Expression QTL analysis in expression genetic data

Shilu Zhang

University of Wisconsin, Madison

Research Questions

1. Are there additive allele effects or dominance effects in the quantitative trait loci (QTL) analysis?
2. Are the effects of QTL different in the two sexes?
3. Do some pairs of tissues have similar effects of QTL?
4. If it is expression QTL in one tissue, is it more likely to be expression QTL in other tissues?

Introduction

Identifying the genetic loci that contribute to variation in quantitative traits in nature is an important problem to biologists. Studying the effects of such quantitative trait loci (QTLs) can help understand the biological basis and evolution of these traits. Expression quantitative trait loci (eQTL) refer to the genomic locations that influence variation in gene expression levels (mRNA abundances) [3]. eQTL near the genomic location of the influenced gene are called local eQTL, and eQTL far away from the influenced gene are called trans-eQTL. Some of the eQTL hotspots show effects in multiple tissues, whereas some are specific to one tissue.

This study is aim to identify genes and pathways that cause obesity-induced type II diabetes. Two different mouse lines: C57BL/6J (abbreviated B6 or B) and BTBR T^{+} *tf*/J (abbreviated BTBR or R) mice were used to conduct an intercross. B6 mice are resistant to diabetes, while BTBR mice are susceptible to diabetes. BTBR females were crossed to B6 males to generate F_1 heterozygous parents, and then F_2 $Lep^{ob/ob}$ mice. F_2 mice were genotyped with the 5K GeneChip (Affymetrix). There were three genotypes in F2 mice, BB (homozygote B6), BR (heterozygote) and RR (homozygote BTBR). Six tissues from F_2 mice were considered for expression profiling: adipose, gastrocnemius muscle, hypothalamus, pancreatic islets, liver, and kidney.

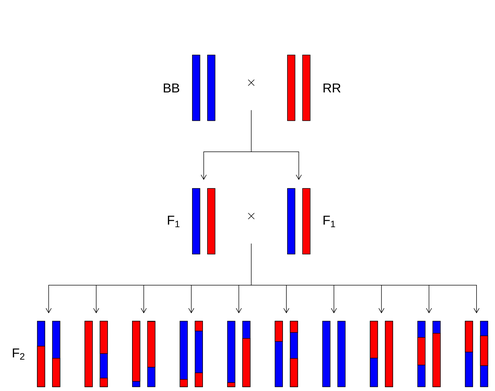

Figure 1: Schematic representation of the autosomes in an intercross experiment (from Broman KW et al 2009) [1]

Materials and Methods

1. Haley-Knott regression

Let B and R denote the two alleles in the cross, and let μ_{BB}, μ_{BR} and μ_{RR} denote the average expression levels for genotypes BB, BR, and RR, respectively. The estimated additive effect is half the difference between the two homozygotes BB and RR:

$$\hat{a} = \frac{(\hat{\mu}_{RR} - \hat{\mu}_{BB})}{2} \tag{1}$$

The estimated dominance effect is the difference between the heterozygote BR and the midpoint between the two homozygotes (BB, RR):

$$\hat{d} = \hat{\mu}_{BR} - \frac{(\hat{\mu}_{RR} + \hat{\mu}_{BB})}{2} \tag{2}$$

For a given QTL model, we have

$$Y = X\beta + \varepsilon \tag{3}$$

where Y is an $n \times 1$ vector of phenotypes, with n as the number of F_2 individuals. X is an $n \times 3$ matrix of covariates, with $(P_{RR} - P_{BB})/2$ in the second column, and $P_{BR} - (P_{RR} + P_{BB})/2$ in the third column. P_{BR}, P_{RR} and P_{BB} are conditional genotype probabilities given observed multipoint marker genotype data. Thus, $\hat{\beta}_1$ is the estimated additive effect $\hat{a}$ and $\hat{\beta}_2$ is the estimated dominance effect $\hat{d}$.

We then plot $\hat{d}$ vs $\hat{a}$ for all expression traits. If $d \approx 0$, the effect is additive. If $d \approx a$, R is dominant. If $d \approx -a$, B is dominant.

