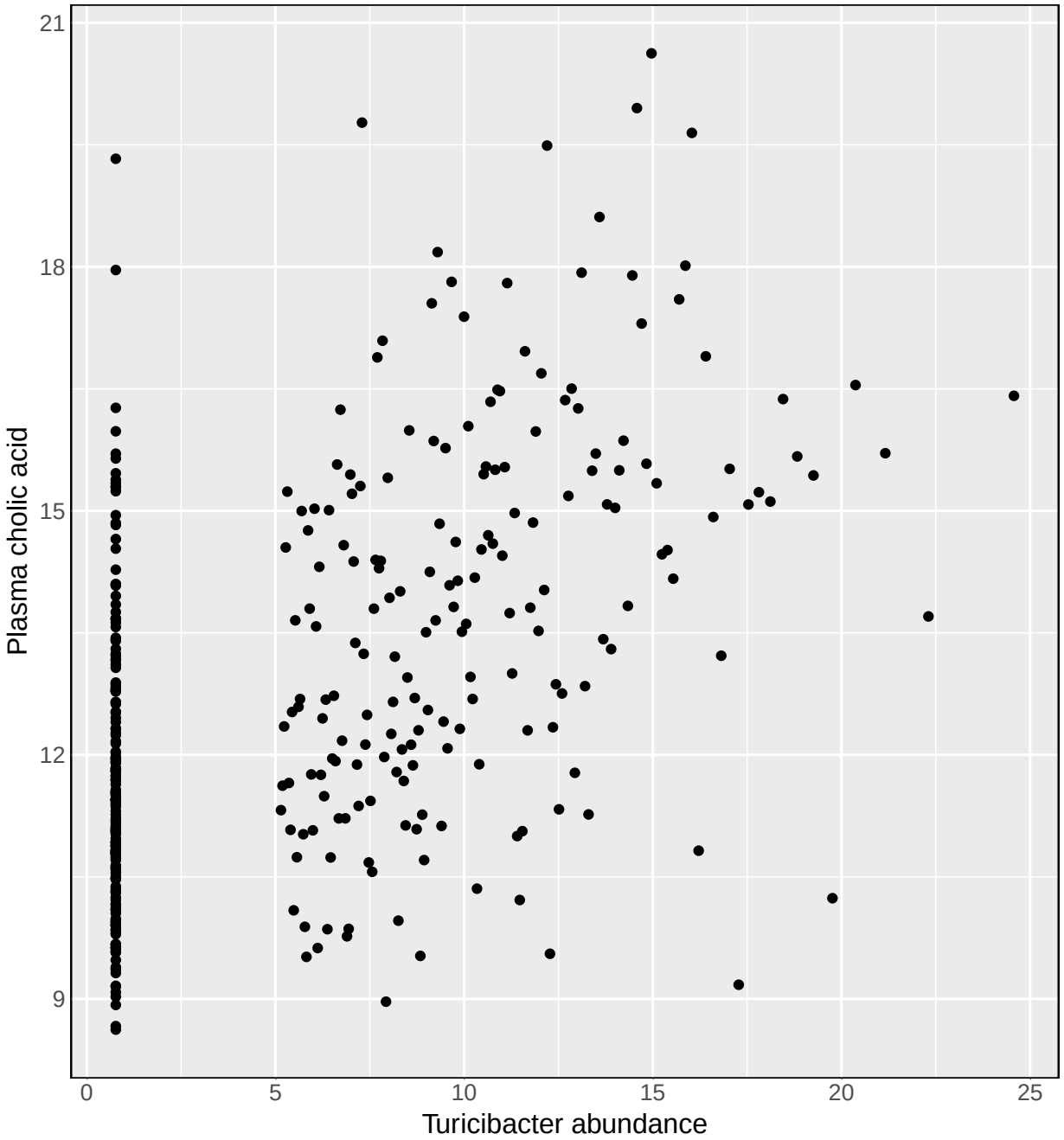


Figure 3.12: Scatter plot of plasma cholic acid levels against *Turicibacter* abundance

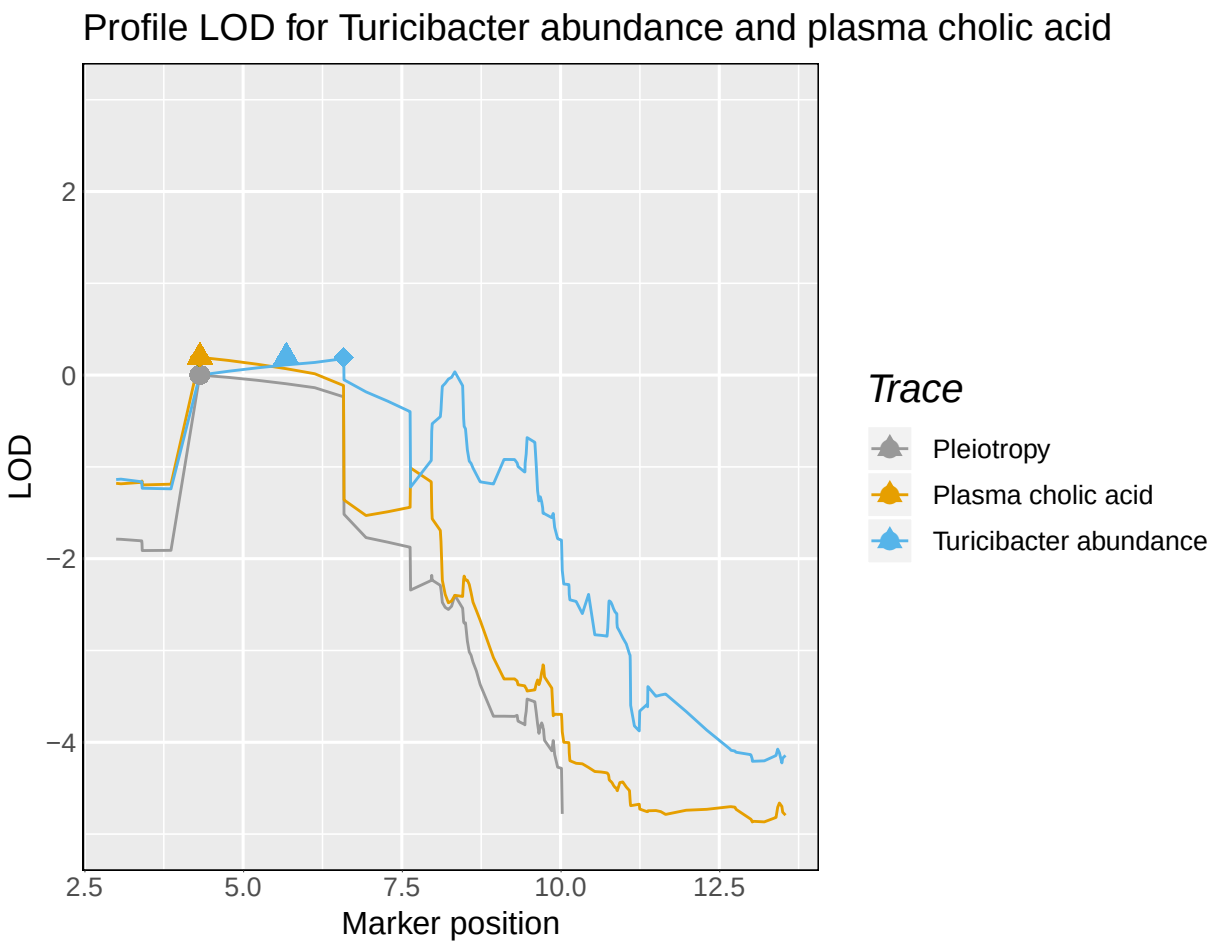


Figure 3.13: Profile LODs for *Turicibacter* abundance and plasma cholic acid levels.

3.3.4 Discussion

Results from our analyses above, including the pleiotropy test, are consistent with a single pleiotropic QTL affecting both *Turicibacter* abundance and plasma cholic acid levels. The similarity of founder allele effects patterns (Figure 3.11) is also consistent with a single pleiotropic QTL (King et al., 2012; Macdonald and Long, 2007).

While our results indicate presence of a single pleiotropic QTL, they leave many questions unanswered. For example, it would be interesting to know if one trait mediates the effects of the QTL on the other trait. Statistical mediation analyses (Chick et al., 2016), like those discussed earlier in Chapter 3, may clarify this issue. Both possibilities are biologically plausible: 1. that plasma cholic acid levels affect *Turicibacter* abundance and 2. that *Turicibacter* abundance affects plasma cholic acid levels.

A consideration of cholic acid biology illuminates this biological plausibility. Plasma cholic acid levels, which are related to absorption rates of cholic acid from the distal gut, may be influenced by cholic acid production rates in the liver. This, in turn, might affect gut colonization by *Turicibacter*. Alternatively, the second possibility above is also biologically plausible, because, for example, *Turicibacter* may affect plasma cholic acid levels by chemically transforming cholic acid while it's in the gut. Mediation analyses and causal model selection tests (Neto et al., 2013) may clarify the biological interactions that affect plasma cholic acid levels and *Turicibacter* abundance in the distal gut.

A second issue that arose in our analysis is the possibility of two nearby peaks for the *Turicibacter* abundance trait. While the univariate LOD plot doesn't indicate two *Turicibacter* peaks (Figure 3.11), the profile LOD plot suggests that there is a second *Turicibacter* abundance peak from approximately 8 to 8.7 Mb. We might clarify this possibility by examining statistical models that explicitly incorporate two QTL for *Turicibacter* while constraining plasma cholic acid levels to a single QTL (Schadt et al., 2005). Distinguishing whether there are two QTL that affect *Turicibacter* abundance offers insights into the trait's genetic architecture and, by suggesting candidate causal genes, may inform subsequent experiments.

In this microbiome study, we've illustrated another scientific application of our pleiotropy test. We plan to perform additional analyses, including mediation studies and causal model selection tests, with these data. Our pleiotropy test tells us that the data are consistent with presence of a single QTL affecting both *Turicibacter* abundance and plasma cholic acid levels. This insight serves as a starting point for follow-up studies into the directions of intertwined causal relationships between plasma cholic acid and *Turicibacter* abundance.

