

Chapter 3B

Frederick Boehm

January 1, 2019

Todo list

■ add annotation info for Asah2 & its eQTL	3
■ check R code - did I check for distance between QTL peak and gene start site? ie, when defining "local gene" QTL peaks	3
■ Did we choose the 20-Mb region to be centered on the midpoint of the Asah2 gene? Check the R code!	3
■ add peak position	3
■ Add 1. annotations table for 4 anchor genes; 2. annotations table for 76 other (not anchor) genes .	3
■ allele effects plots for the four anchor genes. Make them as separate figures in R, then, in latex, combine into a multi-panel figure.	3
■ see R code for details of scan region, etc. How many markers in scan region? What are the start and end positions for the scan?	4

Contents

1 Introduction	2
2 Methods	2
2.1 Data description	2
2.2 Study design	2
2.3 Calculating pleiotropy vs. separate QTL likelihood ratio test statistics	4
2.4 Estimating founder allele effects	4
2.5 Univariate QTL LOD calculations	4

3 Results	4
3.1 Allele effects plots for the four anchor genes	4
3.2 Pleiotropy LRT vs. chromosomal position	4
3.3 Pleiotropy likelihood ratio test statistics vs. univariate LOD	4
3.4 Pleiotropy LRT vs. fitted values correlation	4
4 Discussion	4
5 References	4

List of Figures

1 Caption	5
2 Caption	6
3 Caption	7

List of Tables

1 Introduction

The goal of this section is to characterize the statistical power of our pleiotropy vs. separate QTL 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 chapter 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.

2 Methods

2.1 Data description

2.2 Study design

The goal of the study is to examine statistical power of our test of pleiotropy v separate QTL. To do this, we focus on 80 local expression QTL and their corresponding transcript levels. Here, 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

add annotation info for Asah2 & its eQTL

. Thus, we term the Chromosome 19 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.

We use pancreatic islet gene expression traits from a publicly available data set, which Keller et al. (2018) first collected and analyzed. We examine a collection of 80 local traits on Chromosome 19 and perform our test for pleiotropy vs. separate QTL on pairs of traits.

check R code - did I check for distance between QTL peak and gene start site? ie, when defining "local gene" QTL peaks

Our design involves us first choosing a set of four "anchor" expression traits. Each "anchor" expression trait has a univariate LOD score above 10. The corresponding genes are located near the center of Chromosome 19. We then perform tests of pleiotropy v separate QTL for it paired with each other local trait that has LOD score above 10 and is within 10 Mb of the peak position for the anchor gene Asah2. Gene Asah2 is located near the center of Chromosome 19 and has a local expression QTL with a peak at

-

add peak position

on Chromosome 19. We identified a set of 79 other expression traits that map to the 20-Mb region centered on the peak for Asah2, at

-

Mb. Each trait among the 79 maps to Chromosome 19 with a univariate LOD of at least 10.

In addition to Asah2, we chose 3 other "anchor" genes. These are: Lipo1, Lipo2, and 4933413C19Rik. We chose these because they represent distinct allele effects patterns despite being near the Asah2 gene position.

Did we choose the 20-Mb region to be centered on the midpoint of the Asah2 gene? Check the R code!

Add 1. annotations table for 4 anchor genes; 2. annotations table for 76 other (not anchor)

allele effects plots for the four anchor genes. Make them as separate figures in R, then, in latex, combine into a multi-panel figure.

2.3 Calculating pleiotropy vs. separate QTL likelihood ratio test statistics

see R code for details of scan region, etc. How many markers in scan region? What are the start and end positions for the scan?

2.4 Estimating founder allele effects

2.5 Univariate QTL LOD calculations

For each trait in our analysis, we estimated founder allele effects vector at the local trait's peak position.

