Testing Pleiotropy vs. Separate QTL in Multiparental Populations

Frederick Boehm

February 23, 2019

Contents

List of Figures

List of Tables

Chapter 1

Complex traits in multiparental populations

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, 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.

2.2.1 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 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.

2.2.2 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.

2.2.3 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).

2.2.4 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.

2.2.5 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.

2.2.6 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.

2.2.7 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

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.

2.3 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).

2.3.1 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.

2.3.2 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 **S1**). 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

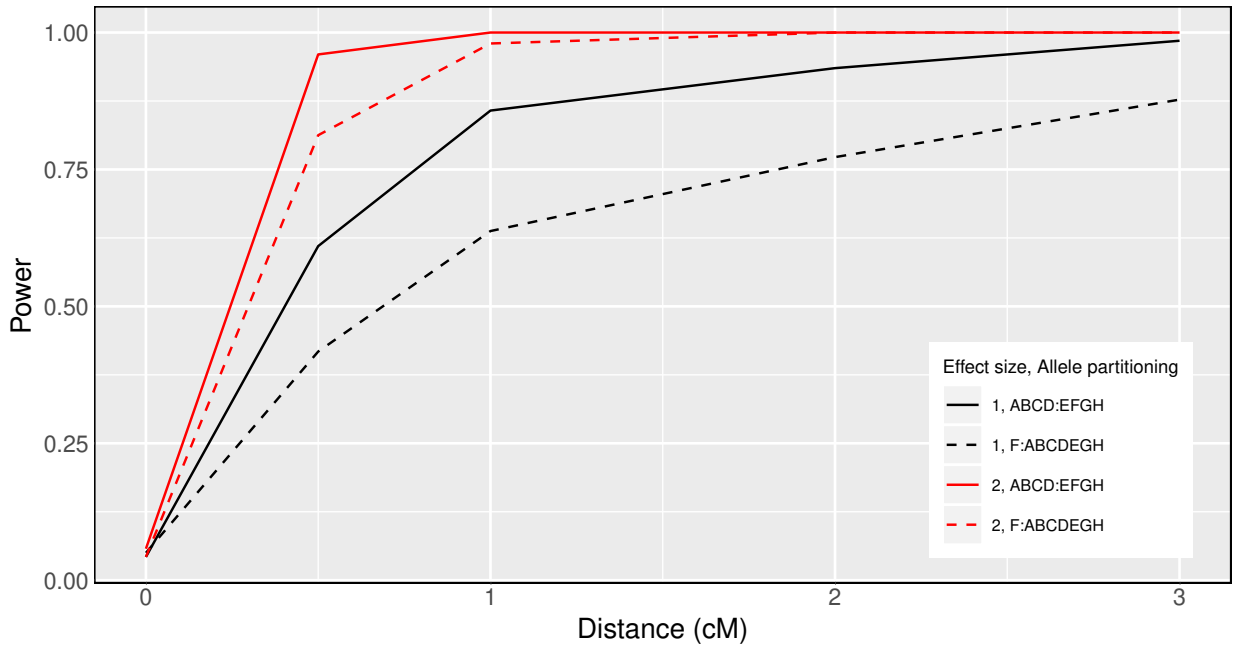


Figure 2.1: Pleiotropy vs. separate QTL power curves for each of four sets of parameter settings. Factors that differ among the four curves are allele effects difference and allele partitioning. Red denotes high allele effects difference, while black is the low allele effects difference. Solid line denotes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

2.4 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 **S1**).

QTL analysis with the LOCO method, and using sex as an additive covariate, showed multiple suggestive QTL for each phenotype (Figure **S2**; Table **S2**). 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.

2.5 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 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

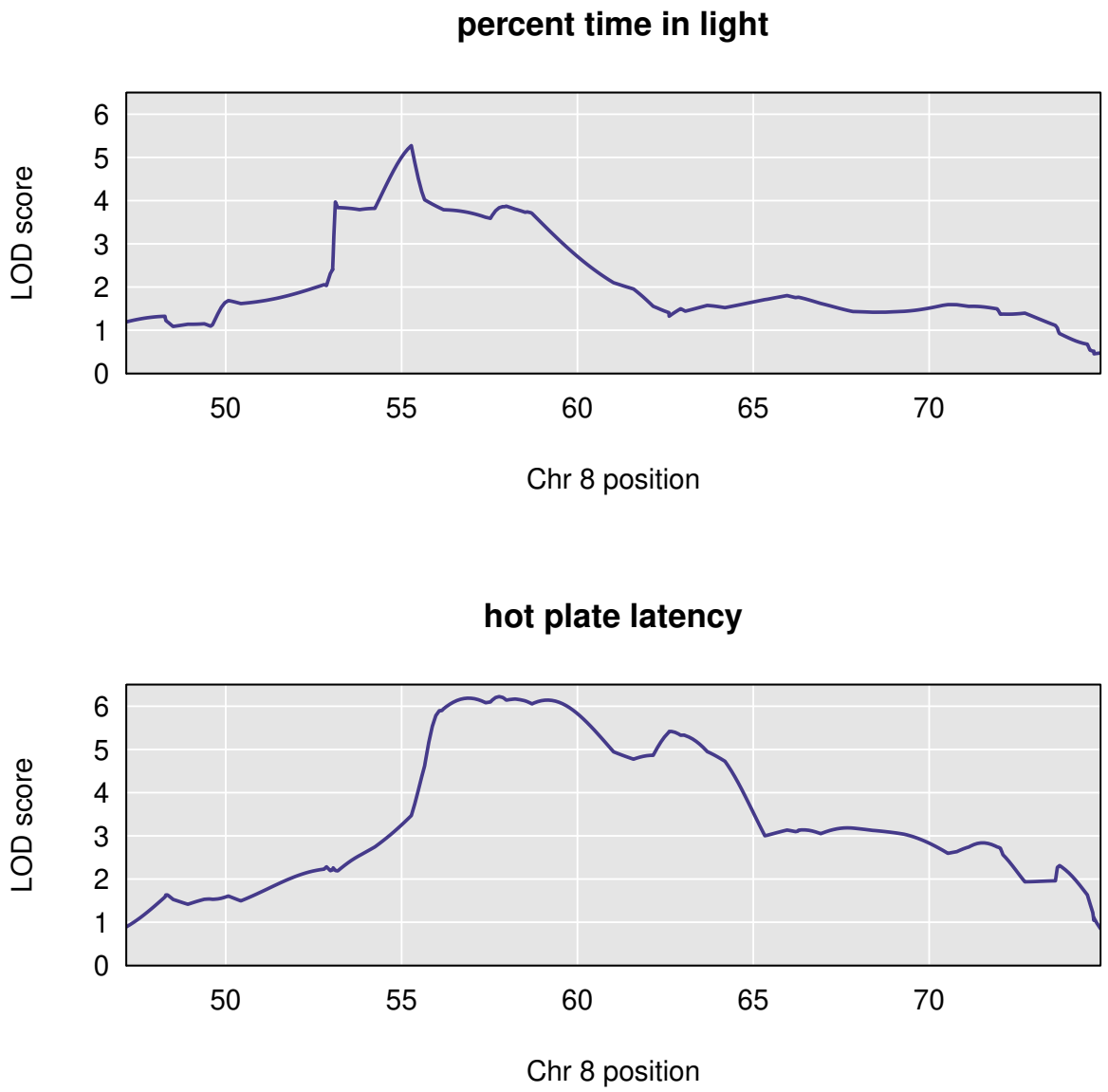


Figure 2.2: Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins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