Chapter 4

Computing vignettes

4.1 Pleiotropy testing

Our setting involves a pair of traits, Y_1 and Y_2, each of which individually (univariately) maps to a single genomic region. Y_1 and Y_2 are both measured on the same subjects. The exact definition of a genomic region is imprecise; in practice, it may be as large as 4 or 5 Mb. We seek to distinguish whether Y_1 and Y_2 associations (in the genomic region of interest) arise due to a single QTL or whether there are two two distinct loci, each of which associates with exactly one of the two traits. We recognize that more complicated association patterns are possible, but we neglect them in this test.

4.1.1 Installing `qtl2pleio`

We install `qtl2pleio` from github via the `devtools` R package, which is available from CRAN.

To install qtl2pleio, use `install_github()` from the devtools package.

```
install.packages("devtools")
devtools::install_github("fboehm/qtl2pleio")
```

You may also wish to install R/qtl2 and the `qtl2convert` package. We will use both below.

```
install.packages(c("qtl2"), repos = "http://rqtl.org/qtl2")
```

The above line only needs to be run once on a given computer (unless you wish to install a newer version of the package).

We then load the library into our R session with the `library` command:

```
library(qtl2pleio)
```

We also load the `qtl2` package with the `library` command.

```
library(qtl2)
```

4.1.2 Reading data from `qtl2data` repository on github

We'll consider the `DOex` data in the `qtl2data` repository. We'll download the DOex.zip file before calculating founder allele dosages.

```
file <- paste0("https://raw.githubusercontent.com/rqtl/", "qtl2data/master/DOex/DOex.zip")
DOex <- read_cross2(file)
```

Let's calculate the founder allele dosages from the 36-state genotype probabilities.

```
probs <- calc_genoprob(DOex)
pr <- genoprob_to_alleleprob(probs)
```

We now have an allele probabilities object stored in `pr`.

```
names(pr)
#> [1] "2" "3" "X"
dim(pr$`3`)
#> [1] 261   8 102
```

We see that `pr` is a list of 3 three-dimensional arrays - one array for each of 3 chromosomes.

We now have an allele probabilities object stored in `pr`.

```
names(pr)
#> [1] "2" "3" "X"
dim(pr$`3`)
#> [1] 261   8 102
```

We see that `pr` is a list of 3 three-dimensional arrays - one array for each of 3 chromosomes.

4.1.3 Kinship calculations

For our statistical model, we need a kinship matrix. Although we don't have genome-wide data - since we have allele probabilities for only 3 chromosomes - let's calculate a kinship matrix using "leave-one-chromosome-out". In practice, one would want to use allele probabilities from a full genome-wide set of markers.

```
kinship <- calc_kinship(probs = pr, type = "loco")
```

```
str(kinship)
#> List of 3
#>  £ 2: num [1:261, 1:261] 0.6934 0.0705 0.2356 0.0558 0.0513 ...
#>   ..- attr(*, "n_pos")= int 195
#>   ..- attr(*, "dimnames")=List of 2
#>   .. ..£ : chr [1:261] "1" "4" "5" "6" ...
#>   .. ..£ : chr [1:261] "1" "4" "5" "6" ...
#>  £ 3: num [1:261, 1:261] 0.6662 0.0647 0.2024 0.1129 0.0772 ...
#>   ..- attr(*, "n_pos")= int 220
#>   ..- attr(*, "dimnames")=List of 2
#>   .. ..£ : chr [1:261] "1" "4" "5" "6" ...
#>   .. ..£ : chr [1:261] "1" "4" "5" "6" ...
#>  £ X: num [1:261, 1:261] 0.4871 0.0831 0.1953 0.1043 0.1125 ...
#>   ..- attr(*, "n_pos")= int 229
#>   ..- attr(*, "dimnames")=List of 2
#>   .. ..£ : chr [1:261] "1" "4" "5" "6" ...
#>   .. ..£ : chr [1:261] "1" "4" "5" "6" ...
```

We see that `kinship` is a list containing 3 matrices. Each matrix is 261 by 261 - where the number of subjects is 261 - and symmetric.

4.1.4 Statistical model

Before we simulate phenotype data, we first specify our statistical model.

We use the model:

$$vec(Y) = Xvec(B) + vec(G) + vec(E)$$

where Y is a n by 2 matrix, where each row is one subject and each column is one quantitative trait. X is a $2n$ by $2f$ design matrix containing n by f allele probabilities matrices for each of two (possibly identical) markers. Thus, X is a block-diagonal matrix, with exactly two n by f blocks on the diagonal. B is a f by 2 matrix. "vec" refers to the vectorization operator. "vec(B)", where B is a f by 2 matrix, is, thus, a (column) vector of length $2f$ that is formed by stacking the second column of B beneath the first column of B.

G is a matrix of random effects. We specify its distribution as matrix-variate normal with mean being a n by 2 matrix of zeros, covariance among row vectors a n by n kinship matrix, K, and covariance among column vectors a 2 by 2 genetic covariance matrix, V_g.

In mathematical notation, we write:

$$G \sim MN_{\text{n by 2}}(0, K, V_g)$$

We also need to specify the distribution of the E matrix, which contains the random errors. E is a random n by 2 matrix that is distributed as a matrix-variate normal distribution with mean being the n by 2 zero matrix, covariance among row vectors I_n, the n by n identity matrix, and covariance among columns the 2 by 2 matrix V_e.

$$E \sim MN_{\text{n by 2}}(0, I_n, V_e)$$

In practice, we typically measure the phenotype matrix Y. We also treat as known the design matrix X and the kinship matrix K. We then infer the values of B, V_g, and V_e.

4.1.5 Simulating phenotypes with `qtl2pleio::sim1`

The function to simulate phenotypes in `qtl2pleio` is `sim1`. By examining its help page, we see that it takes five arguments. The help page also gives the dimensions of the inputs.

```
# set up the design matrix, X
pp <- pr[[2]]  #we'll work with Chr 3's genotype data
```

```
dim(pp)
#> [1] 261   8 102
```

We prepare a block-diagonal design matrix X that contains two nonzero blocks on the diagonal, one for each trait. We use here a function from the `gemma2` R package to set up the needed matrix.

```
# Next, we prepare a design matrix X
X <- gemma2::stagger_mats(pp[, , 50], pp[, , 50])
dim(X)
#> [1] 522  16
```

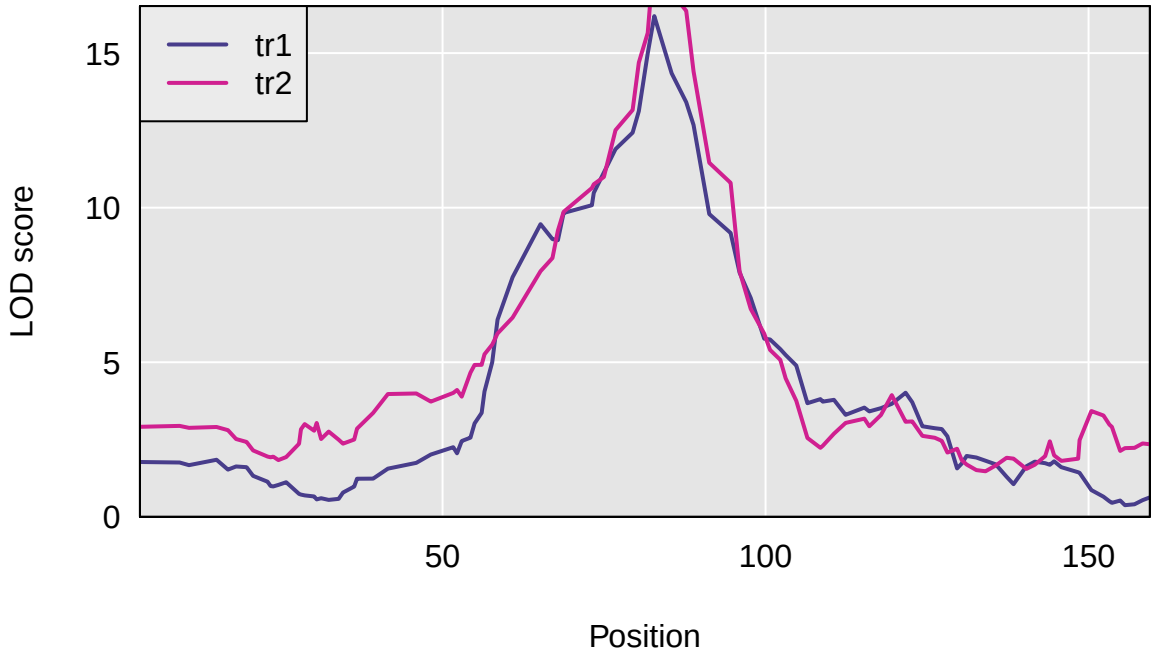
```
# assemble B matrix of allele effects
B <- matrix(data = c(-1, -1, -1, -1, 1, 1, 1, 1, -1, -1, -1,
    -1, 1, 1, 1, 1), nrow = 8, ncol = 2, byrow = FALSE)
# verify that B is what we want:
B
#>      [,1] [,2]
#> [1,]   -1   -1
#> [2,]   -1   -1
#> [3,]   -1   -1
#> [4,]   -1   -1
#> [5,]    1    1
#> [6,]    1    1
#> [7,]    1    1
#> [8,]    1    1
# set.seed to ensure reproducibility
set.seed(2018 - 1 - 30)
# call to sim1
Ypre <- sim1(X = X, B = B, Vg = diag(2), Ve = diag(2), kinship = kinship[[2]])
Y <- matrix(Ypre, nrow = 261, ncol = 2, byrow = FALSE)
rownames(Y) <- rownames(pp)
colnames(Y) <- c("tr1", "tr2")
```

Let's perform univariate QTL mapping for each of the two traits in the Y matrix.

```
s1 <- scan1(genoprobs = pr, pheno = Y, kinship = kinship)
```

Here is a plot of the results.

```
plot(s1, DOex$pmap$`3`)
plot(s1, DOex$pmap$`3`, lod = 2, col = "violetred", add = TRUE)
legend("topleft", colnames(s1), lwd = 2, col = c("darkslateblue",
    "violetred"), bg = "gray92")
```



We see that the two traits share a peak on Chr 3.

