QTL mapping in mice: Review of single QTL methods

Karl W Broman

Dept of Biostatistics
Johns Hopkins University

http://biosun01.biostat.jhsph.edu/~kbroman
Backcross experiment
Data and Goals

Phenotypes

\[ y_i = \text{phenotype for mouse } i \]

Marker genotypes

\[ x_{ik} = 1/0 \text{ if } i \text{ is AB/AA at marker } k \]

Genetic map

Locations of markers

Goals:

- Identify the genomic regions (QTLs) contributing to variation in the phenotype
- Identify at least one QTL
- Form confidence interval for QTL location
- Estimate QTL effects
The simplest method: ANOVA

- Split mice into groups according to their genotypes at a marker
- Do a t-test / ANOVA
- Repeat for each typed marker
ANOVA at marker loci

Advantages

• Simple
• Easily incorporate covariates
• Easily extended to more complex models

Disadvantages

• Must exclude individuals with missing genotype data
• Imperfect information about QTL location
• Suffers in low density scans
• Only considers one QTL at a time
Interval mapping

Lander and Botstein 1989

• Consider any one position in the genome as the location for a putative QTL
• For a particular mouse, let \( z = 1/0 \) if (unobserved) genotype at QTL is AB/AA
• Calculate \( \Pr(z = 1 \mid \text{marker data}) \)
  – Assume no meiotic interference
  – Need only consider flanking typed markers
  – May allow for the presence of genotyping errors
• Given genotype at the QTL, phenotype is distributed as normal\((\mu + \Delta z, \sigma^2)\)
• Given marker data, phenotype follows a mixture of normal distributions
Interval mapping

Estimation and LOD scores

• Use a version of the EM algorithm to obtain estimates of $\mu_{AA}$, $\mu_{AB}$, and $\sigma$ (an iterative algorithm)

• Calculate the LOD score

$$\text{LOD} = \log_{10} \left\{ \frac{\Pr(\text{data} \mid \hat{\mu}_{AA}, \hat{\mu}_{AB}, \hat{\sigma})}{\Pr(\text{data} \mid \text{no QTL})} \right\}$$

• Repeat for all other genomic positions (in practice, at 0.5 cM steps along genome)
A simulated example

A

![Graph A](image)

B

![Graph B](image)
Interval mapping

Advantages

- Make proper account of missing data
- Can allow for the presence of genotyping errors
- Pretty pictures
- Higher power in low-density scans
- Improved estimate of QTL location

Disadvantages

- Greater computational effort
- Requires specialized software
- More difficult to include covariates
- Only considers one QTL at a time
Haley-Knott regression

- An approximation of IM
- Regress phenotypes on \( \Pr(z = 1 \mid \text{marker data}) \)
  - Like assuming Normal(\( \mu + \Delta \rho, \sigma \)) rather than a mixture of Normal(\( \mu, \sigma \)) and Normal(\( \mu + \Delta, \sigma \))
  - Like a half-step of the EM algorithm

Advantages

- Fast
- Nearly IM if not too much missing data
- Can easily incorporate covariates

Disadvantages

- Only nearly IM
Statistical significance

Large LOD score \implies evidence for QTL

Question: How large is large?

Answer 1: Consider distribution of LOD score if there were no QTL

Answer 2: Consider distribution of maximum LOD score
More on LOD thresholds

Appropriate threshold depends on:

• Size of genome
• Number of typed markers
• Pattern of missing data
• Stringency of significance threshold
• Type of cross (e.g. $F_2$ vs BC)
• Etc.

Methods for obtaining thresholds

• Analytical calculations (assuming dense map of markers)
• Computer simulations
• Permutation/randomization test
LOD support intervals

1.5-LOD support interval =

• Interval in which LOD score is within 1.5 of its maximum
• Indicates plausible location for QTL

Plot LOD score with maximum at 0

→ Valuable tool for depicting evidence for QTL location
Selection bias in QTL effects

QTL effect = 5
Bias = 79%

QTL effect = 8
Bias = 18%

QTL effect = 11
Bias = 1%
Summary

• **Simplest method: ANOVA**
  – Suffers in the presence of missing genotype data and/or low-density map

• **Interval mapping**
  – Makes proper account of missing data
  – More computationally intensive
  – Still only considers single QTLs

• **Statistical significance**
  – Adjust for multiple testing
  – Permutation tests

• **QTL location: LOD support intervals**
  – Plot LOD curve re-centered so max=0

• **Selection bias**
  – Est’d QTL effects may be too large
QTL effect = 5
Power = 2%

QTL effect = 8
Power = 41%

QTL effect = 11
Power = 97%
A 100 mice

B 200 mice

Chromosome position (cM)

LOD – max(LOD)