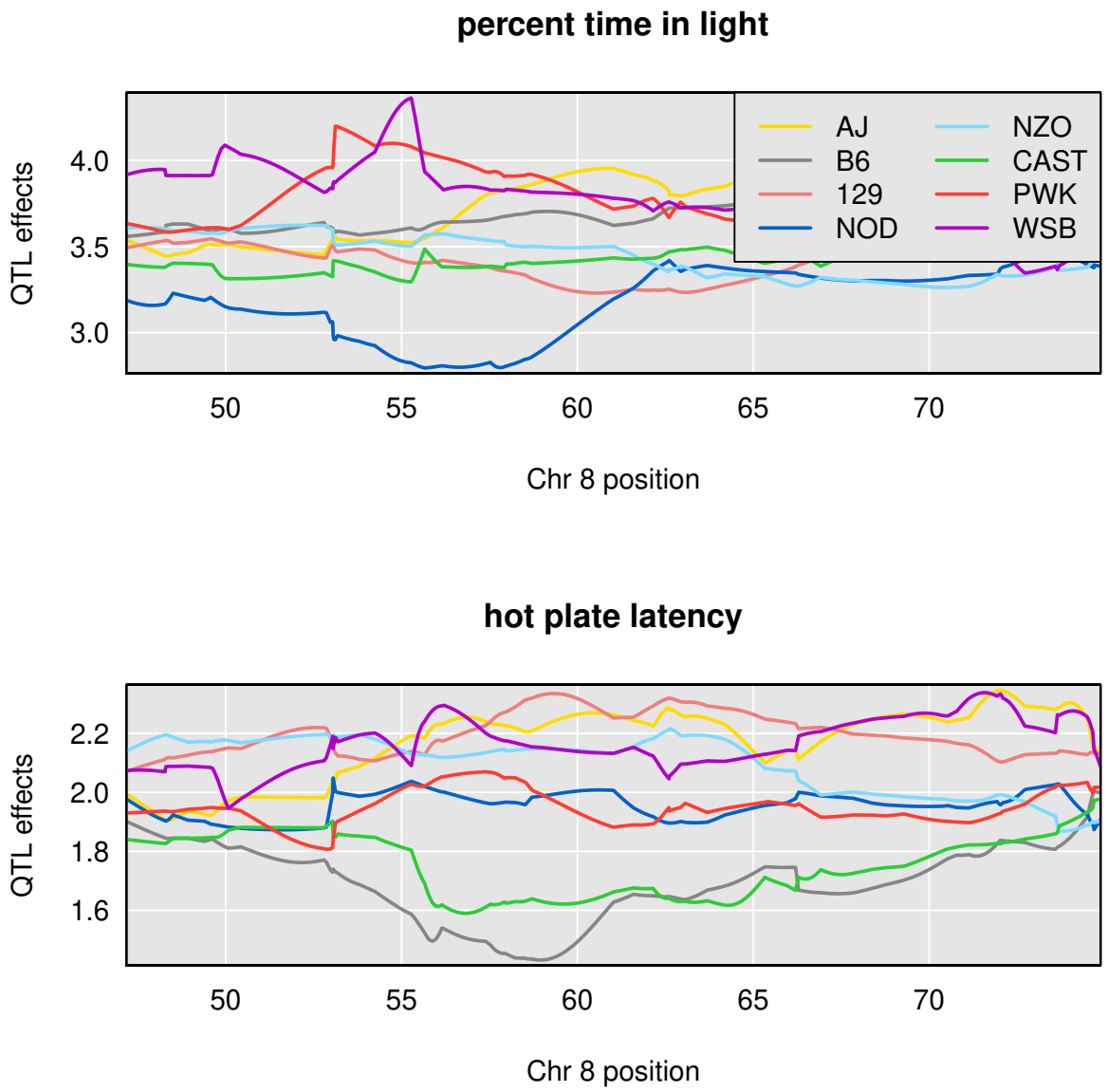


Figure 2.3: Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins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

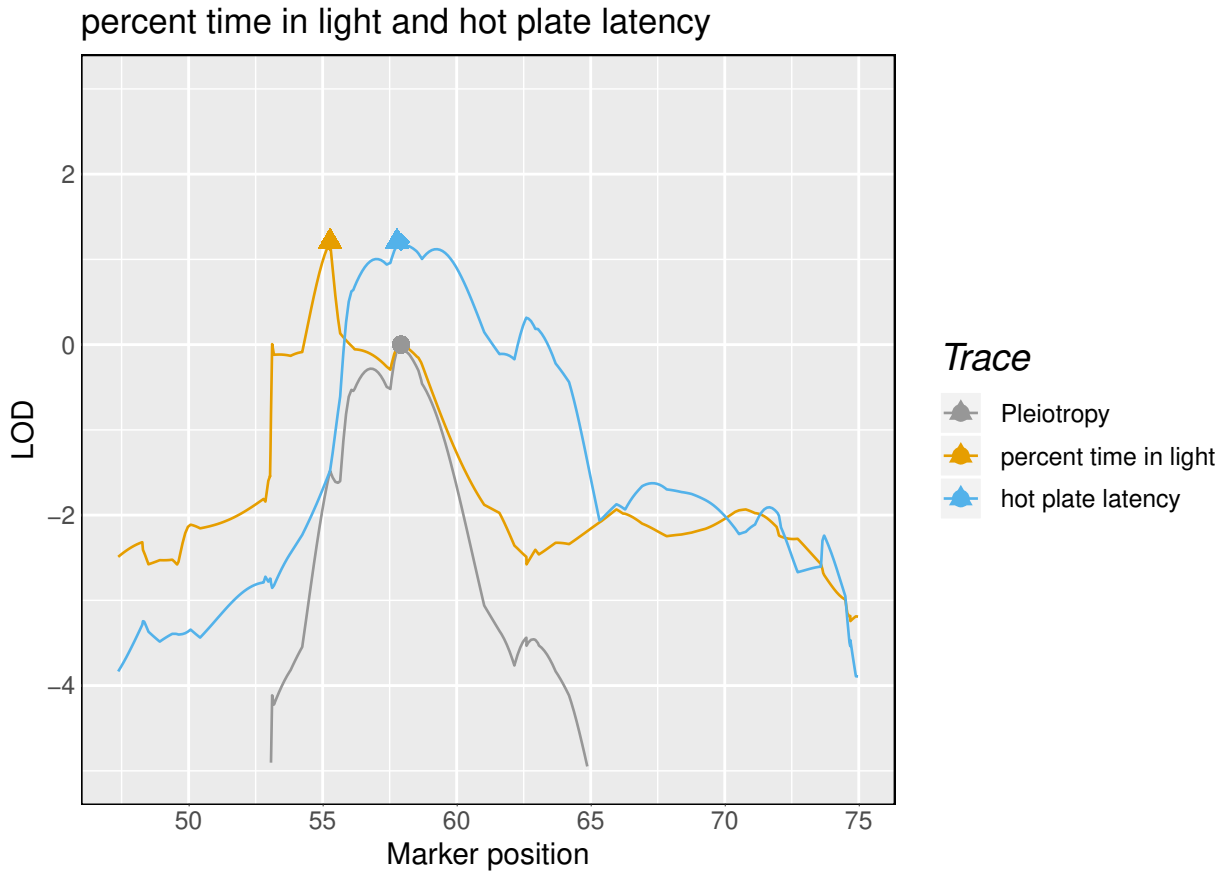


Figure 2.4: Profile LOD curves for the pleiotropy vs. separate QTL hypothesis test for "percent time in light" and "hot plate latency". Gray trace denotes pleiotropy LOD values. Triangles denote the univariate LOD maxima, while diamonds denote the profile LOD maxima. For "percent time in light", the 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, 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; Tisné 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

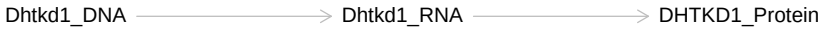


Figure 3.2: A DNA variant in the *Dhtkd1* gene affects *Dhtkd1* transcript abundances which, in turn, affect DHTKD1 protein concentrations.

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

significance 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 that met

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)

three criteria (Frame 3.3).

The 147 nonlocal traits that we studied all had LOD peak heights above 7.18 and LOD peak 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 perform both bivariate QTL scans and mediation analyses of all 13 * 147 = 1911 pairs involving one local expression trait and one nonlocal expression trait.

We used both univariate and bivariate QTL scans. In the univariate scans, we identified LOD peaks for each of 160 (147 nonlocal and 13 local) expression traits that mapped to within 2 Mb of the center of the Chromosome 2 expression trait hotspot. For a given univariate phenotype, we fitted a linear mixed effects model (with polygenic random effect) at every marker across the genome. Keller et al. (2018) performed calculations with the `qtl2` R package (K. W. Broman et al., 2019).

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 * 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.

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 * 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 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 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.

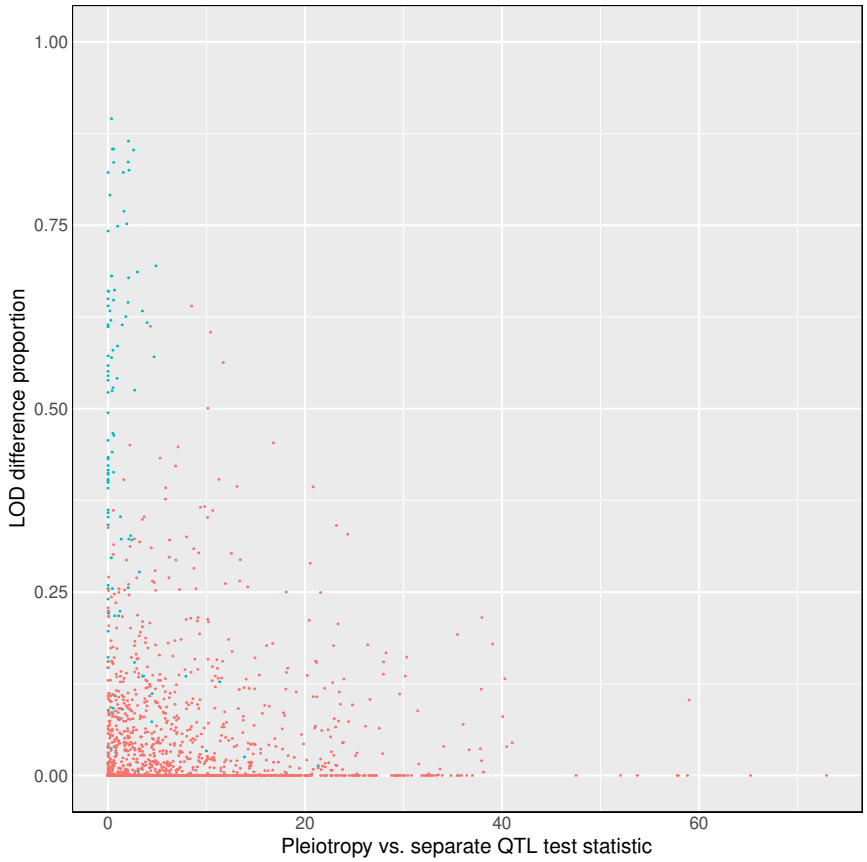


Figure 3.3: LOD difference proportion against pleiotropy vs. separate QTL test statistics for 1911 pairs of traits. Each point represents a single pair. Pairs that involve local expression trait Hnf4a are colored blue, while all others are colored red. Note the high prevalence of blue points in the upper 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.