And here are the observed QTL peaks with LOD > 8. In practice, we could do a permutation test to determine a threshold for family-wise error rate control.

```
find_peaks(s1, map = DOex$pmap, threshold = 8)
#>   lodindex lodcolumn chr       pos      lod
#> 1        1       tr1   3  82.77806 16.19704
#> 2        2       tr2   3  82.77806 18.26406
#> 3        2       tr2   X 103.79061 16.19708
```

4.1.6 Perform two-dimensional scan as first step in pleiotropy vs. separate QTL hypothesis test

We now have the inputs that we need to do a pleiotropy vs. separate QTL test. We have the founder allele dosages for one chromosome, *i.e.*, Chr 3, in the R object `pp`, the matrix of two trait measurements in `Y`, and a LOCO-derived kinship matrix. We also specify, via the `start_snp` argument, the starting point for the two-dimensional scan within the array of founder allele dosages. Here, we choose the 38th marker in the array as the starting point. Via the `n_snp` argument, we specify the number of markers to include in the two-dimensional scan. Here, we input 25, so that we fit the bivariate linear mixed effects model at 25*25 = 625 ordered pairs of markers. In practice, we usually use between 100 and 300 markers for most two-dimensional scans.

Lastly, we specify the number of cores to use, with the `n_cores` argument. We set it to 1 here, to ensure that the vignette can be run by CRAN. However, in practice, you may wish to increase the number of cores to accelerate computing.

```
out <- scan_pvl(probs = pp, pheno = Y, kinship = kinship$`3`,
    start_snp = 38, n_snp = 25, n_cores = 1)
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
```

The number of cores available will vary by computer. For example, on my Macbook pro computer, with 16GB RAM, I have access to 8 cores. If I use all 8, I can't do other computing tasks, so I often set `n_cores` to 7.

To check how many cores are available on your computer, run this code.

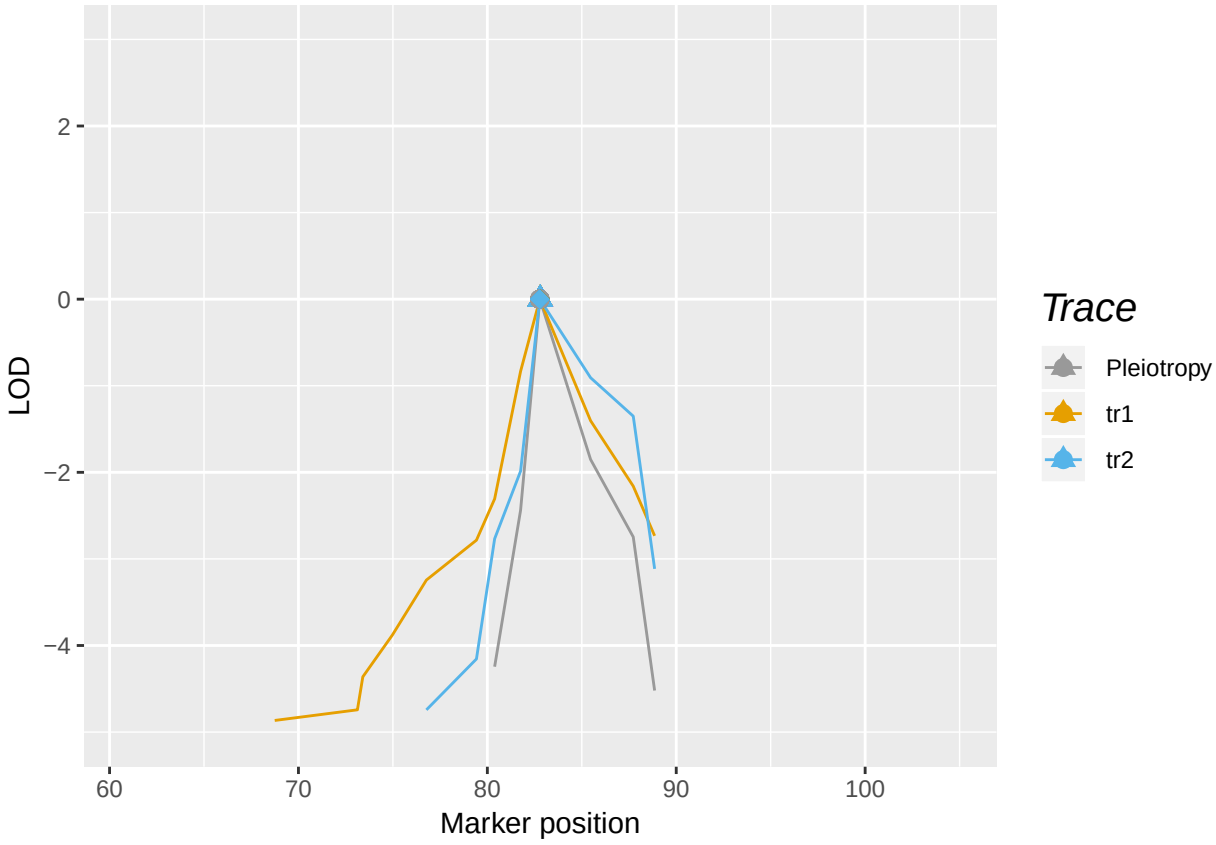
```
parallel::detectCores()
```

Create a profile LOD plot to visualize results of two-dimensional scan To visualize results from our two-dimensional scan, we calculate profile LOD for each trait. The code below makes use of the R package `ggplot2` to plot profile LODs over the scan region.

```
out
#> # A tibble: 625 x 3
#>    Var1                Var2        loglik
#>    <chr>               <chr>        <dbl>
#>  1 JAX00108034         JAX00108034  -866.
#>  2 backupUNC031096286 JAX00108034  -863.
#>  3 JAX00525579         JAX00108034  -865.
#>  4 JAX00525718         JAX00108034  -865.
#>  5 UNC030088171        JAX00108034  -864.
#>  6 backupUNC030474070 JAX00108034  -864.
#>  7 backupUNC030474244 JAX00108034  -863.
#>  8 UNC030103315        JAX00108034  -862.
#>  9 UNC030107226        JAX00108034  -861.
#> 10 JAX00527615         JAX00108034  -861.
#> # ... with 615 more rows
```

We see that `out` is a 625 by 3 tibble, as expected. The first two columns contain the marker ids for each ordered pair of markers. The third column contains the log-likelihood values.

```
library(dplyr)
#>
#> Attaching package: 'dplyr'
#> The following objects are masked from 'package:stats':
#>
#>     filter, lag
#> The following objects are masked from 'package:base':
#>
#>     intersect, setdiff, setequal, union
out %>% tidy_scan_pvl(DOex$pmap$`3`) %>% add_intercepts(intercepts_univariate = c(82.8,
    82.8)) %>% plot_pvl(phenames = c("tr1", "tr2"))
#> Warning: Removed 49 rows containing missing values (geom_path).
```



We first pass the `scan_pvl` output, *i.e.*, `out`, to the function `tidy_scan_pvl` to add the physical map coordinates to the `out` tibble. We pipe that output to the `add_intercepts` function. This function adds columns for the univariate peak positions. Note that we need to specify the univariate peak positions by hand. In the current case, the two traits have identical peak positions.

Finally, the output of `add_intercepts` is piped to `plot_pvl`. This function uses `ggplot2` functions to create a profile LOD plot with three "traces": one for each trait and a third for all ordered pairs under the pleiotropy hypothesis.

Calculate the likelihood ratio test statistic for pleiotropy v separate QTL We use the function `calc_lrt_tib` to calculate the likelihood ratio test statistic value for the specified traits and specified genomic region.

```
(lrt <- calc_lrt_tib(out))

#> [1] 0
```

Bootstrap analysis to get p-values

The calibration of test statistic values to get p-values uses bootstrap methods because we don't know the theoretical distribution of the test statistic under the null hypothesis. Thus, we use a bootstrap approach to obtain an empirical distribution of test statistic values under the null hypothesis of the presence of one pleiotropic locus.

We will use the function `boot_pvl` from our package `qtl2pleio`.

We use a parametric bootstrap strategy in which we first use the studied phenotypes to infer the values of model parameters. Once we have the inferred values of the model parameters, we simulate phenotypes from the pleiotropy model (with the inferred parameter values).

A natural question that arises is "which marker's allele probabilities do we use when simulating phenotypes?" We use the marker that, under the null hypothesis, *i.e.*, under the pleiotropy constraint, yields the greatest value of the log-likelihood.

Before we call `boot_pvl`, we need to identify the index (on the chromosome under study) of the marker that maximizes the likelihood under the pleiotropy constraint. To do this, we use the `qtl2pleio` function `find_pleio_peak_tib`.

```
(pleio_index <- find_pleio_peak_tib(out, start_snp = 38))
#> loglik13
#>       50
```

```
set.seed(2018 - 11 - 25)
system.time(b_out <- boot_pvl(probs = pp, pheno = Y, pleio_peak_index = pleio_index,
    kinship = kinship$`3`, nboot_per_job = 10, start_snp = 38,
    n_snp = 25))
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
```

```
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#> starting covariance matrices estimation with data from 261 subjects.
#> covariance matrices estimation completed.
#>    user  system elapsed
#>  80.916   4.184  85.332
```

The argument `nboot_per_job` indicates the number of bootstrap samples that will be created and analyzed. Here, we set `nboot_per_job = 10`, so we expect to see returned a numeric vector of length 10, where each entry is a LRT statistic value from a distinct bootstrap sample.

Finally, we determine a bootstrap p-value in the usual method. We treat the bootstrap samples' test statistics as an empirical distribution of the test statistic under the null hypothesis of pleiotropy. Thus, to get a p-value, we want to ask "What is the probability, under the null hypothesis, of observing a test statistic value that is at least as extreme as that which we observed?"

```
b_out
#>  [1] 2.8881948 0.0000000 0.9943472 0.0000000 1.1553322 0.6764714 0.0000000
#>  [8] 1.1446433 0.0000000 1.0460109
(pvalue <- mean(b_out >= lrt))
#> [1] 1
```

In practice, one would want to use many more bootstrap samples to achieve an empirical distribution that is closer to the theoretical distribution of the test statistic under the null hypothesis.

