Hopkins Statistical Genetics Working Group

- **Goals:**
  - Get to know each other
  - Foster collaboration
  - See new stuff
  - Get help
- **Scope:**

  statistical, computational, mathematical issues in genetics, genomics, molecular biology
- **Format:**
  - interesting papers, problems, works-in-progress
  - informal and interactive
- **Time:**
  - Mondays, 3:30pm
  - biweekly (next meeting 16 Oct)
- **Webpage:**

  biosun01.biostat.jhsph.edu/~kbroman/hsgwg

Identifying QTLs in experimental crosses

Karl W Broman

Humans vs model organisms

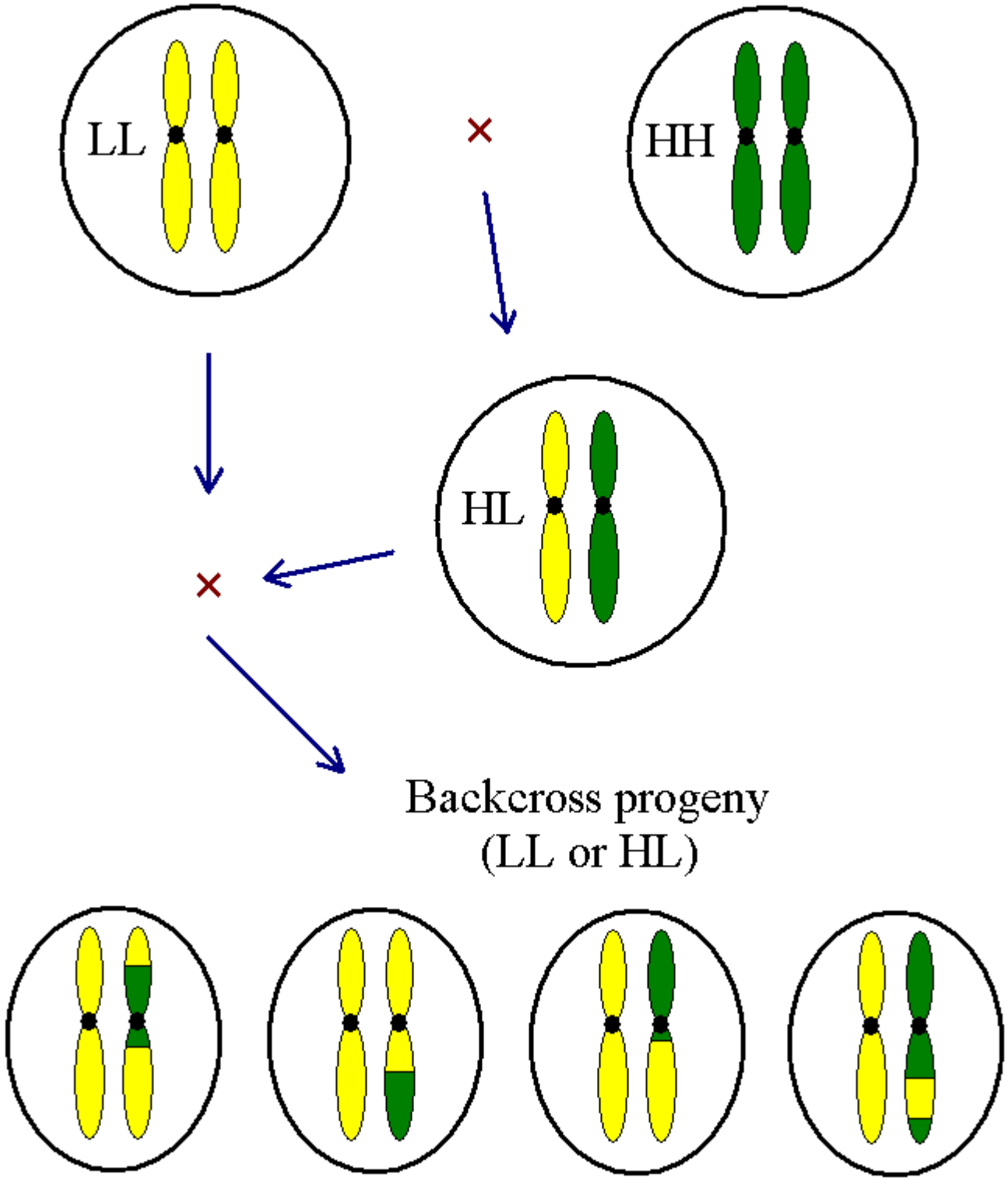
"Model":