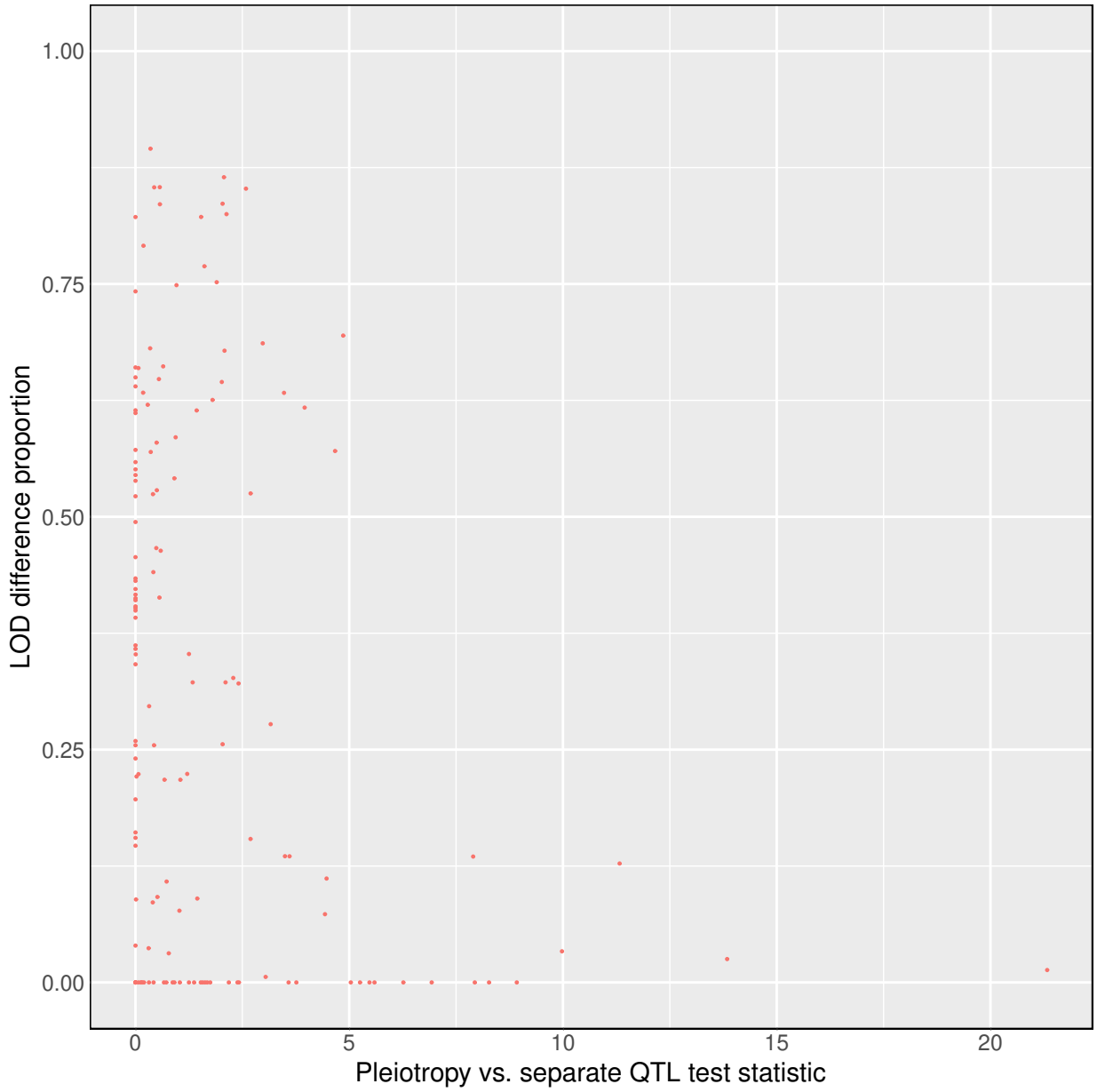


Figure 3.4: Scatter plot of LOD difference proportion against pleiotropy vs. separate QTL test statistic for 147 pairs of traits. Each pair includes *Hnf4a* and one of the nonlocal gene expression traits that map to the Chromosome 2 hotspot.

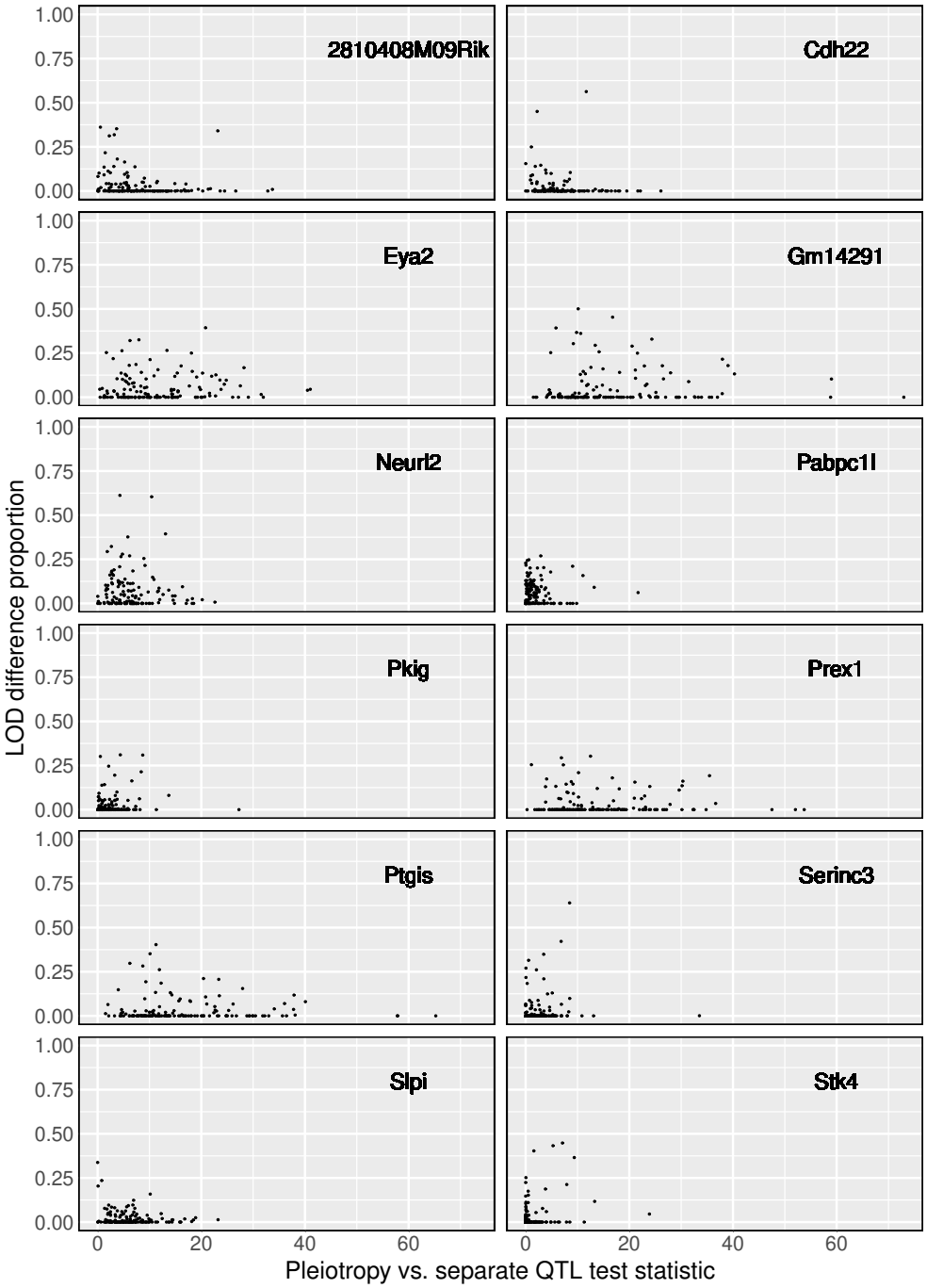
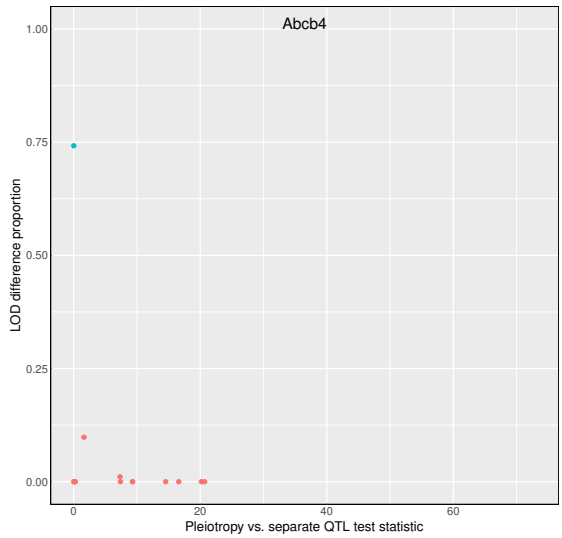
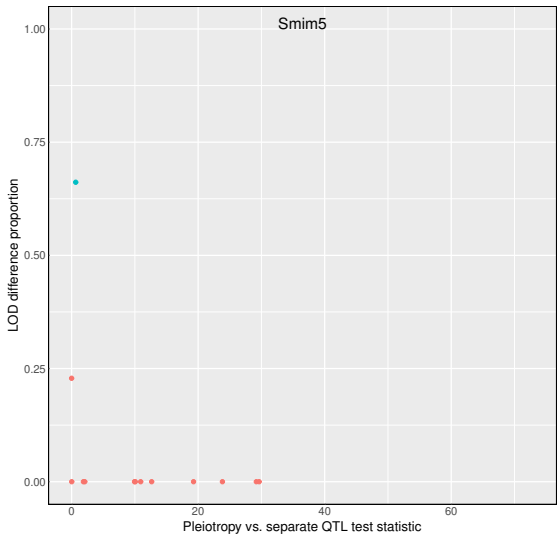