However, if one wants to perform analyses with a reasonable number - say 400 - bootstrap samples, this will take a very long time - many days - on a single laptop computer. We have used a series of computer clusters that are coordinated by the University of Wisconsin-Madison's Center for High-throughput Computing (http://chtc.cs.wisc.edu). We typically are able to analyze 1000 bootstrap samples in less than 24 hours with this service.

4.1.7 Session info

```
devtools::session_info()
#> - Session info -------------------------------------------------------------
#>  setting  value
#>  version  R version 3.5.3 (2019-03-11)
#>  os       macOS Mojave 10.14.3
#>  system   x86_64, darwin15.6.0
#>  ui       RStudio
#>  language (EN)
#>  collate  en_US.UTF-8
#>  ctype    en_US.UTF-8
#>  tz       America/Chicago
#>  date     2019-03-17
#>
#> - Packages -----------------------------------------------------------------
#>  ! package        * version    date       lib
#>    assertthat       0.2.0      2017-04-11 [1]
#>    backports        1.1.3      2018-12-14 [1]
#>    bit              1.1-14     2018-05-29 [1]
#>    bit64            0.9-7      2017-05-08 [1]
#>    blob             1.1.1      2018-03-25 [1]
#>    broman           0.68-2     2018-07-25 [1]
#>    callr            3.2.0      2019-03-15 [1]
#>    cli              1.0.1      2018-09-25 [1]
```

```
#>    colorspace       1.4-0      2019-01-13 [1]
#>    commonmark       1.7        2018-12-01 [1]
#>    crayon           1.3.4      2017-09-16 [1]
#>    curl             3.3        2019-01-10 [1]
#>    data.table       1.12.0     2019-01-13 [1]
#>    DBI              1.0.0      2018-05-02 [1]
#>    desc             1.2.0      2018-05-01 [1]
#>    devtools       * 2.0.1      2018-10-26 [1]
#>    digest           0.6.18     2018-10-10 [1]
#>    dplyr          * 0.8.0.1    2019-02-15 [1]
#>    evaluate         0.13       2019-02-12 [1]
#>    fansi            0.4.0      2018-10-05 [1]
#>    formatR          1.6        2019-03-05 [1]
#>    fs               1.2.6      2018-08-23 [1]
#>    gemma2           0.0.1.1    2019-03-13 [1]
#>    ggplot2          3.1.0      2018-10-25 [1]
#>    glue             1.3.1      2019-03-12 [1]
#>    gtable           0.2.0      2016-02-26 [1]
#>    hms              0.4.2      2018-03-10 [1]
#>    htmltools        0.3.6      2017-04-28 [1]
#>    jsonlite         1.6        2018-12-07 [1]
#>    knitr            1.22       2019-03-08 [1]
#>    labeling         0.3        2014-08-23 [1]
#>    lattice          0.20-38    2018-11-04 [1]
#>    lazyeval         0.2.2      2019-03-15 [1]
#>    magrittr         1.5        2014-11-22 [1]
#>    MASS             7.3-51.1   2018-11-01 [1]
#>    Matrix           1.2-15     2018-11-01 [1]
#>    memoise          1.1.0      2017-04-21 [1]
#>    microbenchmark   1.4-6      2018-10-18 [1]
```

```
#>    munsell          0.5.0      2018-06-12 [1]
#>    mvtnorm          1.0-10     2019-03-05 [1]
#>    packrat          0.5.0      2018-11-14 [1]
#>    pillar           1.3.1      2018-12-15 [1]
#>    pkgbuild         1.0.2      2018-10-16 [1]
#>    pkgconfig        2.0.2      2018-08-16 [1]
#>    pkgload          1.0.2      2018-10-29 [1]
#>    plyr             1.8.4      2016-06-08 [1]
#>    prettyunits      1.0.2      2015-07-13 [1]
#>    processx         3.3.0      2019-03-10 [1]
#>    progress         1.2.0      2018-06-14 [1]
#>    ps               1.3.0      2018-12-21 [1]
#>    purrr            0.3.2      2019-03-15 [1]
#>    qtl2           * 0.19-7     2019-03-17 [1]
#>  P qtl2pleio      * 0.1.2.9001 2019-03-17 [?]
#>    R6               2.4.0      2019-02-14 [1]
#>    rcmdcheck        1.3.2      2018-11-10 [1]
#>  V Rcpp             1.0.0.3    2019-03-17 [1]
#>    RcppEigen        0.3.3.5.0  2019-03-08 [1]
#>    readr            1.3.1      2018-12-21 [1]
#>    remotes          2.0.2      2018-10-30 [1]
#>    rlang            0.3.1      2019-01-08 [1]
#>    rmarkdown      * 1.11       2018-12-08 [1]
#>    roxygen2         6.1.1      2018-11-07 [1]
#>    rprojroot        1.3-2      2018-01-03 [1]
#>    RSQLite          2.1.1      2018-05-06 [1]
#>    rstudioapi       0.9.0      2019-01-09 [1]
#>    scales           1.0.0      2018-08-09 [1]
#>    sessioninfo      1.1.1      2018-11-05 [1]
#>    stringi          1.4.3      2019-03-12 [1]
```

```
#>    stringr          1.4.0      2019-02-10 [1]
#>    testthat         2.0.1      2018-10-13 [1]
#>    tibble           2.1.1      2019-03-16 [1]
#>    tidyselect       0.2.5      2018-10-11 [1]
#>    usethis        * 1.4.0      2018-08-14 [1]
#>    utf8             1.1.4      2018-05-24 [1]
#>    withr            2.1.2      2018-03-15 [1]
#>    xfun             0.5        2019-02-20 [1]
#>    xml2             1.2.0      2018-01-24 [1]
#>    xopen            1.0.0      2018-09-17 [1]
#>    yaml             2.2.0      2018-07-25 [1]
#>  source
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.3)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
```

```
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  Github (fboehm/gemma2@25fb644)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.3)
#>  CRAN (R 3.5.3)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.3)
#>  CRAN (R 3.5.3)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
```

```
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.3)
#>  Github (rqtl/qtl2@3ccd775)
#>  local
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  Github (RcppCore/Rcpp@0837b35)
#>  Github (RcppCore/RcppEigen@ca305c8)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.3)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.2)
#>  CRAN (R 3.5.0)
#>  CRAN (R 3.5.0)
```

```
#>  CRAN (R 3.5.0)
#>
#> [1] /Library/Frameworks/R.framework/Versions/3.5/Resources/library
#>
#>  V -- Loaded and on-disk version mismatch.
#>  P -- Loaded and on-disk path mismatch.
```

4.2 Recla analysis

4.3 Cluster computing for bootstrap test

Chapter 5

Conclusions

I've successfully developed a pleiotropy test for multiparental populations. I discussed our new methods in Chapter 2. In Chapter 3, we illustrated the test's use in three vignettes. The first vignette compared pleiotropy testing with mediation analysis in the dissection of expression trait QTL hotspots. I learned that the pleiotropy test provides information about the number of underlying QTL even when mediation analyses don't identify intermediates. The second vignette examined my test's power to detect separate QTL in pairs of local expression traits. I learned that both interlocus distance and univariate LOD scores impact test statistic values. In the last vignette, I applied my test to two gut microbiome-related traits. From this analysis, I learned that the two traits share a pleiotropic QTL and, thus, it is reasonable to conduct further causal modeling studies for these two traits. Chapter 4 demonstrates features of the 'qtl2pleio' R package. This package implements my pleiotropy test and uses the data structures in the R package 'qtl2' (K. W. Broman, D. M. Gatti, et al., 2019).

5.1 A multivariate pleiotropy test

Pleiotropy tests for two traits at a time have a valuable role in complex trait genetics. However, to fully use the tens of thousands of experimentally measured traits, we need to consider testing more than two traits at a time. Suppose that five traits map to a single region spanned by 100 markers. One might perform a series of $\binom{5}{2} = 10$ bivariate QTL scans and 10 pairwise tests for pleiotropy. Each bivariate scan would require $100^2 = 10,000$ model fits by generalized least squares. Alternatively, one could perform a d-variate QTL scan, with $d = 5$ in this case. With the results of the d-variate scan, a variety of statistical

hypotheses could be tested. For example, one could formulate a test for the null hypothesis that all five traits share a pleiotropic QTL against the alternative that the first two traits share a single QTL and the last three traits share a distinct, pleiotropic QTL. The d-variate scan over the 100-marker region, would require $100^d = 100^5 = 10$ billion model fits via generalized least squares. With distributed computing resources, including the resources at the University of Wisconsin's Center for High-throughput Computing, this is not an unreasonable volume of computing. The use of C++ for generalized least squares calculations decreases the computing time for each model fit.

5.2 X chromosome

5.3 Sensitivity analysis for unmeasured confounding in mediation analysis

Bibliography

Akaike, Hirotugu (1974). "A new look at the statistical model identification". In: *IEEE transactions on automatic control* 19.6, pp. 716–723 (cit. on p. 36).

Attie, Alan D, Gary A Churchill, and Joseph H Nadeau (2017). "How mice are indispensable for understanding obesity and diabetes genetics". In: *Current opinion in endocrinology, diabetes, and obesity* 24.2, p. 83 (cit. on p. 2).

Baron, Reuben M and David A Kenny (1986). "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." In: *Journal of personality and social psychology* 51.6, p. 1173 (cit. on pp. 35, 38).

Baum, Leonard E, Ted Petrie, George Soules, and Norman Weiss (1970). "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains". In: *The annals of mathematical statistics* 41.1, pp. 164–171 (cit. on p. 13).

Boehm, Frederick (2018a). *gemma2: Zhou & Stephens (2014) GEMMA multivariate linear mixed model*. R package version 0.0.1. URL: https://github.com/fboehm/gemma2 (cit. on p. 26).

— (2018b). *qtl2pleio: Hypothesis test of close linkage vs pleiotropy in multiparental populations*. R package version 0.1.0 (cit. on pp. 44, 52, 53, 64).

