QTL mapping 2: Special topics

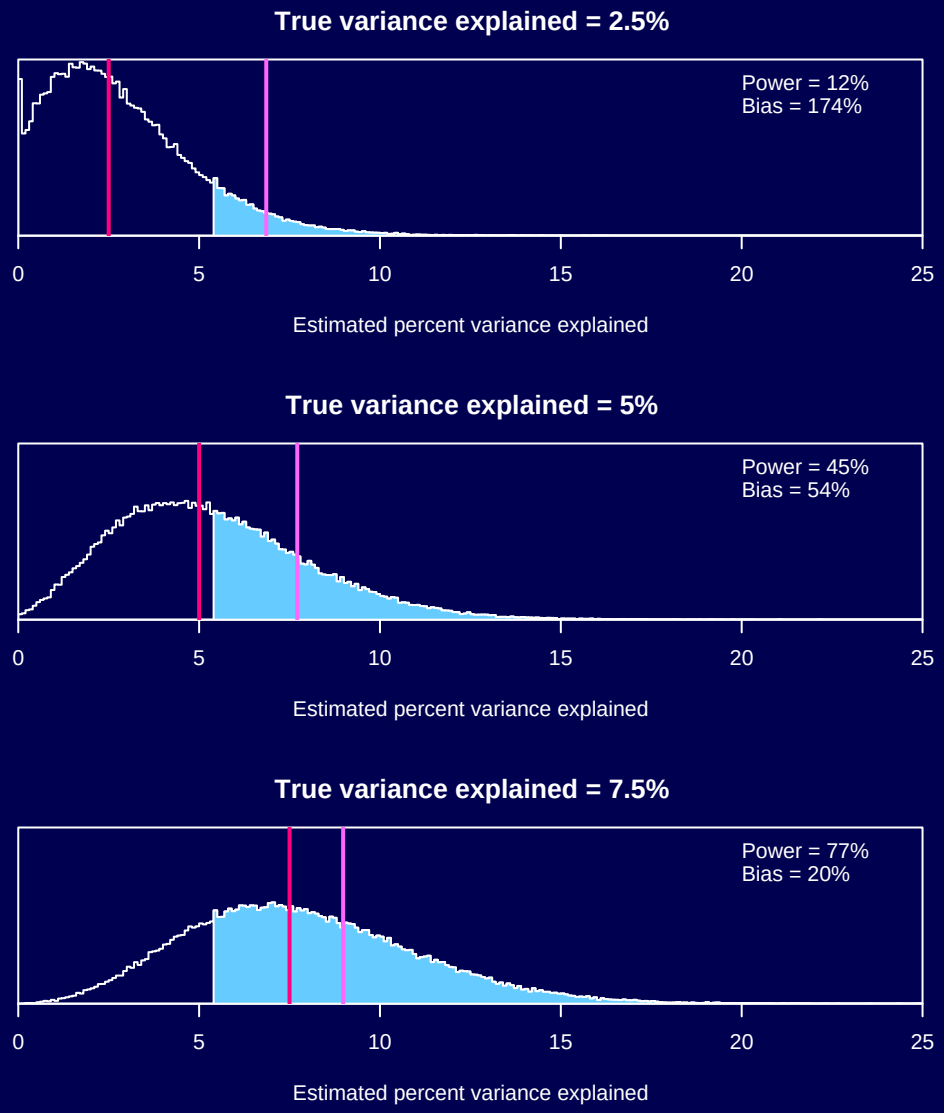
Karl Broman

Biostatistics and Medical Informatics
University of Wisconsin – Madison

kbroman.org
github.com/kbroman
@kwbroman

Selection bias

- The estimated effect of a QTL will vary somewhat from its true effect.
- Only when the estimated effect is large will the QTL be detected.
- Among those experiments in which the QTL is detected, the estimated QTL effect will be, on average, larger than its true effect.
- This is selection bias.
- Selection bias is largest in QTLs with small or moderate effects.
- The true effects of QTLs that we identify are likely smaller than was observed.