Figure 3.5: Scatter plots of LOD difference proportion against pleiotropy vs. separate QTL test statistic with 147 pairs of traits per panel. Each panel includes a local gene expression trait (per the label in the upper right quadrant) and one of the 147 nonlocal gene expression traits that map to the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

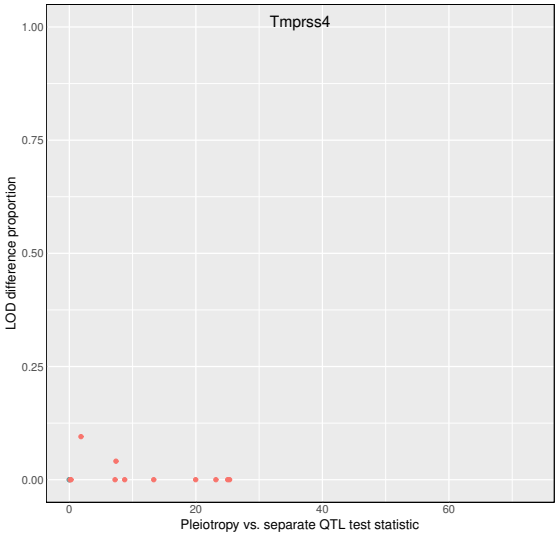
Figure 3.6: Scatter plots for four nonlocal expression traits. Each plot features 13 points, one for each local gene expression trait. The vertical axis denotes LOD difference proportion values, while the horizontal axis corresponds to pleiotropy test statistics. Blue points represent the pairing with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



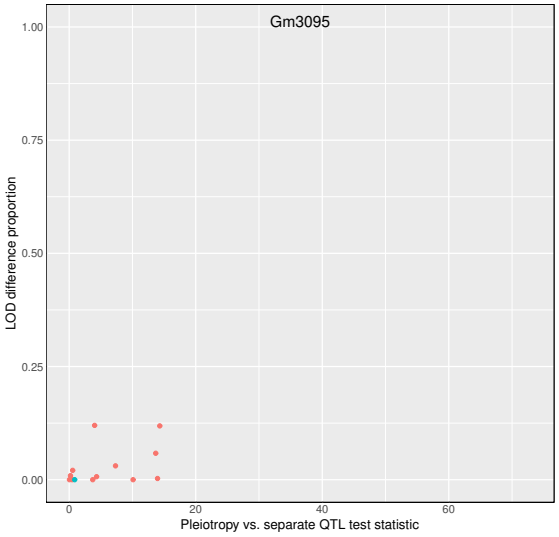
(a) *Hnf4a* transcript levels mediate *Abcb4* transcript levels. The high LOD difference proportion and the very small pleiotropy test statistic together provide evidence that *Hnf4a* mediates *Abcb4* transcript levels.



(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*.

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.

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 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 **S2**).

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

For each Chromosome 19 marker, we estimated founder allele and covariate effects. We calculated:

$$\widehat{(B:C)} = \left((X:W)^T\hat{\Gamma}^{-1}(X:W)\right)^{-1}(X:W)^T\hat{\Gamma}^{-1}Y \tag{3.11}$$

where $B:C$ denotes the concatenation of B and C, and $X:W$ refers to the n by 12 matrix resulting from appending the columns of W to the X matrix. $\hat{\Gamma}$ is the covariance matrix defined by Equation 3.12.

$$\hat{\Gamma} = \hat{\sigma}_g^2 K + \hat{\sigma}_e^2 I_n \tag{3.12}$$

We denote the restricted maximum likelihood estimates of the variance components by $\hat{\sigma}_g^2$ and $\hat{\sigma}_e^2$.

For each of the 80 expression traits, we calculated fitted values for each subject with the estimated founder

allele and covariate effects (Equation 3.11). We then calculated correlations between fitted values for pairs of traits. Each pairing involved one anchor gene expression trait and one other gene expression trait.

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). Macdonald and Long (2007) and King et al. (2012) 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 defined in Chapter 2. These methods are implemented in the R package `qtl2pleio` (Boehm, 2018b).

For each of the 80 expression traits, we used the fitted founder allele and covariate effects ($\widehat{B:C}$ in Equation 3.11) to obtain fitted values vectors for every subject and all 80 traits (Equation 3.13). For each of the 310 pairings of traits, we calculated correlations among the fitted values vectors. Our motivation for working with the fitted values vectors (instead of the $\hat{B}$ estimated founder allele effects vectors) is that the fitted values approximately weight the allele effects by allele frequency. An alternative analysis might neglect the covariates when calculating fitted values.

$$\hat{Y} = X\hat{B} + W\hat{C} \tag{3.13}$$

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 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