- Genes
- Genetic architecture
- Analysis methods

Differences:

- Complexity of pedigrees
- Complexity of genet. architec.
- Environmental variation

Backcross experiment



Data

Phenotypes (trait values)

y_i = phenotype for individual *i*

Marker genotypes

x_{ij} = 1/0 if *i* is HL/LL at marker *j*

Genetic map

Locations of markers

Models

Recombination: No interference

Phenotype/genotype connection

$$y = \mu + \sum \beta_j z_j + \varepsilon$$

$$\varepsilon \sim \text{Normal}(0, \sigma^2)$$

Problem

100 to 1000 backcross progeny

100 to 400 markers

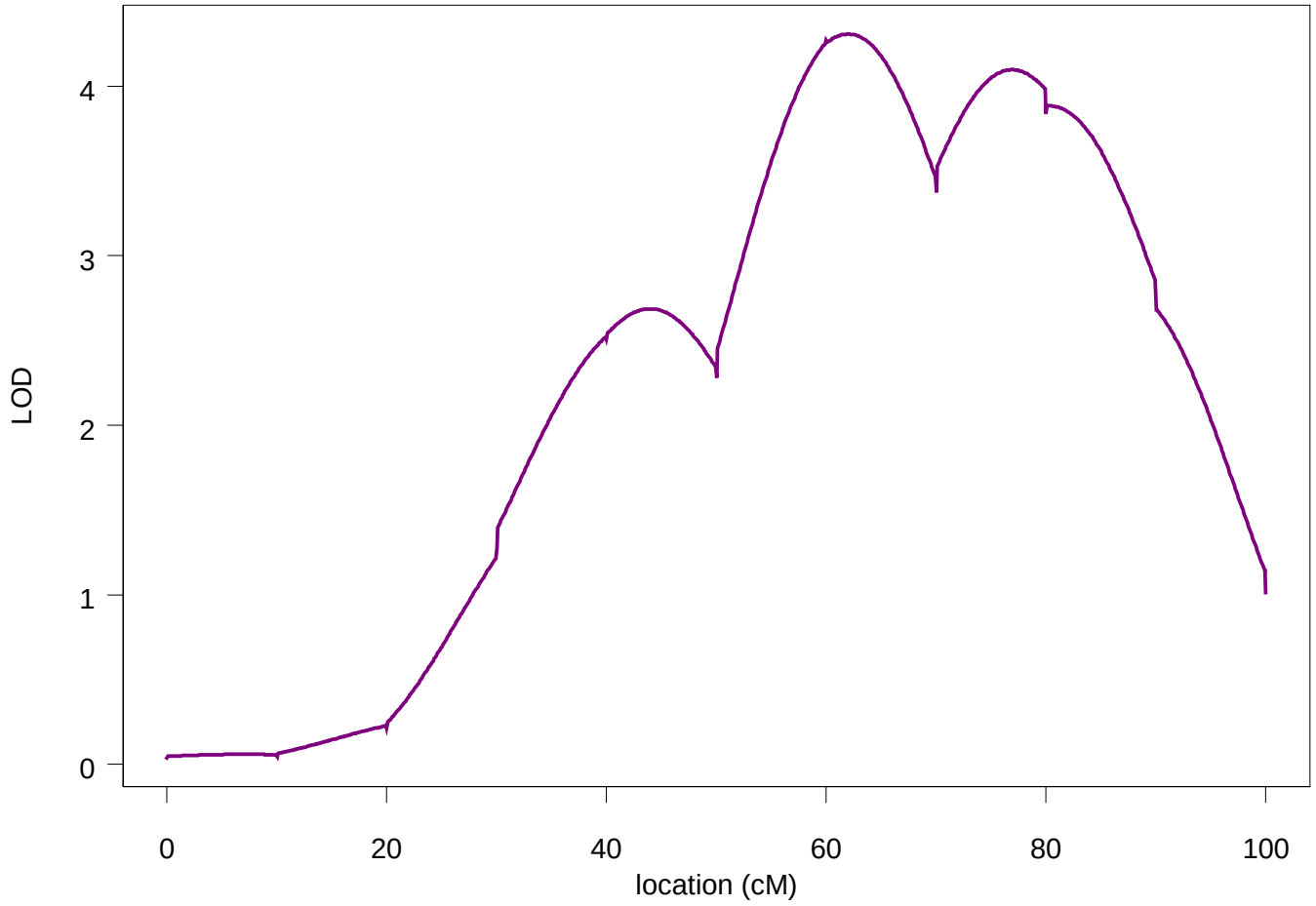
$y = \mu + \sum \beta_j x_j + \varepsilon$