Implications

- Estimated % variance explained by identified QTLs
- Repeating an experiment
- Congenics (aka near isogenic lines)
- Marker-assisted selection

Non-normal traits

- Standard interval mapping assumes normally distributed residual variation. (Thus the phenotype distribution is a mixture of normals.)
- In reality: we see dichotomous traits, counts, skewed distributions, outliers, and all sorts of odd things.
- Interval mapping, with LOD thresholds derived from permutation tests, generally performs just fine anyway.
- Alternatives to consider:
  - Nonparametric approaches (Kruglyak & Lander 1995)
  - Transformations (*e.g.*, log, square root, normal quantiles)
  - Specially-tailored models (*e.g.*, a generalized linear model, the Cox proportional hazard model, and the two-part model in Broman 2003)

Haley-Knott regression

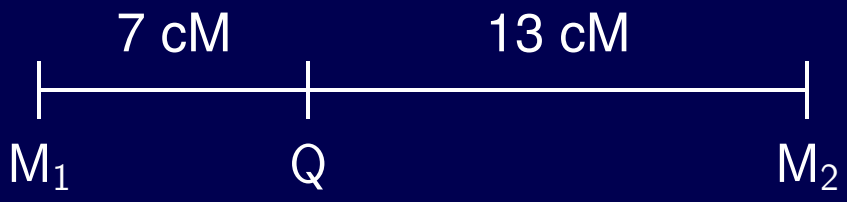
A quick approximation to Interval Mapping.

$$
\begin{aligned}
E(y_i|q_i) &= \mu_q \\
E(y_i|M_i) &= E[\, E(y_i|q_i) \,|M_i] = \textstyle\sum_j \Pr(q = j|M_i)\mu_j \\
&= \textstyle\sum_j p_{ij}\mu_j
\end{aligned}
$$

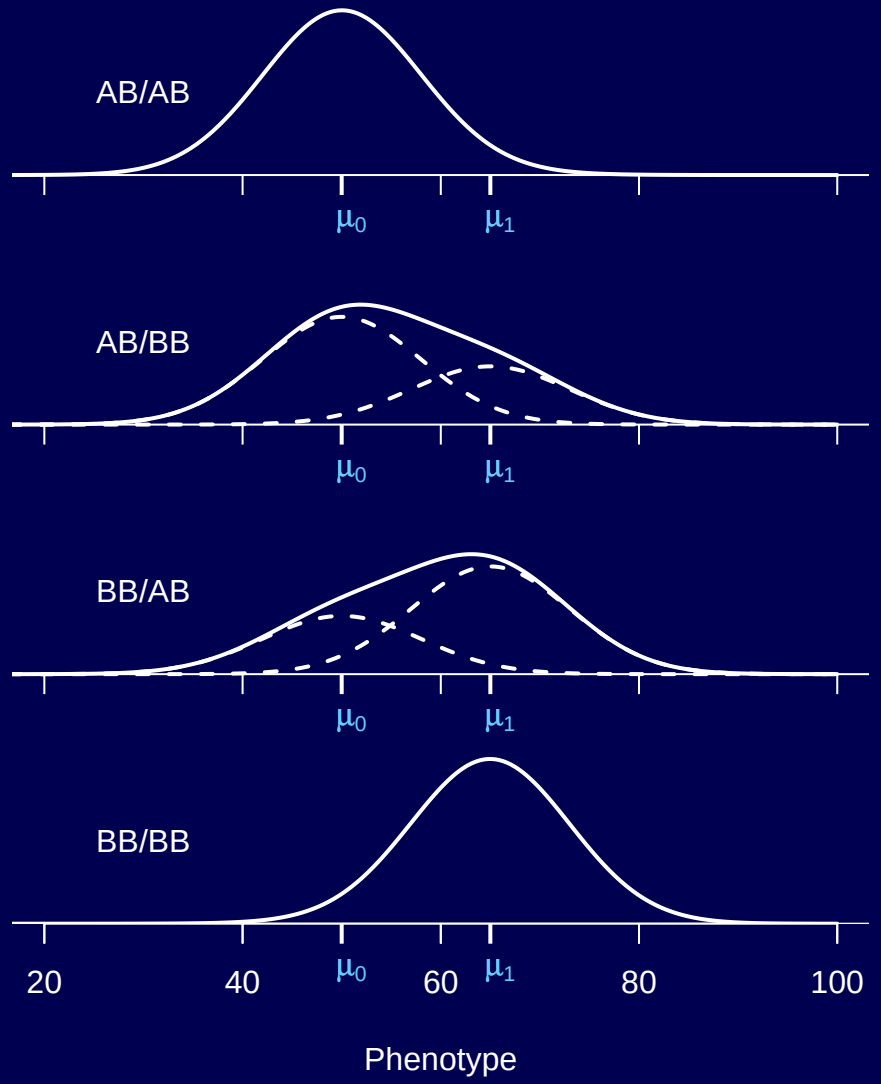
Regress y on p_i, pretending the residual variation is normally distributed (with constant variance).