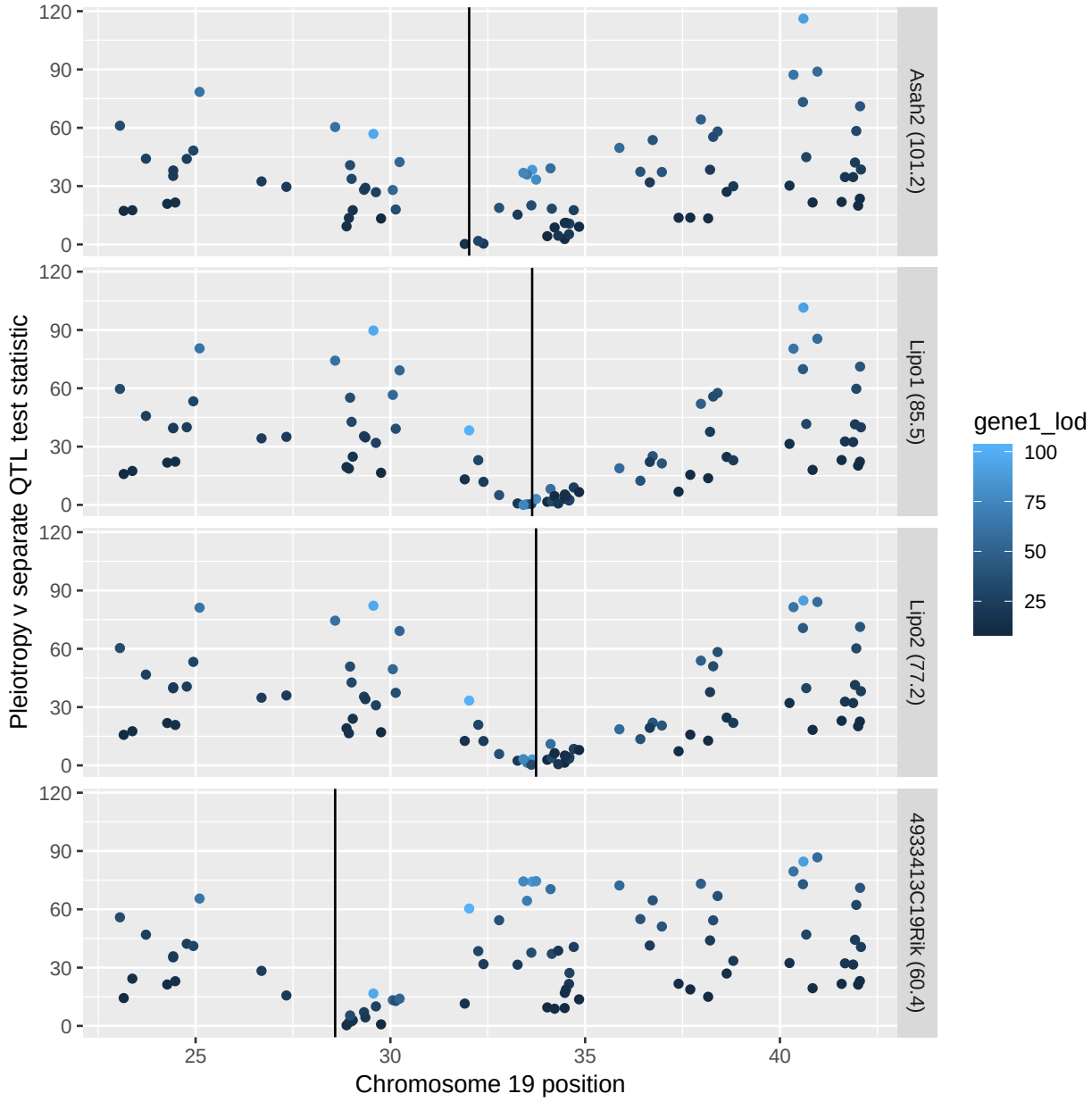


Figure 3.7: Each anchor gene has its own panel. Along the horizontal axis is Chromosome 19 position. The vertical axis is for pleiotropy test statistic value. Each point corresponds to a local gene expression trait (paired with the anchor gene). Point color corresponds to the local gene's univariate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

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 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.

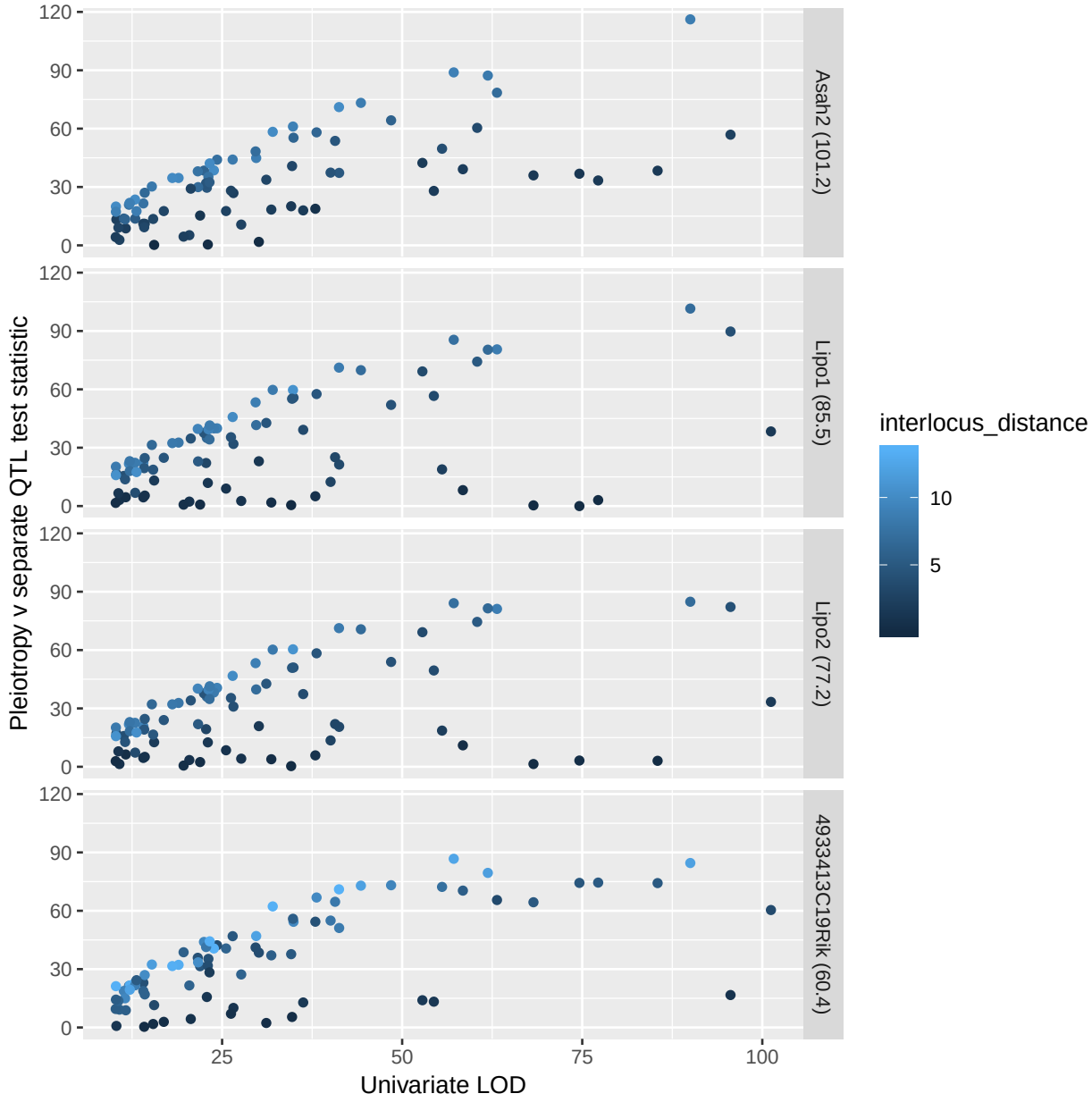


Figure 3.8: Vertical axis denotes pleiotropy test statistic value, while horizontal axis denotes univariate LOD score. Each point corresponds to a single gene expression trait. Panels correspond to the anchor gene expression trait. The pleiotropy test statistics correspond to analyses involving a single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

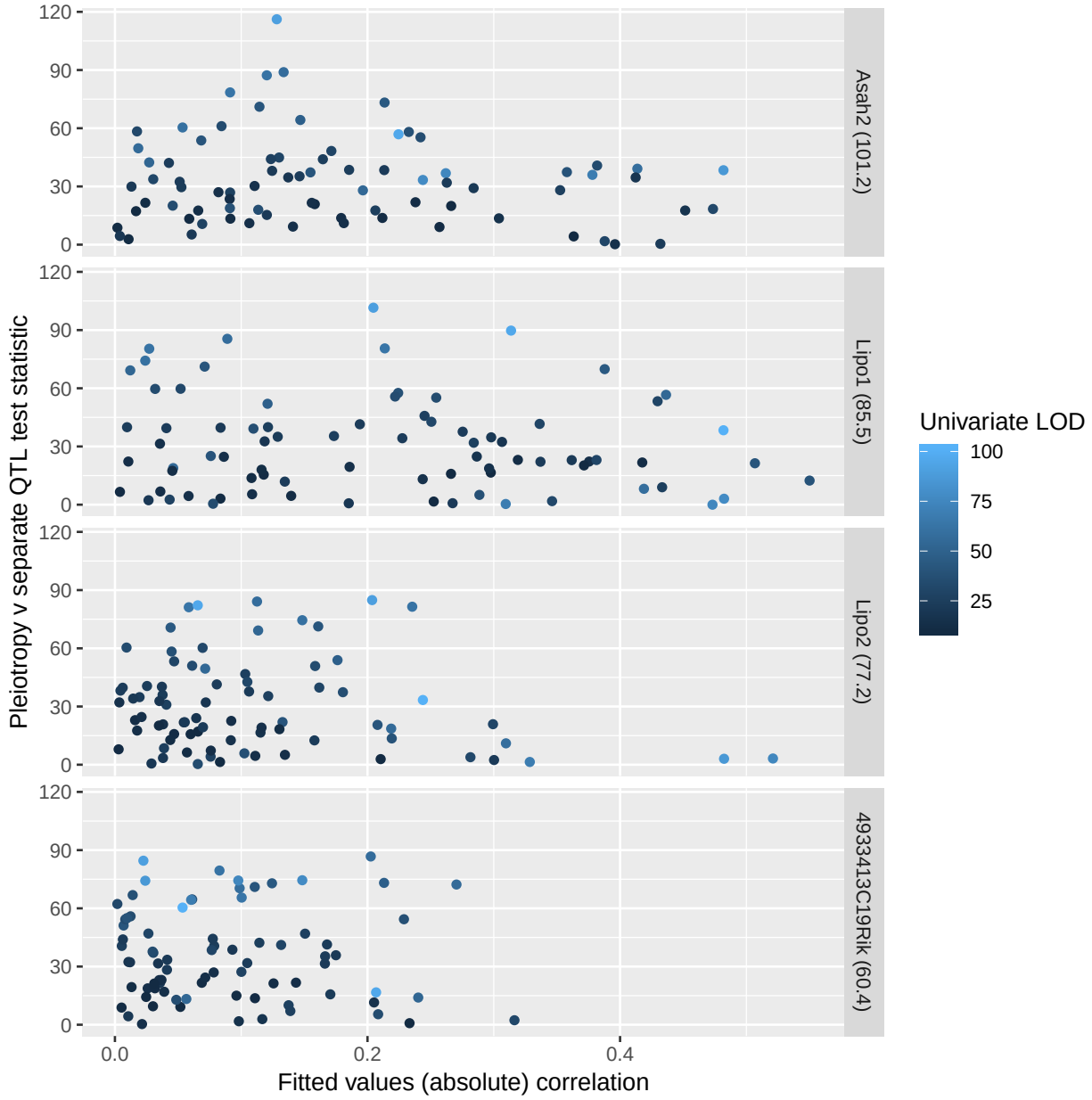


Figure 3.9: Vertical axis denotes the pleiotropy test statistic value, and horizontal axis indicates absolute value of the correlation between vectors of fitted values. Each point corresponds to a pairing between the specified anchor expression trait and one of the 79 other expression traits.