Bolyen, Evan, Jai Ram Rideout, Matthew R Dillon, Nicholas A Bokulich, Christian Abnet, Gabriel A Al-Ghalith, Harriet Alexander, Eric J Alm, Manimozhiyan Arumugam, Francesco Asnicar, et al. (2018). *QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science*. Tech. rep. PeerJ Preprints (cit. on p. 61).

Broman, Karl W. (2006). *Use of hidden Markov models for QTL mapping*. Tech. rep. 125. Johns Hopkins University Department of Biostatistics. URL: http://biostats.bepress.com/jhubiostat/paper125 (cit. on p. 13).

Broman, Karl W. (2012a). "Genotype probabilities at intermediate generations in the construction of recombinant inbred lines". In: *Genetics* 190.2, pp. 403–412 (cit. on pp. 20, 61).

— (2012b). "Haplotype probabilities in advanced intercross populations". In: *G3: Genes, Genomes, Genetics* 2.2, pp. 199–202 (cit. on pp. 20, 61).

Broman, Karl W., D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, S. Sen, B. S. Yandell, and G. A. Churchill (2019). "R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multi-parent populations". In: *Genetics, to appear* (cit. on pp. 20, 21, 24, 52, 61, 62, 86).

Broman, Karl W. and Saunak Sen (2009). *A Guide to QTL Mapping with R/qtl*. Vol. 46. Springer (cit. on pp. 1, 4, 5, 7–9, 12, 13, 20).

Callahan, Benjamin J, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes (2016). "DADA2: high-resolution sample inference from Illumina amplicon data". In: *Nature methods* 13.7, p. 581 (cit. on p. 61).

Chesler, Elissa J, Daniel M Gatti, Andrew P Morgan, Marge Strobel, Laura Trepanier, Denesa Oberbeck, Shannon McWeeney, Robert Hitzemann, Martin Ferris, Rachel McMullan, et al. (2016). "Diversity Outbred Mice at 21: Maintaining Allelic Variation in the Face of Selection". In: *G3: Genes— Genomes— Genetics*, g3–116 (cit. on pp. 18, 23).

Chick, Joel M, Steven C Munger, Petr Simecek, Edward L Huttlin, Kwangbom Choi, Daniel M Gatti, Narayanan Raghupathy, Karen L Svenson, Gary A Churchill, and Steven P Gygi (2016). "Defining the consequences of genetic variation on a proteome-wide scale". In: *Nature* 534.7608, p. 500 (cit. on pp. 1, 35, 38, 39, 41, 43, 67).

Churchill, Gary A, David C Airey, Hooman Allayee, Joe M Angel, Alan D Attie, Jackson Beatty, William D Beavis, John K Belknap, Beth Bennett, Wade Berrettini, et al. (2004). "The Collaborative Cross, a community resource for the genetic analysis of complex traits". In: *Nature genetics* 36.11, pp. 1133–1137 (cit. on pp. 1, 18, 23).

Churchill, Gary A and Rebecca W Doerge (1994). "Empirical threshold values for quantitative trait mapping." In: *Genetics* 138.3, pp. 963–971 (cit. on p. 9).

Churchill, Gary A, Daniel M Gatti, Steven C Munger, and Karen L Svenson (2012). "The diversity outbred mouse population". In: *Mammalian genome* 23.9-10, pp. 713–718 (cit. on pp. 18, 23).

Crick, Francis (1970). "Central dogma of molecular biology". In: *Nature* 227.5258, p. 561 (cit. on p. 38).

Crick, Francis HC (1958). "On protein synthesis". In: *Symp Soc Exp Biol*. Vol. 12. 138-63, p. 8 (cit. on p. 38).

Cubillos, Francisco A, Leopold Parts, Francisco Salinas, Anders Bergstrom, Eugenio Scovacricchi, Amin Zia, Christopher JR Illingworth, Ville Mustonen, Sebastian Ibstedt, Jonas Warringer, et al. (2013). "High-resolution mapping of complex traits with a four-parent advanced intercross yeast population". In: *Genetics* 195.3, pp. 1141–1155 (cit. on p. 37).

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38 (cit. on pp. 16, 20).

DeSantis, Todd Z, Philip Hugenholtz, Neils Larsen, Mark Rojas, Eoin L Brodie, Keith Keller, Thomas Huber, Daniel Dalevi, Ping Hu, and Gary L Andersen (2006). "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB". In: *Applied and environmental microbiology* 72.7, pp. 5069–5072 (cit. on p. 61).

Eddelbuettel, Dirk, Romain Francois, J Allaire, John Chambers, Douglas Bates, and Kevin Ushey (2011). "Rcpp: Seamless R and C++ integration". In: *Journal of Statistical Software* 40.8, pp. 1–18 (cit. on p. 35).

Efron, B (1979). "Bootstrap methods: another look at the jackknife". In: *The Annals of Statistics* 7.1, pp. 1–26 (cit. on pp. 24, 27).

Fani, L, S Bak, Patric Delhanty, EFC Van Rossum, and ELT Van Den Akker (2014). "The melanocortin-4 receptor as target for obesity treatment: a systematic review of emerging pharmacological therapeutic options". In: *International journal of obesity* 38.2, p. 163 (cit. on p. 2).

Farooqi, I Sadaf, Julia M Keogh, Giles SH Yeo, Emma J Lank, Tim Cheetham, and Stephen O'rahilly (2003). "Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene". In: *New England Journal of Medicine* 348.12, pp. 1085–1095 (cit. on p. 2).

Gatti, Daniel M, Karen L Svenson, Andrey Shabalin, Long-Yang Wu, William Valdar, Petr Simecek, Neal Goodwin, Riyan Cheng, Daniel Pomp, Abraham Palmer, et al. (2014). "Quantitative trait locus mapping methods for diversity outbred mice". In: *G3: Genes, Genomes, Genetics* 4.9, pp. 1623–1633 (cit. on pp. 20, 21).

Han, Xianlin, Kui Yang, and Richard W Gross (2012). "Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses". In: *Mass spectrometry reviews* 31.1, pp. 134–178 (cit. on p. 36).

1575 Huang, Bevan E, Andrew W George, Kerrie L Forrest, Andrzej Kilian, Matthew J Hayden, Matthew K
1576 Morell, and Colin R Cavanagh (2012). "A multiparent advanced generation inter-cross population for
1577 genetic analysis in wheat". In: *Plant biotechnology journal* 10.7, pp. 826–839 (cit. on p. 1).
1578 Huang, Xueqing, Maria-Joao Paulo, Martin Boer, Sigi Effgen, Paul Keizer, Maarten Koornneef, and Fred A
1579 van Eeuwijk (2011). "Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant
1580 inbred line population". In: *Proceedings of the National Academy of Sciences* 108.11, pp. 4488–4493 (cit.
1581 on p. 1).
1582 Huszar, Dennis, Catherine A Lynch, Victoria Fairchild-Huntress, Judy H Dunmore, Qing Fang, Lucy R
1583 Berkemeier, Wei Gu, Robert A Kesterson, Bruce A Boston, Roger D Cone, et al. (1997). "Targeted
1584 disruption of the melanocortin-4 receptor results in obesity in mice". In: *Cell* 88.1, pp. 131–141 (cit. on
1585 p. 2).
1586 Jansen, Ritsert C. (2007). "Quantitative trait loci in inbred lines". In: *Handbook of statistical genetics*. Ed. by
1587 David J Balding, Martin Bishop, and Chris Cannings. John Wiley & Sons. Chap. 18, pp. 587–622 (cit. on
1588 p. 1).
1589 Jiang, Changjian and Zhao-Bang Zeng (1995). "Multiple trait analysis of genetic mapping for quantitative
1590 trait loci." In: *Genetics* 140.3, pp. 1111–1127 (cit. on pp. 14–16, 23–25, 27, 31, 36).
1591 Kang, Hyun Min, Jae Hoon Sul, Susan K Service, Noah A Zaitlen, Sit-yee Kong, Nelson B Freimer, Chiara
1592 Sabatti, Eleazar Eskin, et al. (2010). "Variance component model to account for sample structure in
1593 genome-wide association studies". In: *Nature genetics* 42.4, pp. 348–354 (cit. on p. 31).
1594 Kang, Hyun Min, Noah A Zaitlen, Claire M Wade, Andrew Kirby, David Heckerman, Mark J Daly, and
1595 Eleazar Eskin (2008). "Efficient control of population structure in model organism association mapping".
1596 In: *Genetics* 178.3, pp. 1709–1723 (cit. on p. 21).
1597 Katoh, Kazutaka and Daron M Standley (2013). "MAFFT multiple sequence alignment software version
1598 7: improvements in performance and usability". In: *Molecular biology and evolution* 30.4, pp. 772–780
1599 (cit. on p. 61).
1600 Keller, Mark P, Daniel M Gatti, Kathryn L Schueler, Mary E Rabaglia, Donnie S Stapleton, Petr Simecek,
1601 Matthew Vincent, Sadie Allen, Aimee Teo Broman, Rhonda Bacher, et al. (2018). "Genetic Drivers of
1602 Pancreatic Islet Function". In: *Genetics*, genetics–300864 (cit. on pp. 1, 28, 38, 39, 43, 51, 52, 60, 105).
1603 Kemis, Julia H. et al. (submitted). "Genetic determinants of gut microbiota composition and bile acid profiles
1604 in mice". In: (cit. on p. 60).

King, Elizabeth G, Chris M Merkes, Casey L McNeil, Steven R Hoofer, Saunak Sen, Karl W. Broman, Anthony D Long, and Stuart J Macdonald (2012). "Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource". In: *Genome research*, gr–134031 (cit. on pp. 35, 53, 59, 67).

Klebig, ML, JE Wilkinson, JG Geisler, and RP Woychik (1995). "Ectopic expression of the agouti gene in transgenic mice causes obesity, features of type II diabetes, and yellow fur". In: *Proceedings of the National Academy of Sciences* 92.11, pp. 4728–4732 (cit. on p. 2).

Knott, Sara A and Chris S Haley (2000). "Multitrait least squares for quantitative trait loci detection". In: *Genetics* 156.2, pp. 899–911 (cit. on pp. 23, 24).