Find the x's with $\beta_j \neq 0$

Errors:

- Miss important loci
- Include extraneous loci

The usual method



At each location:

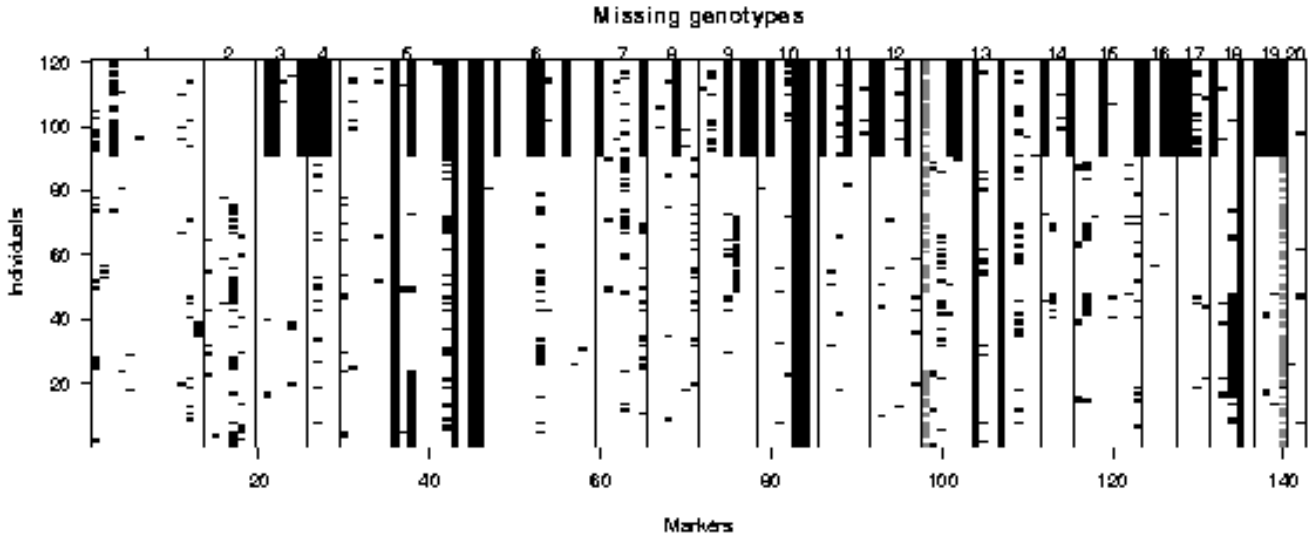
- Imagine a single QTL
- Infer genotypes
- Regression of phenotype on genotype