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 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

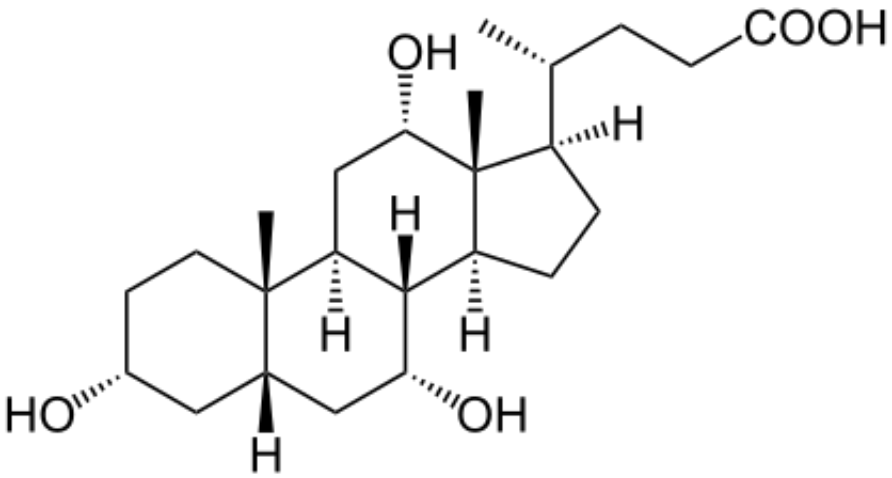


Figure 3.10: Cholic acid chemical structure. From Wikipedia, https://upload.wikimedia.org/wikipedia/commons/2/21/Chols%C3%A4ure.svg.

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. (in preparation), 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 Karl W Broman (2012a) and Karl W Broman (2012b) and implemented in the R package `qtl2` (K. W. Broman 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 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.

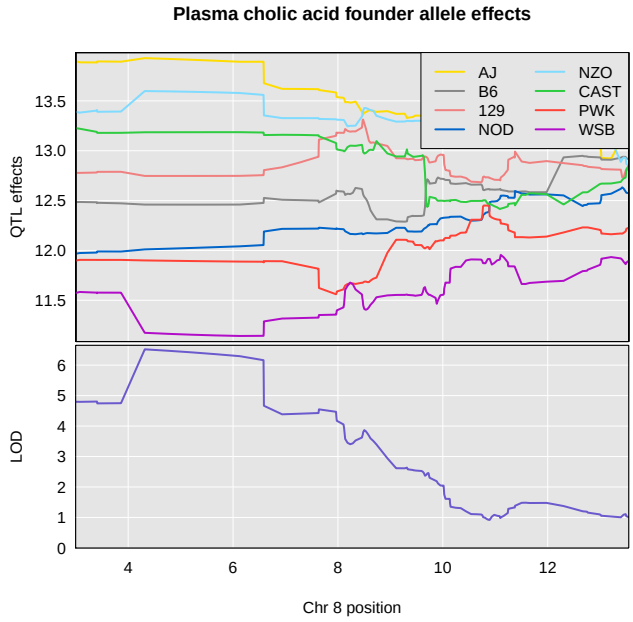
3.3.3 Results

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

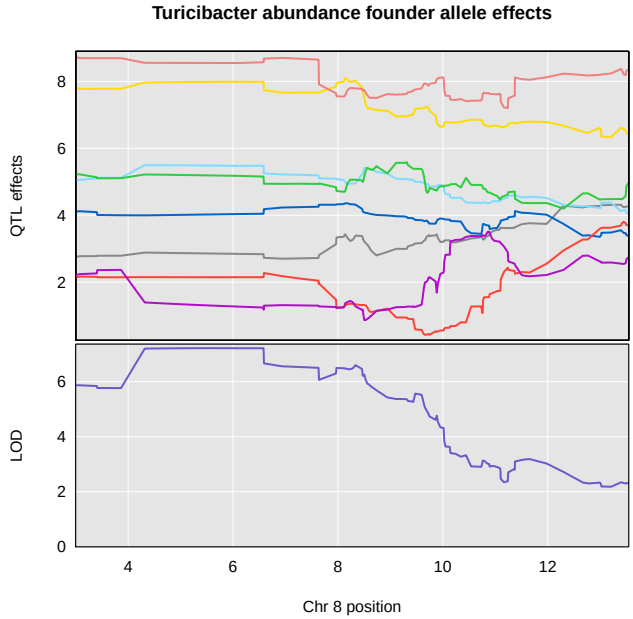
Table 3.4: Genome-wide LOD peaks greater than 5 for plasma cholic acid levels and *Turicibacter* abundance. Both traits map to approximately the same region on Chromosome 8. Chromosome positions are in Mb units.

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 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.



(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.

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.

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.

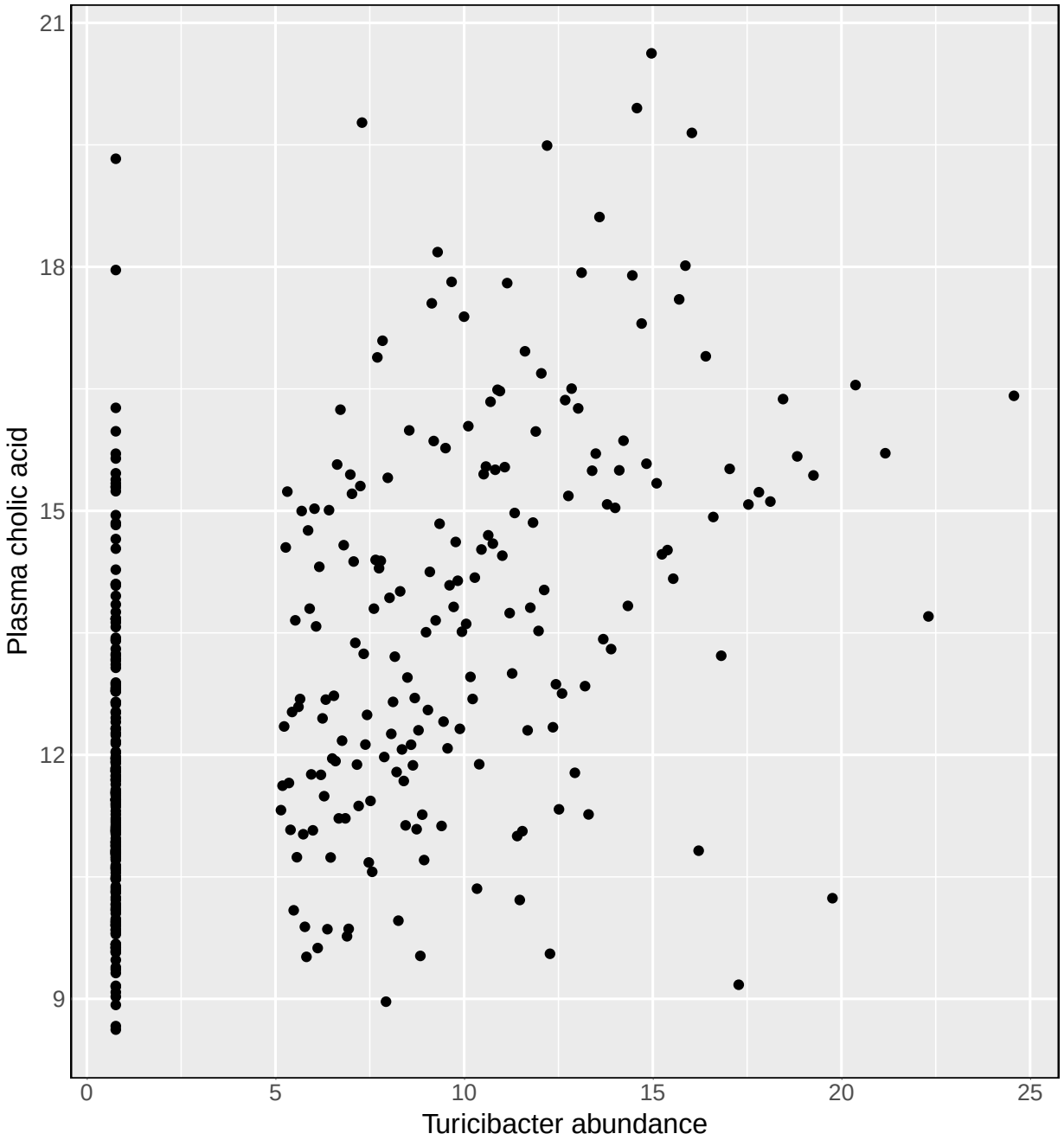


Figure 3.12: Scatter plot of plasma cholic acid levels against *Turicibacter* abundance