Koning, Dirk-Jan de and Lauren M McIntyre (2014). "GENETICS and G3: Community-Driven Science, Community-Driven Journals". In: *Genetics* 198.1, pp. 1–2 (cit. on pp. 18, 23).

— (2017). "Back to the Future: Multiparent Populations Provide the Key to Unlocking the Genetic Basis of Complex Traits". In: *Genetics* 206, pp. 527–529 (cit. on p. 1).

Korol, Abraham B, Yefim I Ronin, and Valery M Kirzhner (1995). "Interval mapping of quantitative trait loci employing correlated trait complexes." In: *Genetics* 140.3, pp. 1137–1147 (cit. on p. 14).

Kover, Paula X, William Valdar, Joseph Trakalo, Nora Scarcelli, Ian M Ehrenreich, Michael D Purugganan, Caroline Durrant, and Richard Mott (2009). "A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana". In: *PLoS genetics* 5.7, e1000551 (cit. on pp. 1, 37).

Lander, Eric S and David Botstein (1989). "Mapping mendelian factors underlying quantitative traits using RFLP linkage maps." In: *Genetics* 121.1, pp. 185–199 (cit. on pp. 1, 12).

Logan, Ryan W, Raymond F Robledo, Jill M Recla, Vivek M Philip, Jason A Bubier, Jeremy J Jay, Carter Harwood, Troy Wilcox, Daniel M Gatti, Carol J Bult, et al. (2013). "High-precision genetic mapping of behavioral traits in the diversity outbred mouse population". In: *Genes, Brain and Behavior* 12.4, pp. 424–437 (cit. on pp. 24, 28, 31, 35).

Lu, Dongsi, Derril Willard, Indravadan R Patel, Sue Kadwell, Laurie Overton, Tom Kost, Michael Luther, Wenbiao Chen, Richard P Woychik, William O Wilkison, et al. (1994). "Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor". In: *Nature* 371.6500, p. 799 (cit. on p. 2).

Macdonald, Stuart J and Anthony D Long (2007). "Joint Estimates of QTL Effect and Frequency Using Synthetic Recombinant Populations of Drosophila melanogaster". In: *Genetics* 176.2, pp. 1261–1281. URL: http://www.genetics.org/content/176/2/1261 (cit. on pp. 35, 53, 59, 67).

Mackay, Trudy FC, Stephen Richards, Eric A Stone, Antonio Barbadilla, Julien F Ayroles, Dianhui Zhu, Sonia Casillas, Yi Han, Michael M Magwire, Julie M Cridland, et al. (2012). "The Drosophila melanogaster genetic reference panel". In: *Nature* 482.7384, p. 173 (cit. on p. 37).

MacNeil, Douglas J, Andrew D Howard, Xiaoming Guan, Tung M Fong, Ravi P Nargund, Maria A Bednarek, Mark T Goulet, David H Weinberg, Alison M Strack, Donald J Marsh, et al. (2002). "The role of melanocortins in body weight regulation: opportunities for the treatment of obesity". In: *European journal of pharmacology* 450.1, pp. 93–109 (cit. on p. 2).

Meng, Xiao-Li and Donald B Rubin (1993). "Maximum likelihood estimation via the ECM algorithm: A general framework". In: *Biometrika* 80.2, pp. 267–278 (cit. on p. 16).

Meyer, Hannah Verena, Francesco Paolo Casale, Oliver Stegle, and Ewan Birney (2018). "LiMMBo: a simple, scalable approach for linear mixed models in high-dimensional genetic association studies". In: *bioRxiv*, p. 255497 (cit. on p. 35).

Meyer, Karin (1989). "Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm". In: *Genetics Selection Evolution* 21.3, p. 317 (cit. on p. 26).

— (1991). "Estimating variances and covariances for multivariate animal models by restricted maximum likelihood". In: *Genetics Selection Evolution* 23.1, p. 67 (cit. on p. 26).

Morgan, Andrew P, Chen-Ping Fu, Chia-Yu Kao, Catherine E Welsh, John P Didion, Liran Yadgary, Leeanna Hyacinth, Martin T Ferris, Timothy A Bell, Darla R Miller, et al. (2015). "The mouse universal genotyping array: from substrains to subspecies". In: *G3: Genes— Genomes— Genetics*, g3–115 (cit. on pp. 20, 43, 61).

— (2016). "The mouse universal genotyping array: from substrains to subspecies". In: *G3: Genes, Genomes, Genetics* 6.2, pp. 263–279 (cit. on p. 52).

Neto, Elias Chaibub, Aimee T Broman, Mark P Keller, Alan D Attie, Bin Zhang, Jun Zhu, and Brian S Yandell (2013). "Modeling causality for pairs of phenotypes in system genetics". In: *Genetics* 193.3, pp. 1003–1013 (cit. on p. 67).

Ozsolak, Fatih and Patrice M Milos (2011). "RNA sequencing: advances, challenges and opportunities". In: *Nature reviews genetics* 12.2, p. 87 (cit. on p. 36).

Paulson, Joseph Nathaniel, M Pop, and HC Bravo (2013). "metagenomeSeq: Statistical analysis for sparse high-throughput sequencing". In: *Bioconductor package* 1.0 (cit. on p. 61).

Price, Morgan N, Paramvir S Dehal, and Adam P Arkin (2010). "FastTree 2 approximately maximum-likelihood trees for large alignments". In: *PloS one* 5.3, e9490 (cit. on p. 61).

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/ (cit. on pp. 52, 59).

Raghavan, Chitra, Ramil Mauleon, Vanica Lacorte, Monalisa Jubay, Hein Zaw, Justine Bonifacio, Rakesh Kumar Singh, B Emma Huang, and Hei Leung (2017). "Approaches in Characterizing Genetic Structure and Mapping in a Rice Multiparental Population". In: *G3: Genes, Genomes, Genetics* 7.6, pp. 1721–1730 (cit. on p. 37).

Recla, Jill M, Raymond F Robledo, Daniel M Gatti, Carol J Bult, Gary A Churchill, and Elissa J Chesler (2014). "Precise genetic mapping and integrative bioinformatics in Diversity Outbred mice reveals Hydin as a novel pain gene". In: *Mammalian genome* 25.5-6, pp. 211–222 (cit. on pp. 24, 28, 31, 35).

Sax, Karl (1923). "The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris". In: *Genetics* 8.6, p. 552 (cit. on p. 1).

Schadt, Eric E, John Lamb, Xia Yang, Jun Zhu, Steve Edwards, Debraj GuhaThakurta, Solveig K Sieberts, Stephanie Monks, Marc Reitman, Chunsheng Zhang, et al. (2005). "An integrative genomics approach to infer causal associations between gene expression and disease". In: *Nature genetics* 37.7, p. 710 (cit. on pp. 35, 36, 67).

Shivakumar, M, Gireesh C Giriraj Kumawat, SV Ramesh, and SM Husain (2018). "Soybean MAGIC population: a novel resource for genetics and plant breeding". In: (cit. on p. 1).

Stanley, Patrick D, Enoch Ngoma, Siri ODay, and Elizabeth G King (2017). "Genetic dissection of nutrition-induced plasticity in insulin/insulin-like growth factor signaling and median life span in a Drosophila multiparent population". In: *Genetics* 206.2, pp. 587–602 (cit. on pp. 1, 37).

Sugiyama, Fumihiro, Gary A Churchill, David C Higgins, Conrado Johns, Konstatinos P Makaritsis, Haralambos Gavras, and Beverly Paigen (2001). "Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci". In: *Genomics* 71.1, pp. 70–77 (cit. on p. 8).

Svenson, Karen L, Daniel M Gatti, William Valdar, Catherine E Welsh, Riyan Cheng, Elissa J Chesler, Abraham A Palmer, Leonard McMillan, and Gary A Churchill (2012). "High-resolution genetic mapping using the Mouse Diversity outbred population". In: *Genetics* 190.2, pp. 437–447 (cit. on pp. 1, 18).

Tian, Jianan, Mark P Keller, Aimee Teo Broman, Christina Kendziorski, Brian S Yandell, Alan D Attie, and Karl W. Broman (2016). "The dissection of expression quantitative trait locus hotspots". In: *Genetics* 202.4, pp. 1563–1574 (cit. on pp. 23, 24, 27, 28).

Tisne, Sebastien, Virginie Pomies, Virginie Riou, Indra Syahputra, Benoit Cochard, and Marie Denis (2017). "Identification of Ganoderma disease resistance loci using natural field infection of an oil palm multi-parental population". In: *G3: Genes, Genomes, Genetics* 7.6, pp. 1683–1692 (cit. on pp. 1, 37).

Tyler, Anna L, Leah Rae Donahue, Gary A Churchill, and Gregory W Carter (2016). "Weak epistasis generally stabilizes phenotypes in a mouse intercross". In: *PLoS genetics* 12.2, e1005805 (cit. on p. 36).

Tyler, Anna L, Bo Ji, Daniel M Gatti, Steven C Munger, Gary A Churchill, Karen L Svenson, and Gregory W Carter (2017). "Epistatic networks jointly influence phenotypes related to metabolic disease and gene expression in diversity outbred mice". In: *Genetics* 206.2, pp. 621–639 (cit. on p. 36).

Tyler, Anna L, Wei Lu, Justin J Hendrick, Vivek M Philip, and Gregory W Carter (2013). "CAPE: an R package for combined analysis of pleiotropy and epistasis". In: *PLoS computational biology* 9.10, e1003270 (cit. on p. 36).

Vaisse, Christian, Karine Clement, Emmanuelle Durand, Serge Hercberg, Bernard Guy-Grand, and Philippe Froguel (2000). "Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity". In: *The Journal of clinical investigation* 106.2, pp. 253–262 (cit. on p. 2).

VanderWeele, Tyler (2015). *Explanation in causal inference* (cit. on pp. 42, 51).

Yang, Jian, Noah A Zaitlen, Michael E Goddard, Peter M Visscher, and Alkes L Price (2014). "Advantages and pitfalls in the application of mixed-model association methods". In: *Nature genetics* 46.2, pp. 100–106 (cit. on pp. 21, 24, 25, 31).

Yu, Jianming, James B Holland, Michael D McMullen, and Edward S Buckler (2008). "Genetic design and statistical power of nested association mapping in maize". In: *Genetics* 178.1, pp. 539–551 (cit. on p. 37).