3 Results

3.1 Allele effects plots for the four anchor genes

3.2 Pleiotropy LRT vs. chromosomal position

3.3 Pleiotropy likelihood ratio test statistics vs. univariate LOD

3.4 Pleiotropy LRT vs. fitted values correlation

4 Discussion

5 References

References

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

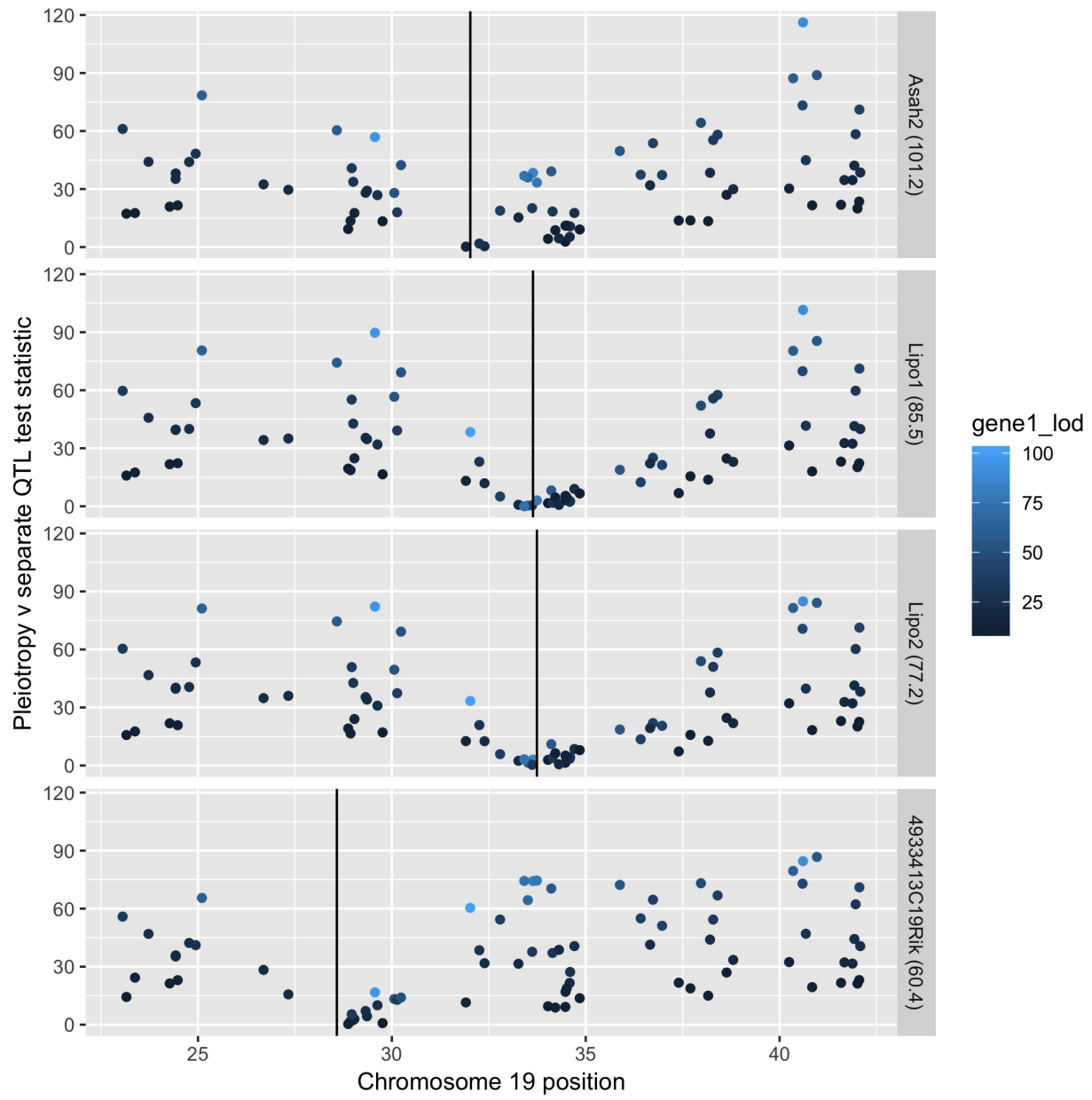


Figure 1: Caption

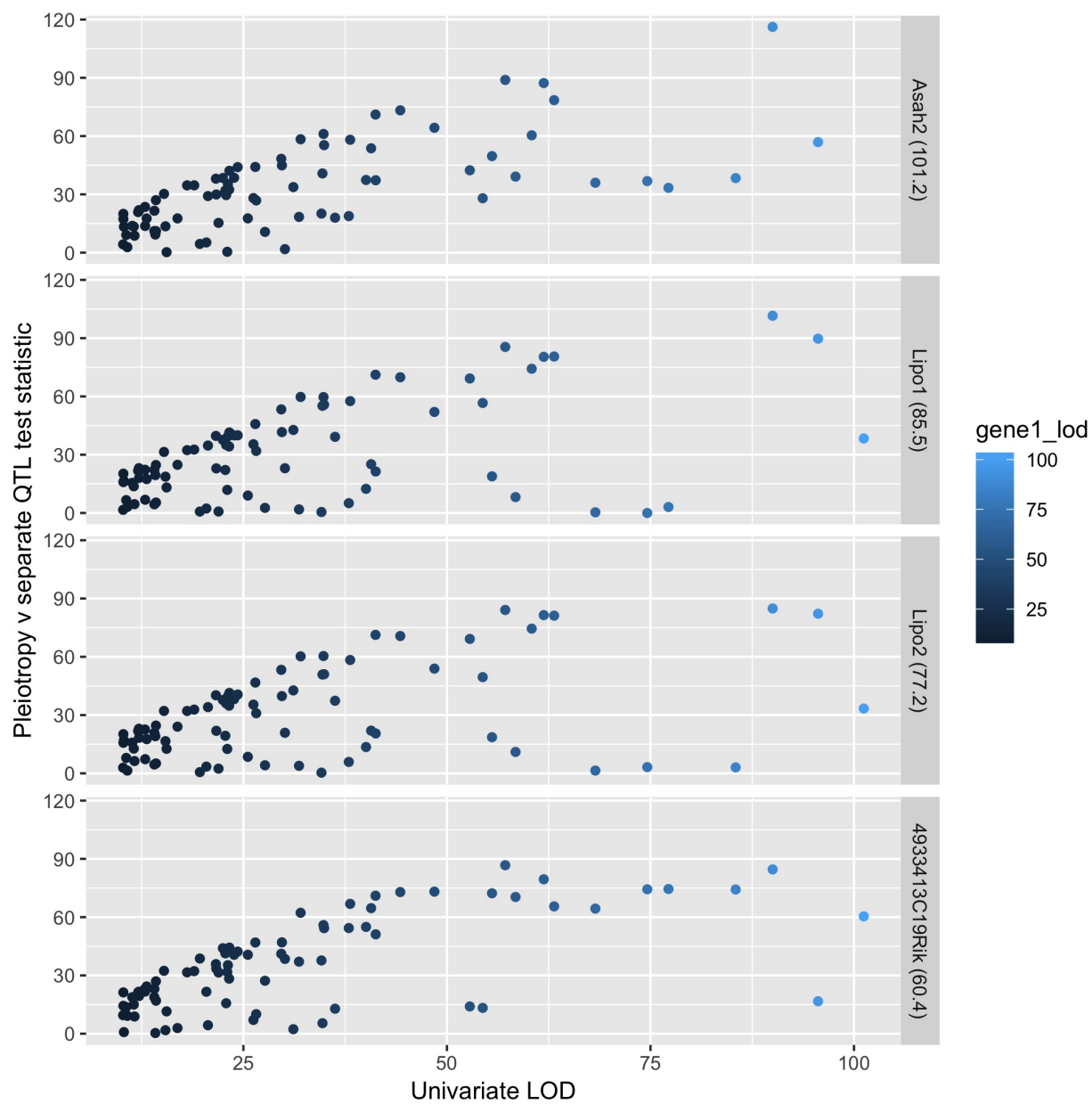


Figure 2: Caption

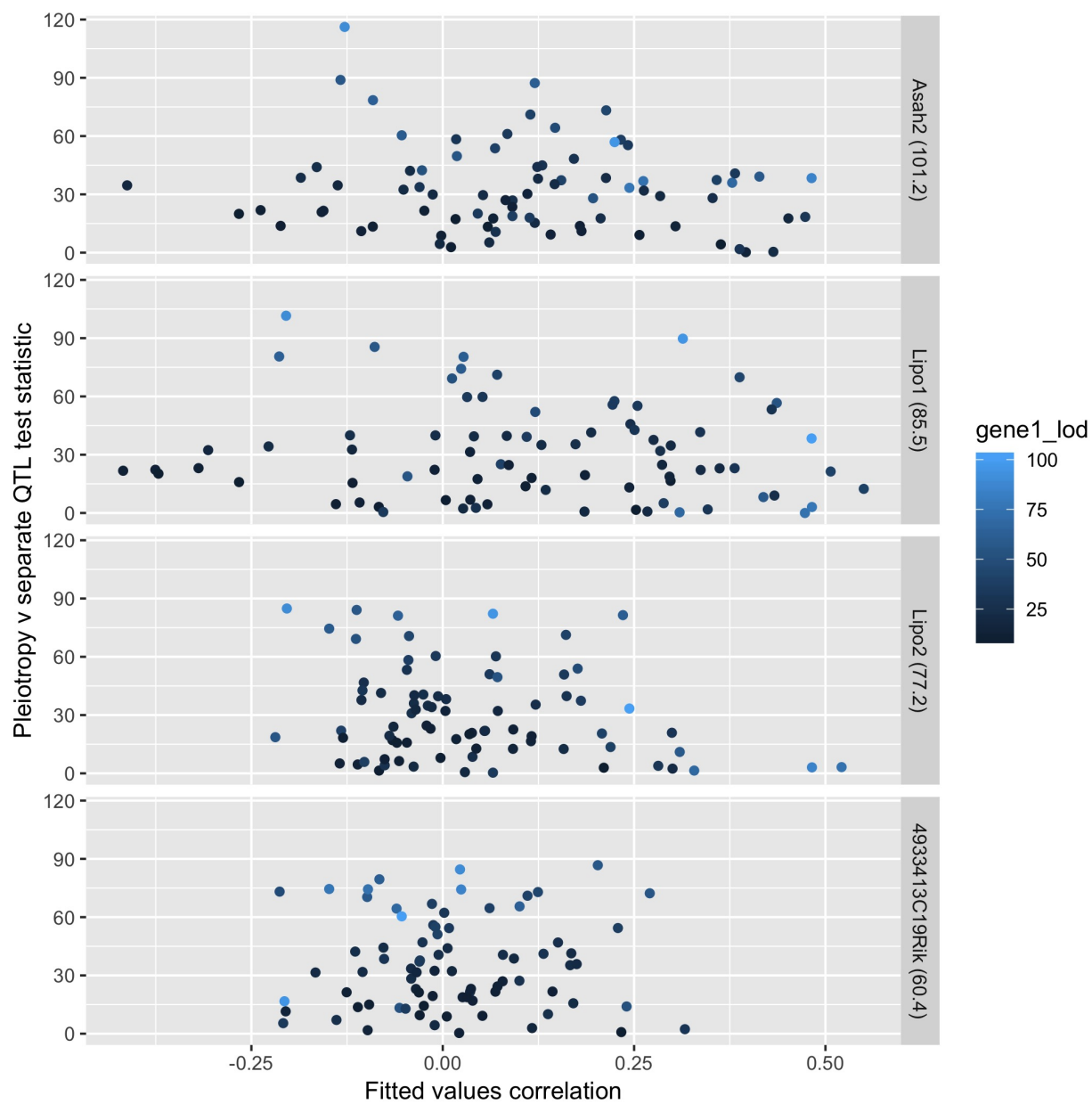


Figure 3: Caption