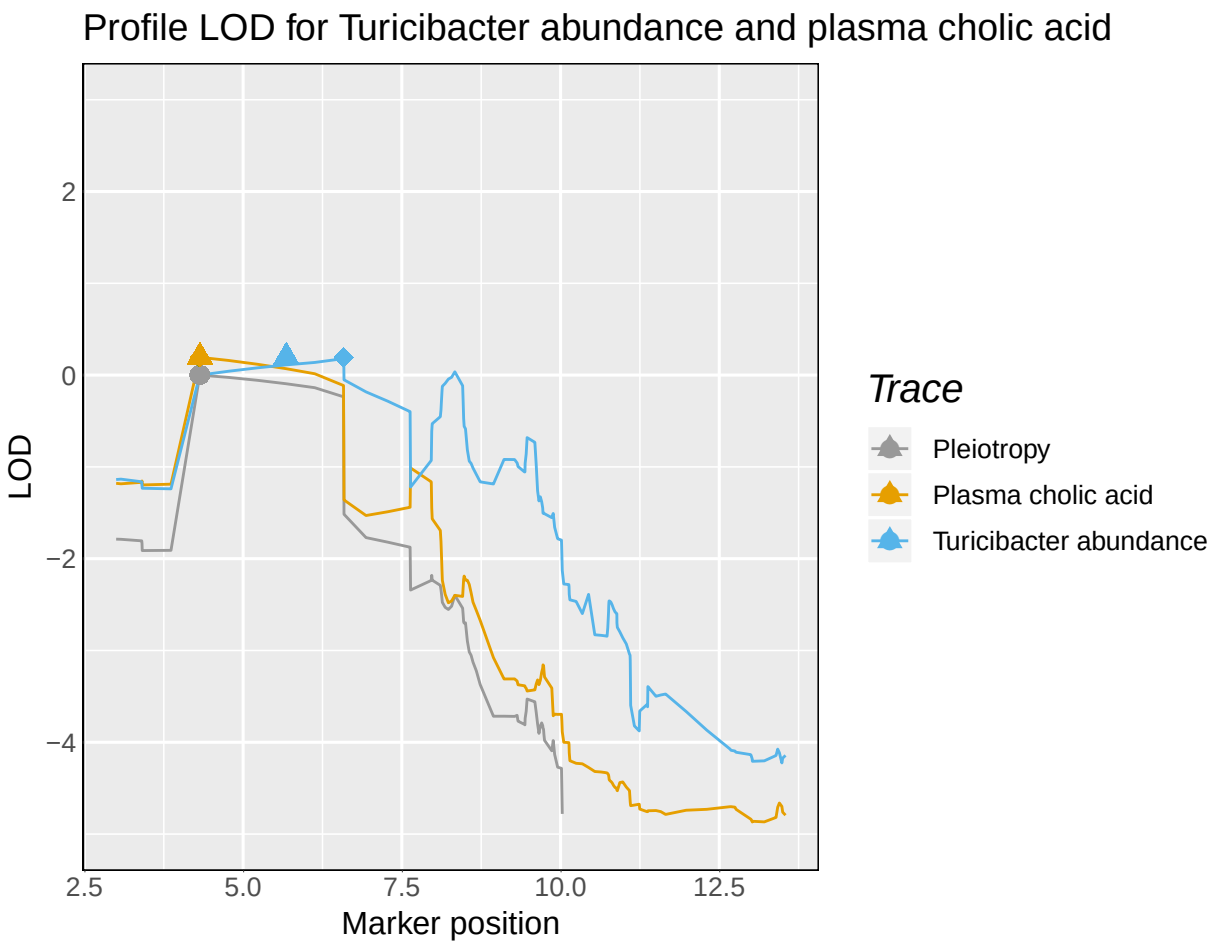


Figure 3.13: Profile LODs for *Turicibacter* abundance and plasma cholic acid levels.

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 vignette

Bibliography

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

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.

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

— (2018b). *qtl2pleio: Hypothesis test of close linkage vs pleiotropy in multiparental populations*. R package version 0.1.0.

Bolyen, Evan et al. (2018). *QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science*. Tech. rep. PeerJ Preprints.

Broman, K. W. et al. (2019). "R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multi-parent populations". In: *Genetics, to appear*.

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

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

Callahan, Benjamin J et al. (2016). "DADA2: high-resolution sample inference from Illumina amplicon data". In: *Nature methods* 13.7, p. 581.

Chesler, Elissa J et al. (2016). "Diversity Outbred Mice at 21: Maintaining Allelic Variation in the Face of Selection". In: *G3: Genes| Genomes| Genetics*, g3–116.

Chick, Joel M et al. (2016). "Defining the consequences of genetic variation on a proteome-wide scale". In: *Nature* 534.7608, p. 500.

Churchill, Gary A, David C Airey, et al. (2004). “The Collaborative Cross, a community resource for the genetic analysis of complex traits”. In: *Nature genetics* 36.11, pp. 1133–1137.

Churchill, Gary A, Daniel M Gatti, et al. (2012). “The diversity outbred mouse population”. In: *Mammalian genome* 23.9-10, pp. 713–718.

Crick, Francis (1970). “Central dogma of molecular biology”. In: *Nature* 227.5258, p. 561.

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

Cubillos, Francisco A et al. (2013). “High-resolution mapping of complex traits with a four-parent advanced intercross yeast population”. In: *Genetics* 195.3, pp. 1141–1155.

DeSantis, Todd Z et al. (2006). “Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB”. In: *Applied and environmental microbiology* 72.7, pp. 5069–5072.

Eddelbuettel, Dirk et al. (2011). “Rcpp: Seamless R and C++ integration”. In: *Journal of Statistical Software* 40.8, pp. 1–18.

Efron, B (1979). “Bootstrap methods: another look at the jackknife”. In: *The Annals of Statistics* 7.1, pp. 1–26.

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.

Jiang, Changjian and Zhao-Bang Zeng (1995). “Multiple trait analysis of genetic mapping for quantitative trait loci.” In: *Genetics* 140.3, pp. 1111–1127.

Kang, Hyun Min et al. (2010). “Variance component model to account for sample structure in genome-wide association studies”. In: *Nature genetics* 42.4, pp. 348–354.

Katoh, Kazutaka and Daron M Standley (2013). “MAFFT multiple sequence alignment software version 7: improvements in performance and usability”. In: *Molecular biology and evolution* 30.4, pp. 772–780.

Keller, Mark P et al. (2018). “Genetic Drivers of Pancreatic Islet Function”. In: *Genetics*, genetics–300864.

Kemis, Julia H. et al. (in preparation). “Genetic determinants of gut microbiota composition and bile acid profiles in mice”. In:

King, Elizabeth G et al. (2012). “Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource”. In: *Genome research*, gr–134031.

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

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.

Kover, Paula X et al. (2009). “A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana”. In: *PLoS genetics* 5.7, e1000551.

Logan, Ryan W 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.

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.

Mackay, Trudy FC et al. (2012). “The Drosophila melanogaster genetic reference panel”. In: *Nature* 482.7384, p. 173.

Meyer, Hannah Verena et al. (2018). “LiMMBo: a simple, scalable approach for linear mixed models in high-dimensional genetic association studies”. In: *bioRxiv*, p. 255497.

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.

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

Morgan, Andrew P et al. (2015). “The mouse universal genotyping array: from substrains to subspecies”. In: *G3: Genes| Genomes| Genetics*, g3–115.

— (2016). “The mouse universal genotyping array: from substrains to subspecies”. In: *G3: Genes, Genomes, Genetics* 6.2, pp. 263–279.

Neto, Elias Chaibub et al. (2013). “Modeling causality for pairs of phenotypes in system genetics”. In: *Genetics* 193.3, pp. 1003–1013.

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

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

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.

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/.

Raghavan, Chitra et al. (2017). "Approaches in Characterizing Genetic Structure and Mapping in a Rice Multiparental Population". In: *G3: Genes, Genomes, Genetics* 7.6, pp. 1721–1730.

Recla, Jill M et al. (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.

Schadt, Eric E et al. (2005). "An integrative genomics approach to infer causal associations between gene expression and disease". In: *Nature genetics* 37.7, p. 710.

Stanley, Patrick D et al. (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.

Tian, Jianan et al. (2016). "The dissection of expression quantitative trait locus hotspots". In: *Genetics* 202.4, pp. 1563–1574.

Tisné, Sébastien et al. (2017). "Identification of Ganoderma disease resistance loci using natural field infection of an oil palm multiparental population". In: *G3: Genes, Genomes, Genetics* 7.6, pp. 1683–1692.