Zeng, Zhao-Bang, Jianjun Liu, Lynn F Stam, Chen-Hung Kao, John M Mercer, and Cathy C Laurie (2000). "Genetic architecture of a morphological shape difference between two Drosophila species". In: *Genetics* 154.1, pp. 299–310 (cit. on p. 27).

Zhou, Xiang and Matthew Stephens (2014). "Efficient multivariate linear mixed model algorithms for genome-wide association studies". In: *Nature methods* 11.4, pp. 407–409 (cit. on pp. 24, 26).

Appendices

Appendix A

Molecular biology primer

A.1 DNA structure

A.2 Cell division

A.3 Central dogma of molecular biology

Appendix B

Statistical inference primer

B.1 Overview of statistics

One characterization of statistics is that it is the science of learning about natural phenomena. The process of learning about natural phenomena proceeds in a cyclical fashion. One first posits a "model" for a natural phenomenon. She designs an experiment to measure a quantity related to the phenomenon of interest. She then collects data, summarizes it as a statistic, and performs a test of competing hypotheses before updating her hypothesis about the natural phenomenon. The newly updated hypothesis replaces the original hypothesis in the second iteration of the cycle.

For example, we might wish to study the genetics of body weight in mice. We first posit the existence of body weight quantitative trait loci (QTL), which are regions of the genome that affect mouse body weight. We quantitatively state the competing possibilities as the null hypothesis that there is no QTL against the alternative hypothesis of presence of a QTL. We then design an experiment, perhaps using genetically diverse mice, like those from the Diversity Outbred mouse population. We measure body weight in, say, 100 Diversity Outbred mice and obtain genome-wide genetic data from each. We then summarize our observations by performing multiple instances of a statistical technique called "linear regression". We perform statistical hypothesis tests to quantify the evidence against the null hypothesis.

B.2 Maximum likelihood methods

B.3 Restricted maximum likelihood methods

B.4 Genotype inference with expectation-maximization

B.5 Genotype inference with hidden Markov models

Appendix C

Supplementary materials for Chapter 2

Table C.1: Eight founder lines and their one-letter abbreviations.

Founder allele	One-letter abbreviation
A/J	A
C57BL/6J	B
129S1/SvImJ	C
NOD/ShiLtJ	D
NZO/H1LTJ	E
Cast/EiJ	F
PWK/PhJ	G
WSB/EiJ	H

Table C.2: Both "hot plate latency" and "percent time in light" demonstrate multiple QTL peaks with LOD scores above 5.

phenotype	chr	pos	LOD score
percent time in light	8	55.28	5.27
hot plate latency	8	57.77	6.22
percent time in light	9	36.70	5.42
hot plate latency	9	46.85	5.22
percent time in light	11	63.39	6.46
hot plate latency	12	43.52	5.13
percent time in light	15	15.24	5.67
hot plate latency	19	47.80	5.48

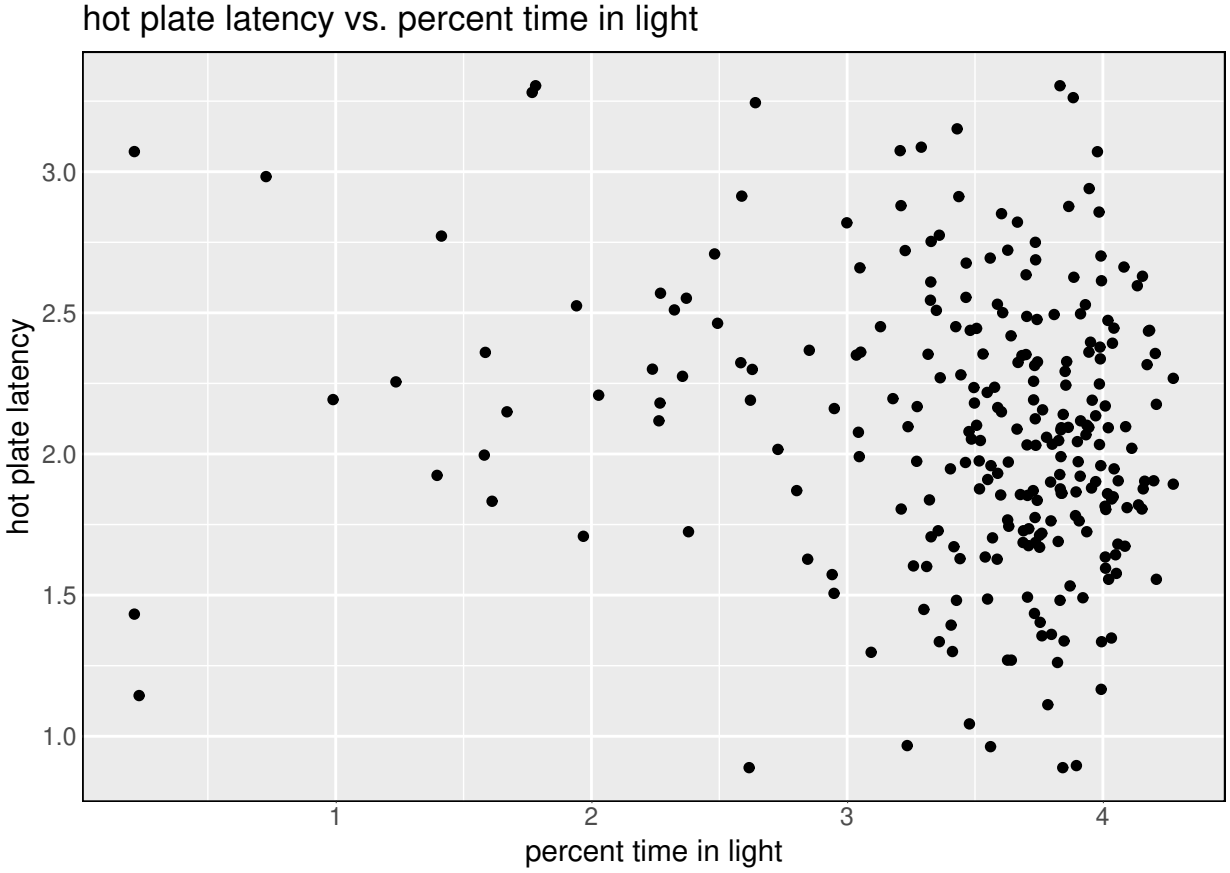


Figure C.1: Scatter plot of "hot plate latency" against "percent time in light", after applying logarithm transformations and winsorizing both traits.

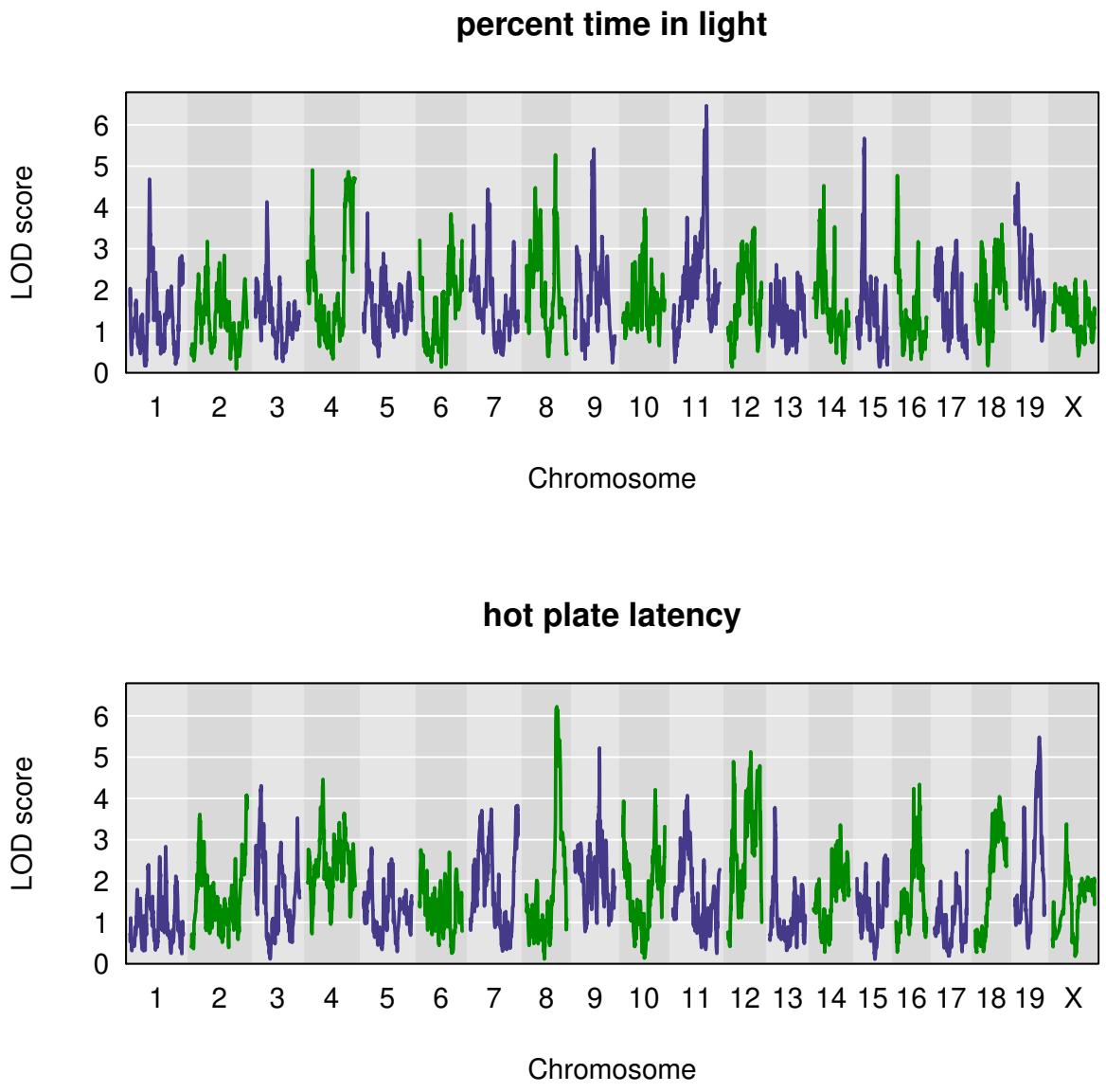


Figure C.2: Genome-wide QTL scan for percent time in light reveals multiple QTL, including one on Chromosome 8.