2. QTL analysis

We performed single-QTL genome scans separately for each probe in each tissue, by Haley-Knott regression [2] with sex included as an interactive covariate. Calculations were performed at the genetics markers and at a set of pseudomarkers inserted into marker intervals, selected so that adjacent positions were separated by less than 0.5 cM. We calculated conditional genotype probabilities, given observed multipoint marker genotype data, using a hidden Markov mdoel assuming a genotyping error rate of 0.2%.

3. Simulations

We use R^2 from Haley-Knott regression as the criteria to infer eQTL. If R^2 is larger then 0.1, we say there is eQTL in the tissue. To identify the effects of eQTL in different tissues, we generate binary data indicating if there is eQTL or not for each probe in each tissue. We assume the distribution of eQTL ($X_i = 1$) and non-eQTL ($X_i = 0$) in tissue i is independent Bernoulli with probability p_i (Table 1). What is the distribution of number of tissues with eQTL, $\sum_{k=1}^{6} X_i$? We did 100,000 simulations using independent Bernoulli distribution with the probability of eQTL in tissue i, p_i (Table 1). Then, we compare the simulated distribution of $\sum_{k=1}^{6} X_i$ to true distribution.

	adipose	islet	liver	kidney	hypo	gastroc
p_i	0.056	0.064	0.042	0.062	0.059	0.046

Table 1: The probability of eQTL in each tissue

Results

1. Estimate the effects of QTL

From the scatterplots, we can see that the effects of QTL are nearly additive because most of the data points are distributed on $d = 0$ line.

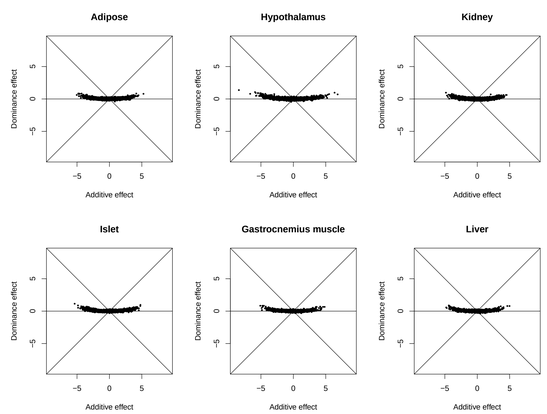

Figure 2: Scatterplots of the estimated dominance effects versus the estimated additive effects

2. Sex differences

The correlation between estimated additive effects in females and males for each tissue is shown in Table 2. As the correlation is very high in each tissue, we can say the effects of QTL are not different in the two sexes for each of the six tissues.

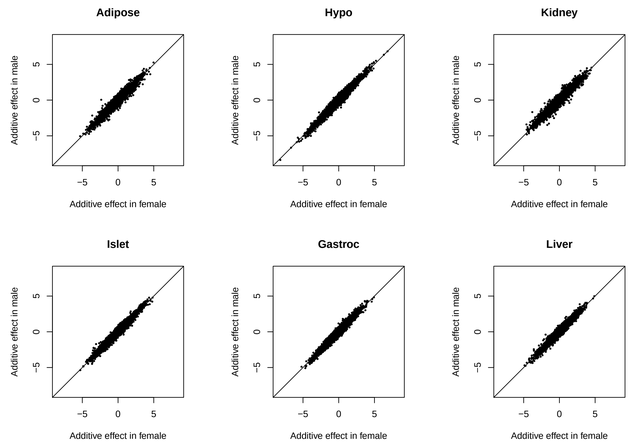

Figure 3: Scatterplots of the estimated additive effects in females vs males for each of the six tissues

	adipose	islet	liver	kidney	hypo	gastroc
correlation	0.96	0.98	0.96	0.96	0.97	0.96

Table 2: The correlation between estimated additive effects in females and males for each tissue

References

[1] Karl W Broman and Saunak Sen. *A Guide to QTL Mapping with R/qtl*, volume 46. Springer, 2009.

[2] Chris S Haley and Sarah A Knott. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69(4):315–324, 1992.

[3] Jianan Tian, Mark P Keller, Angie T Oler, Mary E Rabaglia, Kathryn L Schueler, Donald S Stapleton, Aimee Teo Broman, Wen Zhao, Christina Kendziorski, Brian S Yandell, et al. Identification of the bile acid transporter slco1a6 as a candidate gene that broadly affects gene expression in mouse pancreatic islets. *Genetics*, 201(3):1253–1262, 2015.