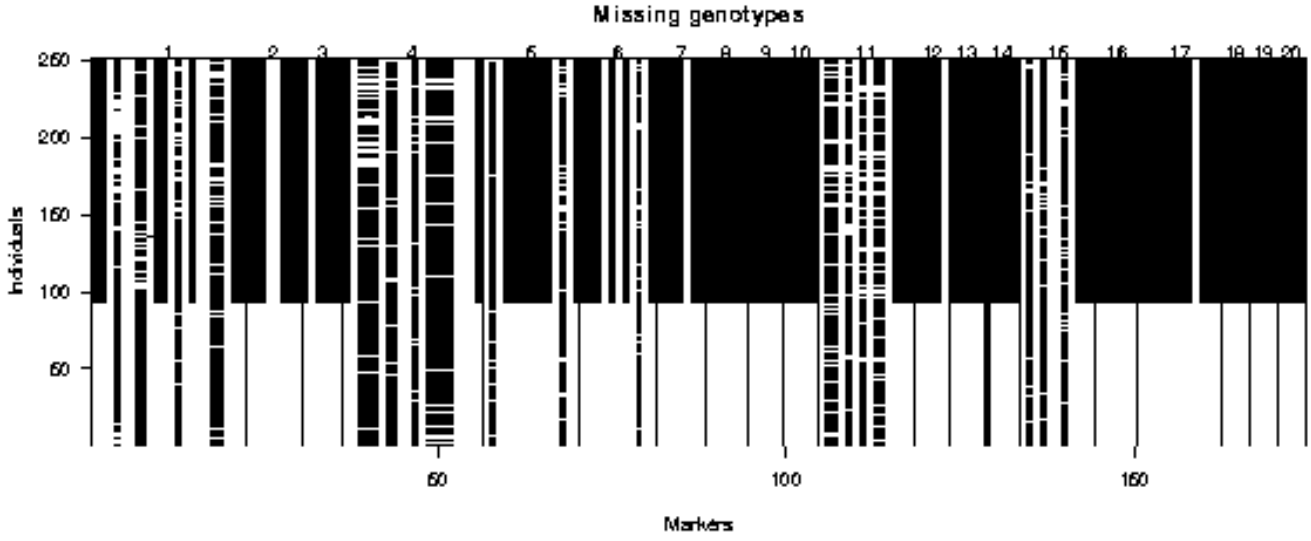
Major issues

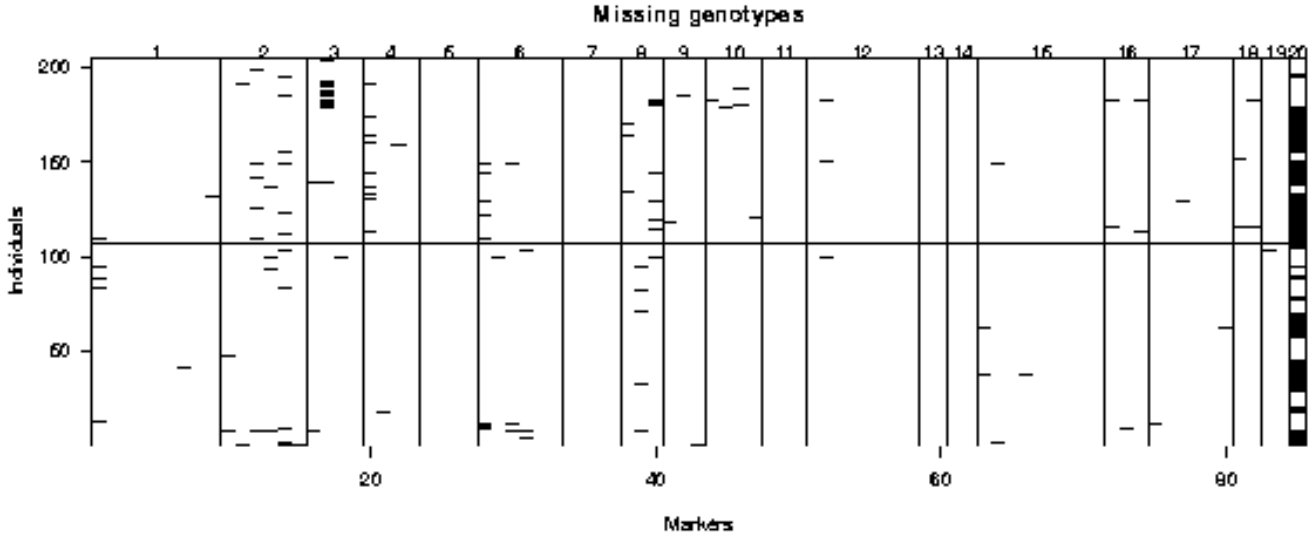
- Missing genotype information
- Model selection
- Estimation of QTL effects
- Estimation of QTL location