Appendix D

Supplementary materials for Chapter 3

Table D.1: LOD peak positions and peak heights for 147 expression traits that map to the Chromosome 2 expression trait hotspot. We see that some transcript levels have LOD scores, per our calculations, below the 7.18 genome-wide threshold. We believe that this is due to differences in statistical modeling between our analyses and those of Keller et al. (2018).

Gene	Peak position	LOD
Mtfp1	163.52	17.79
Slc12a7	163.58	8.30
Gpa33	163.58	26.38
Tmem25	163.58	11.26
Klhl29	163.58	11.03
Slc9a3r1	163.58	8.40
Kif1c	163.58	11.12
Gata4	163.58	8.72
Ppp2r5b	163.58	7.80
Vil1	163.58	26.48
Aldh4a1	163.58	9.38
Cotl1	163.58	16.25
Tmprss4	163.58	18.30

Fhod3	163.58	17.62
Zfp750	163.58	7.83
Svop	163.58	10.12
Abcb4	163.58	16.59
Ccnjl	163.58	6.70
Sephs2	163.58	27.21
Pcdh1	163.58	12.48
Fat1	163.58	7.49
Sox4	163.58	14.71
Gm8206	163.58	11.33
Gm8492	163.58	10.47
Clcn5	164.02	10.49
Slc6a8	164.02	8.35
Bcmo1	164.02	47.31
Arrdc4	164.02	8.75
Cacnb3	164.02	47.06
Sec14l2	164.02	12.00
Pgrmc1	164.02	15.88
Baiap2l2	164.02	20.33
Recql5	164.02	33.97
Cpd	164.02	13.70
Degs2	164.02	15.20
Muc13	164.02	18.91
Clic5	164.02	10.39
Tff3	164.02	16.38
Myo7b	164.02	13.70
Afg3l2	164.02	19.12
Sema4g	164.02	23.39
Agap2	164.02	33.99
Plxna2	164.02	8.61
Aldob	164.02	20.99
Epb4.1l4b	164.02	9.21
Sel1l3	164.02	17.92

Sult1b1	164.02	17.63
Hpgds	164.02	31.14
Ush1c	164.02	21.90
Calml4	164.02	12.62
Fam83b	164.02	16.12
Myo15b	164.02	71.57
Inpp5j	164.02	11.57
Ttyh2	164.02	15.20
Cdhr2	164.02	30.00
Myrf	164.02	37.55
Sh3bp4	164.02	9.04
Vgf	164.02	15.05
Grtp1	164.02	23.88
B4galnt3	164.02	20.43
Gucy2c	164.02	12.02
Smim5	164.02	18.20
Nrip1	164.02	9.50
Clrn3	164.02	20.13
Acot4	164.02	12.17
Hunk	164.02	18.50
Zbtb16	164.02	6.69
Osgin1	164.02	13.51
Zfp541	164.02	25.28
2610042L04Rik	164.02	8.14
Gm9429	164.02	9.13
Agxt2	164.02	14.54
Gm17147	164.02	26.94
Gm8281	164.02	8.40
Ddx23	164.03	9.29
Map2k6	164.03	18.54
Npr1	164.03	9.68
Hdac6	164.03	11.28
Vav3	164.03	13.51

Atp11b	164.03	11.77
Aadat	164.03	8.08
Tmem19	164.03	20.60
Gbp4	164.03	7.14
Papola	164.03	12.14
Als2	164.03	18.35
Sult1d1	164.03	11.23
Myo6	164.03	10.27
Dnajc22	164.03	13.72
Unc5d	164.03	6.96
Gm3095	164.03	10.90
Gm26886	164.03	8.76
Eps8	164.06	12.49
Ddc	164.06	13.61
Fras1	164.06	10.06
Card11	164.06	7.20
Glyat	164.06	10.79
Pipox	164.08	15.90
Iyd	164.08	23.23
Man1a	164.26	8.77
Cdc42ep4	164.26	8.11
Ppara	164.26	8.84
Galr1	164.26	10.10
Ctdsp1	164.26	7.13
Eif2ak3	164.26	8.93
Misp	164.26	9.24
Sun1	164.26	7.35
Kctd8	164.26	9.76
Dcaf12l1	164.26	8.95
4930539E08Rik	164.26	8.25
Slc29a4	164.26	7.94
Gm3239	164.26	9.21
Gm3629	164.26	9.77

Gm3252	164.26	9.87
Gm3002	164.26	7.62
Ctsh	164.29	9.76
Dao	164.29	7.06
Ak7	164.31	10.07
Pcp4l1	164.35	9.11
Gm12929	164.62	37.80
Mgat1	164.63	10.10
Atg7	164.76	7.82
Gm11549	165.05	10.23
Ccdc111	165.12	7.77
Fam20a	165.15	7.59
Oscp1	165.16	7.99
Dsc2	165.28	8.06
Adam10	165.28	9.73
Plb1	165.28	8.85
Ccdc89	165.28	7.84
9930013L23Rik	165.28	7.93
Gm13648	165.28	7.08
Igfbp4	165.45	8.68
Ldlrap1	165.45	7.18
Cdh18	165.45	7.96
Arhgef10l	165.46	7.57
Cib3	165.48	8.70
Macf1	165.57	9.10
Fam63b	165.58	8.33
1190002N15Rik	165.63	9.02
Bcl2l14	165.71	7.54
Hist2h2be	165.88	8.96
Gm6428	165.89	7.32
Dennd5b	166.18	9.42
Gm12168	166.18	7.77
Gm12230	166.42	7.46

Acat3	166.61	7.17
Gpr20	166.84	8.41

Table D.2: Annotations for 76 non-anchor genes on Chromosome 19.

Gene	Start	End	Peak position	LOD
C030046E11Rik	29.52	29.61	29.55	95.58
Tctn3	40.60	40.61	40.59	90.00
Gm7237	33.41	33.42	33.67	74.61
Lipo4	33.50	33.52	34.00	68.23
Dock8	25.00	25.20	25.07	63.17
Sorbs1	40.30	40.40	40.48	61.89
Lipm	34.10	34.12	34.06	58.43
Blnk	40.93	40.99	40.76	57.16
A830019P07Rik	35.84	35.92	35.60	55.54
Uhrf2	30.03	30.09	29.96	54.40
Mbl2	30.23	30.24	30.18	52.81
Myof	37.90	38.04	38.05	48.46
Gm27042	40.59	40.59	40.61	44.27
Btaf1	36.93	37.01	36.90	41.25
Hoga1	42.05	42.07	42.09	41.23
Ppp1r3c	36.73	36.74	36.53	40.69
Pcgf5	36.38	36.46	36.24	40.06
Slc35g1	38.40	38.41	38.35	38.11
Pten	32.76	32.83	32.77	37.95
Gldc	30.10	30.18	30.17	36.26
Lgi1	38.26	38.31	38.17	34.91
C330002G04Rik	23.04	23.08	23.34	34.84
Ppapdc2	28.96	28.97	29.09	34.71
Gm8978	33.61	33.63	33.03	34.59
Mms19	41.94	41.98	41.98	32.03
Ankrd22	34.12	34.17	34.04	31.83

Cdc37l1	28.99	29.02	29.03	31.14
Sgms1	32.12	32.39	32.11	30.10
Entpd1	40.61	40.74	40.50	29.73
Cbwd1	24.92	24.96	24.73	29.65
Gm14446	34.59	34.60	34.28	27.65
Ermp1	29.61	29.65	29.70	26.57
Gm9938	23.72	23.73	23.87	26.46
Insl6	29.32	29.33	29.37	26.23
Slc16a12	34.67	34.75	34.71	25.54
Pgm5	24.68	24.86	25.00	24.30
Morn4	42.07	42.09	41.79	23.86
Exosc1	41.92	41.93	42.10	23.28
Smarca2	26.61	26.78	26.59	23.25
4930418C01Rik	24.42	24.43	23.92	23.10
2700046G09Rik	32.39	32.39	32.25	23.02
Kcnv2	27.32	27.34	27.14	22.88
1500017E21Rik	36.61	36.71	37.07	22.78
Fra10ac1	38.19	38.22	38.35	22.48
Rnls	33.14	33.39	34.17	21.94
Noc3l	38.79	38.82	40.20	21.67
Pip5k1b	24.29	24.56	24.15	21.62
Plgrkt	29.35	29.37	29.37	20.65
Ifit3	34.58	34.59	34.28	20.45
Fas	34.29	34.33	34.20	19.65
Slit1	41.60	41.74	41.70	18.95
Rrp12	41.86	41.90	41.71	18.09
Ak3	29.02	29.05	29.55	16.90
A1cf	31.87	31.95	32.11	15.56
4430402I18Rik	28.90	28.97	29.37	15.43
Pdlim1	40.22	40.27	40.25	15.25
Gm26902	34.47	34.48	36.15	14.26
Plce1	38.48	38.79	38.42	14.26
Slc1a1	28.84	28.91	28.97	14.18

Fam122a	24.48	24.48	24.08	14.07
Lipa	34.49	34.53	34.29	14.06
Mamdc2	23.30	23.45	23.35	13.12
Kif11	37.38	37.42	37.33	12.93
4933411K16Rik	42.05	42.05	42.08	12.92
Ccnj	40.83	40.85	40.59	12.19
Gm340	41.58	41.59	41.30	12.17
Fxn	24.26	24.28	24.31	12.07
Stambpl1	34.19	34.24	34.28	11.62
Pde6c	38.13	38.18	38.07	11.54
Cyp26a1	37.70	37.70	37.48	11.35
Ch25h	34.47	34.48	32.50	10.74
Pank1	34.81	34.88	35.55	10.61
9930021J03Rik	29.71	29.81	28.71	10.32
Klf9	23.14	23.17	23.34	10.26
Ubtd1	41.98	42.03	41.71	10.25
Lipk	34.01	34.05	34.29	10.23