Tyler, Anna L, Leah Rae Donahue, et al. (2016). "Weak epistasis generally stabilizes phenotypes in a mouse intercross". In: *PLoS genetics* 12.2, e1005805.

Tyler, Anna L, Bo Ji, et al. (2017). "Epistatic networks jointly influence phenotypes related to metabolic disease and gene expression in diversity outbred mice". In: *Genetics* 206.2, pp. 621–639.

Tyler, Anna L, Wei Lu, et al. (2013). "CAPE: an R package for combined analysis of pleiotropy and epistasis". In: *PLoS computational biology* 9.10, e1003270.

VanderWeele, Tyler (2015). *Explanation in causal inference*.

Yang, Jian et al. (2014). "Advantages and pitfalls in the application of mixed-model association methods". In: *Nature genetics* 46.2, pp. 100–106.

Yu, Jianming et al. (2008). "Genetic design and statistical power of nested association mapping in maize". In: *Genetics* 178.1, pp. 539–551.

Zeng, Zhao-Bang et al. (2000). "Genetic architecture of a morphological shape difference between two Drosophila species". In: *Genetics* 154.1, pp. 299–310.

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.

Appendices

Appendix A

Supplementary materials for Chapter 2

Table **S1**: 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 **S2**: 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

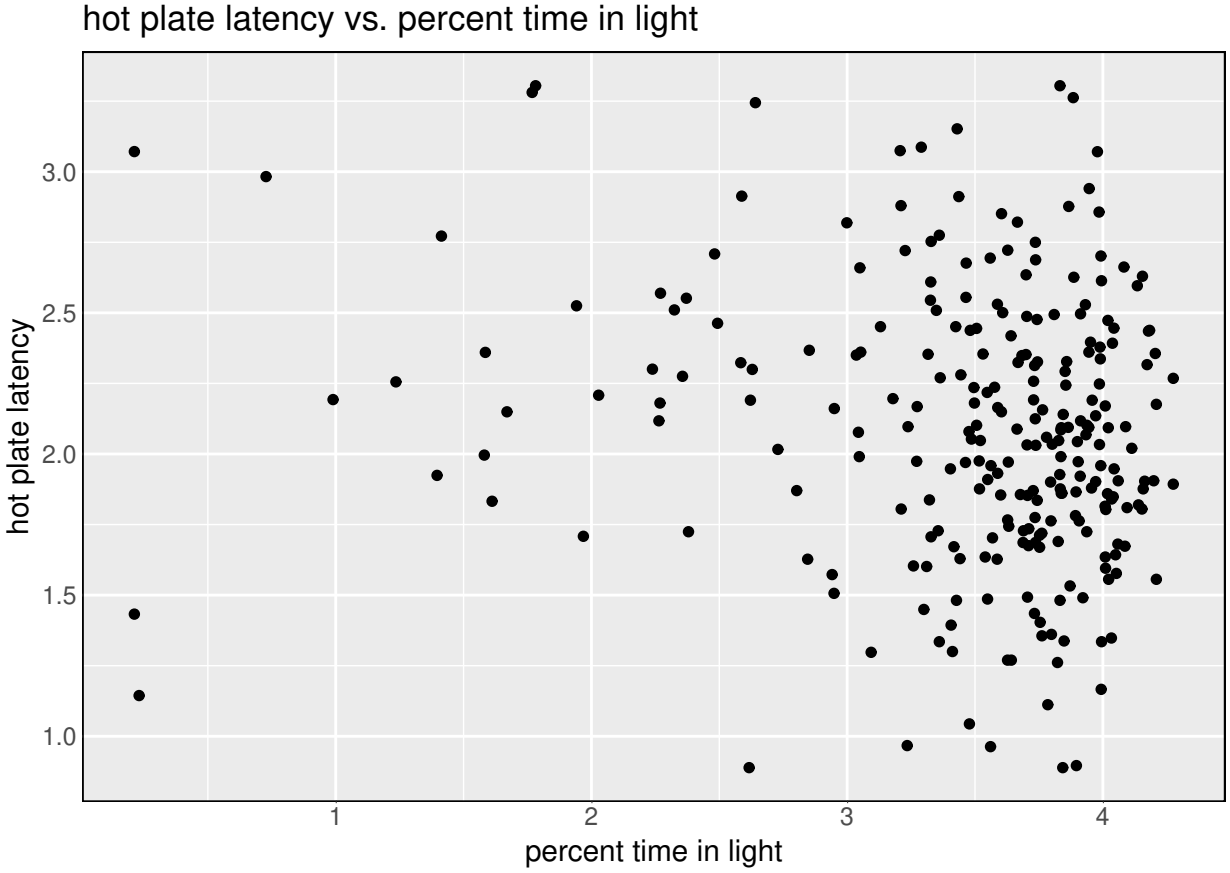


Figure **S1**: Scatter plot of "hot plate latency" against "percent time in light", after applying logarithm transformations and winsorizing both traits.

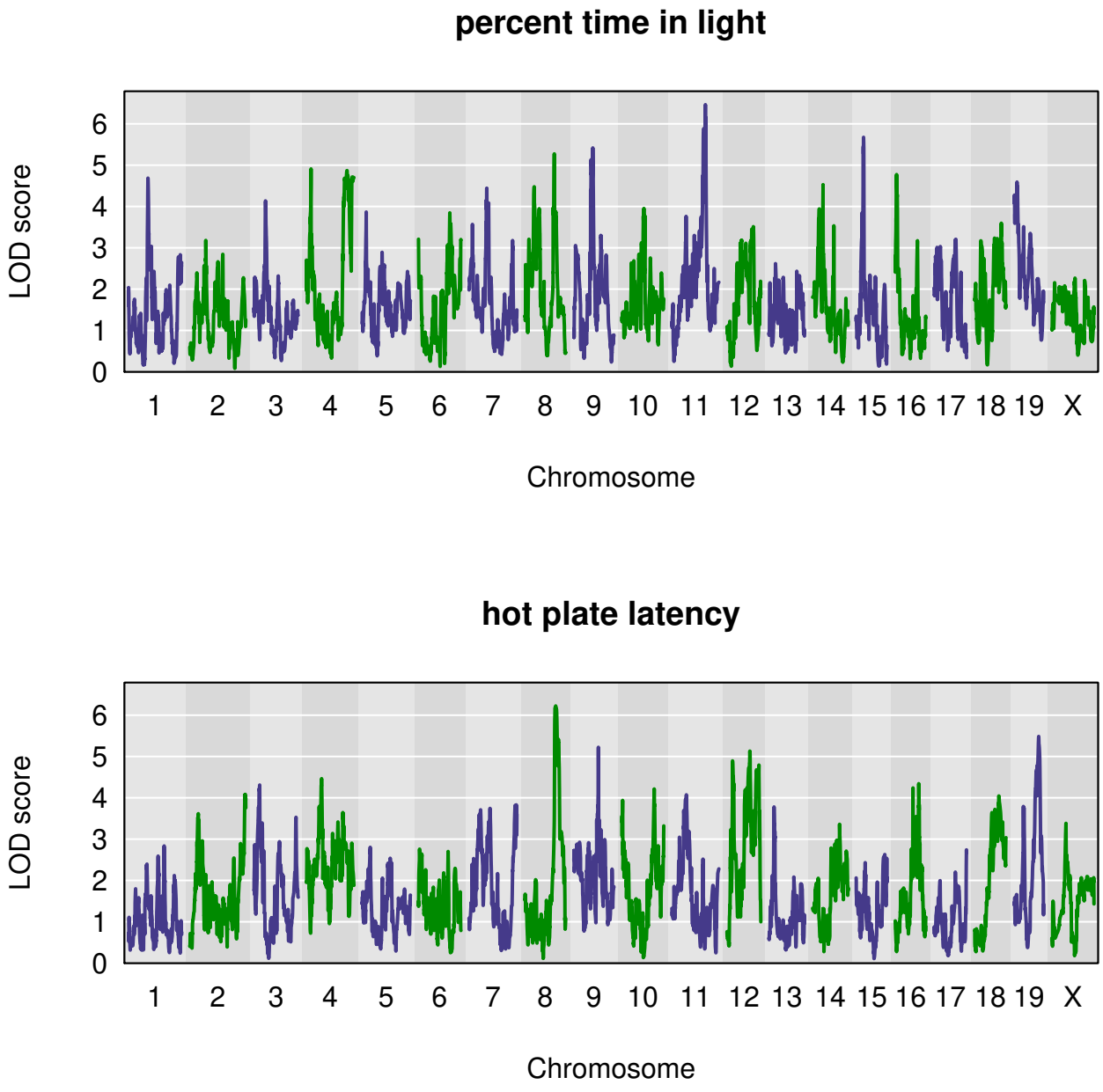


Figure **S2**: Genome-wide QTL scan for percent time in light reveals multiple QTL, including one on Chromosome 8.

Appendix B

Supplementary materials for Chapter 3

Table **S1**: 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 below the 7.18 genome-wide threshold. We believe that this is due to variations 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 **S2**: 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