Model Selection:

- Space of models
- Searching through models
- Comparing models
- Assessing performance


Missing genotypes
Individuals
Markers


Missing genotypes
Individuals
Markers


Missing genotypes
Individuals
Markers

Śaunak Sen and Gary Churchill

A statistical framework for quantitative trait mapping

http://www.jax.org/research/churchill

---

y = phenotypes

m = marker genotypes

H = genetic model

g = QTL genotypes

μ = parameters of genetic model

θ = locations of QTLs

$$p_H(y, m, g, \mu, \theta) = \{ p_H(y \mid g, \mu)\, p_H(\mu) \} \{ p(g \mid m, \theta)\, p(m)\, p(\theta) \}$$

Multiple imputation

Grid G of "pseudomarker" positions

(e.g., equally spaced at 2 cM over entire genome)

$r_i(u)$ = realization of matrix of genotypes simulated from $p(g \mid m, \theta = G)$

- u = locations on grid
- i = 1…q realizations

Weights:

$$W_H[r_i(u)] = p_H[y \mid g = r_i(u)]\, p(\theta = u)$$

- integrate over μ
- $p(\theta = u)$ may be ignored

e.g., Normal model (v = # QTLs):

$$-2 \log W_H[r_i(u)] = v \log n + n \log RSS$$

Estimating QTL locations

$p_H(\theta = u \mid y, m) \quad \alpha \quad \Sigma_i W_H[r_i(u)]$

Estimating QTL model parameters

$p_H(\mu \mid y, m) \quad \alpha$

$\Sigma_i \Sigma_u \, p_H[\mu \mid y, g=r_i(u)] \, W_H[r_i(u)]$

Model selection

Bayes factor $B(H,K) = p_H(y,m) / p_K(y,m)$

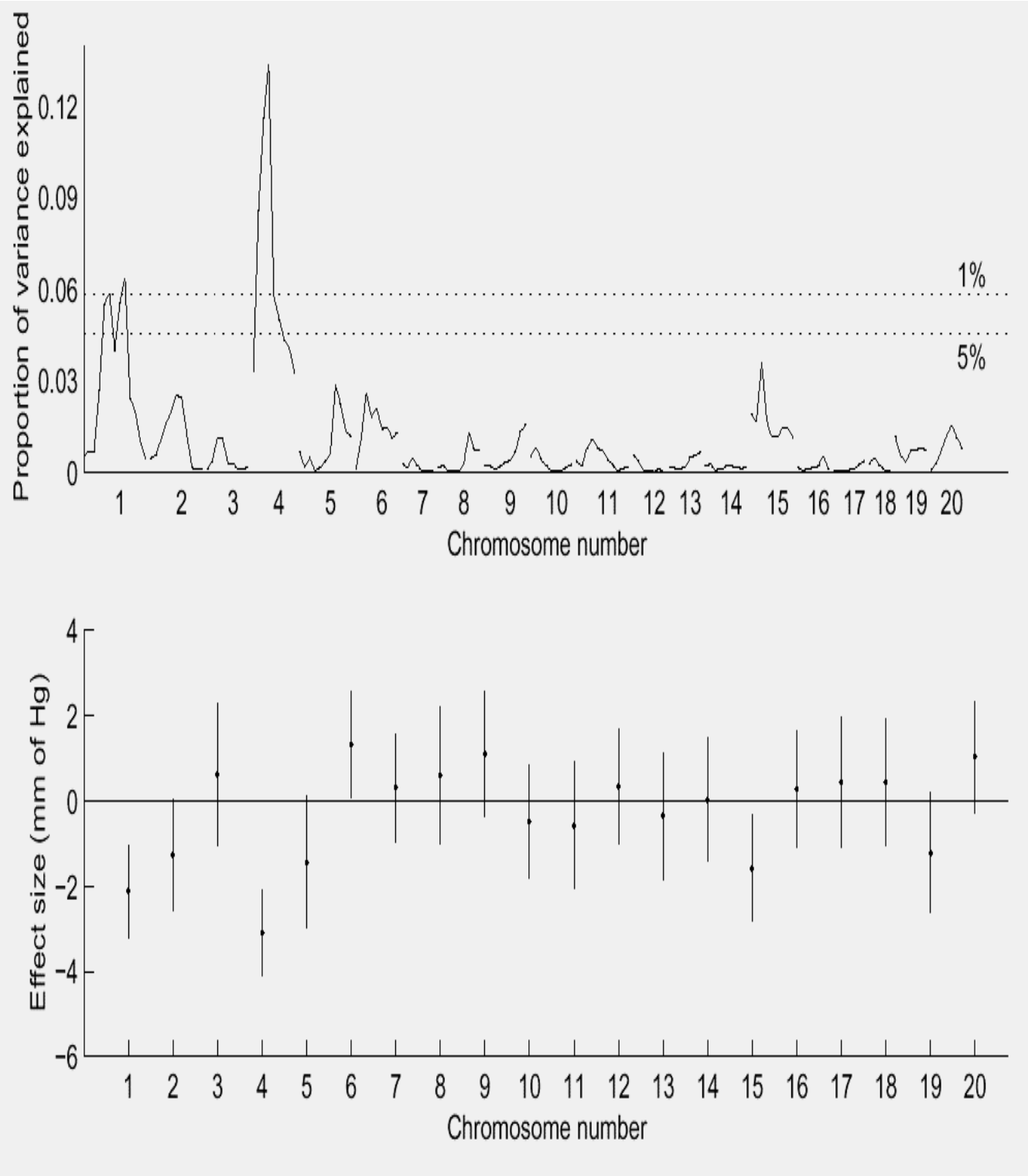
$p_H(y \mid m) \approx \Sigma_i \Sigma_u \, W_H[r_i(u)] / (q \, s)$

q = number of realizations
s = number of p-tuples of qtl loc

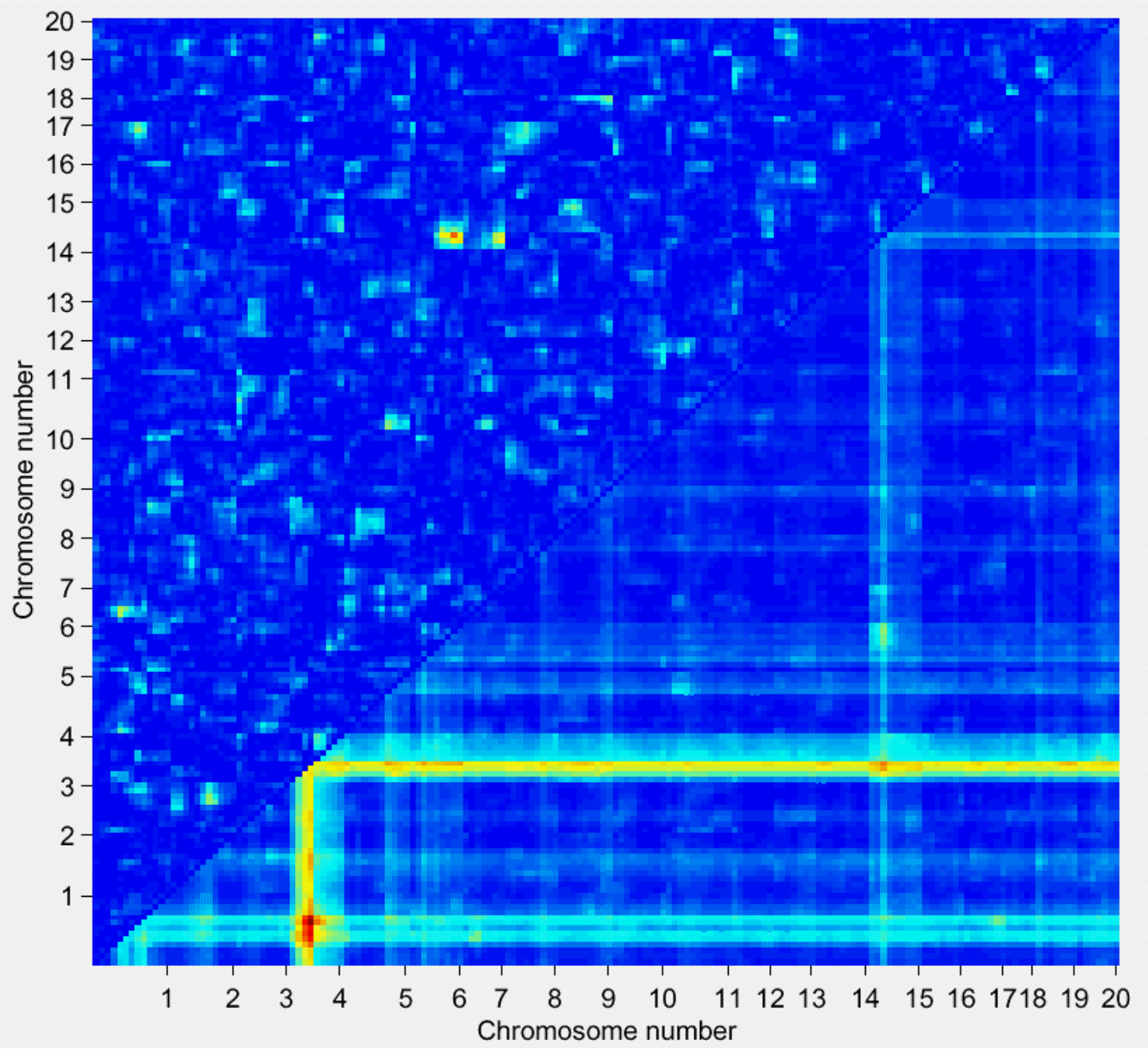
General approach:

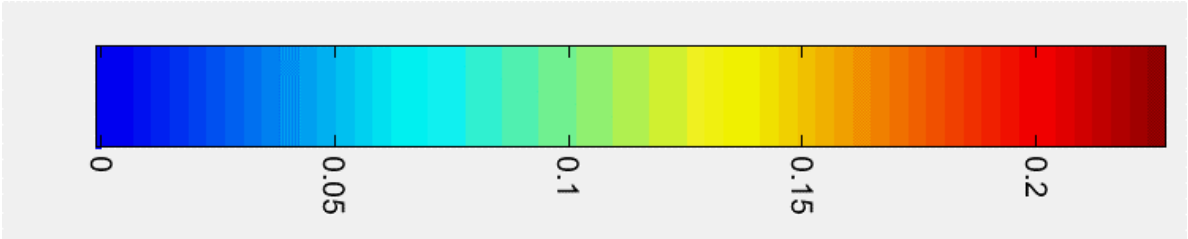
Main-scan
Pair-scan for interactions

Main scan


Proportion of variance explained
0.12
0.09
0.06
0.03
0
1%
5%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome number
Effect size (mm of Hg)
4
2
0
-2
-4
-6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome number

Pair scan


Chromosome number
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome number


0
0.05
0.1
0.15
0.2

Pair scan: closeup view