$$
\text{LOD} = \frac{n}{2}\log_{10}\left(\frac{\text{RSS}_0}{\text{RSS}_1}\right)
$$

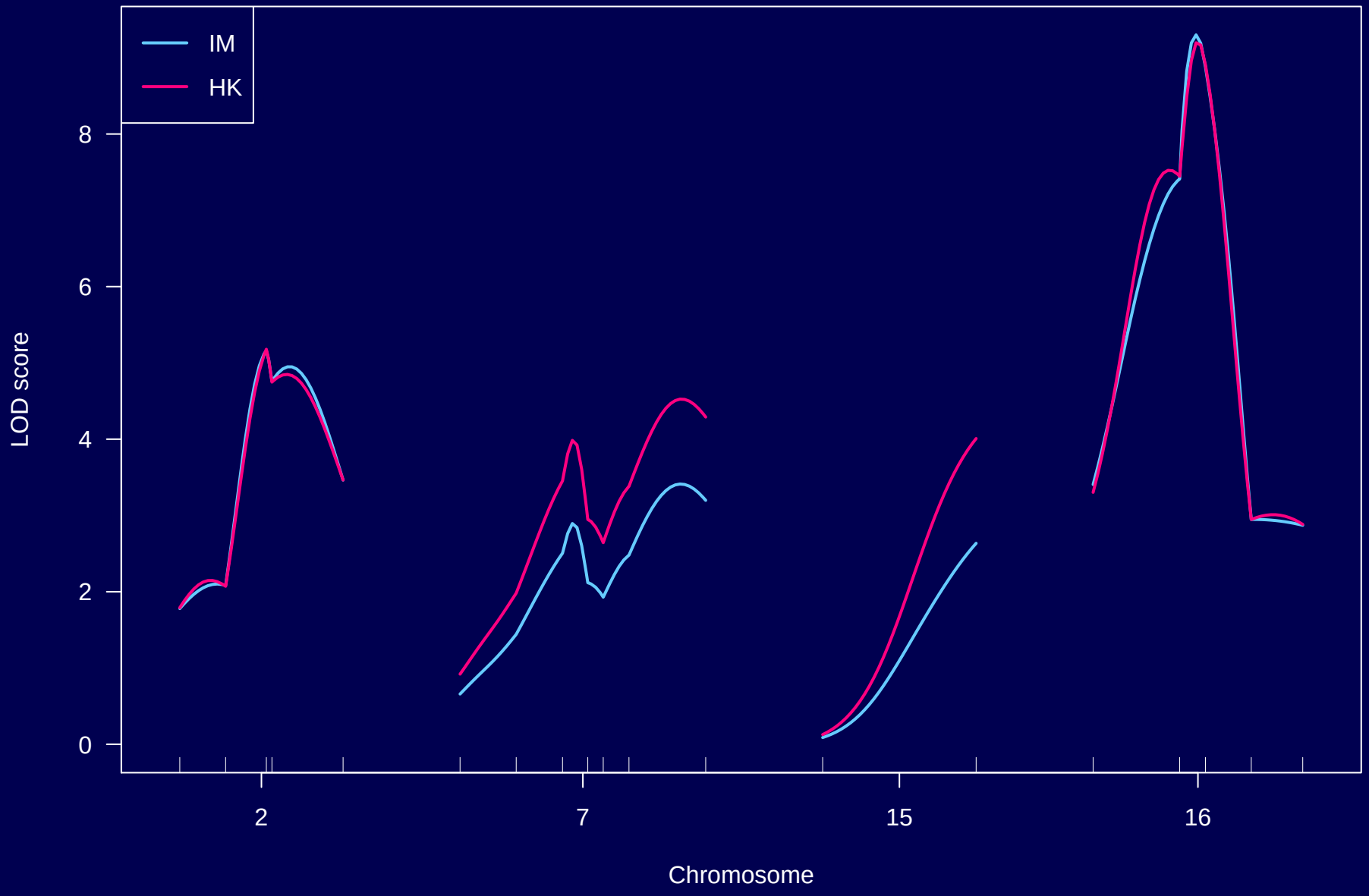
The normal mixtures



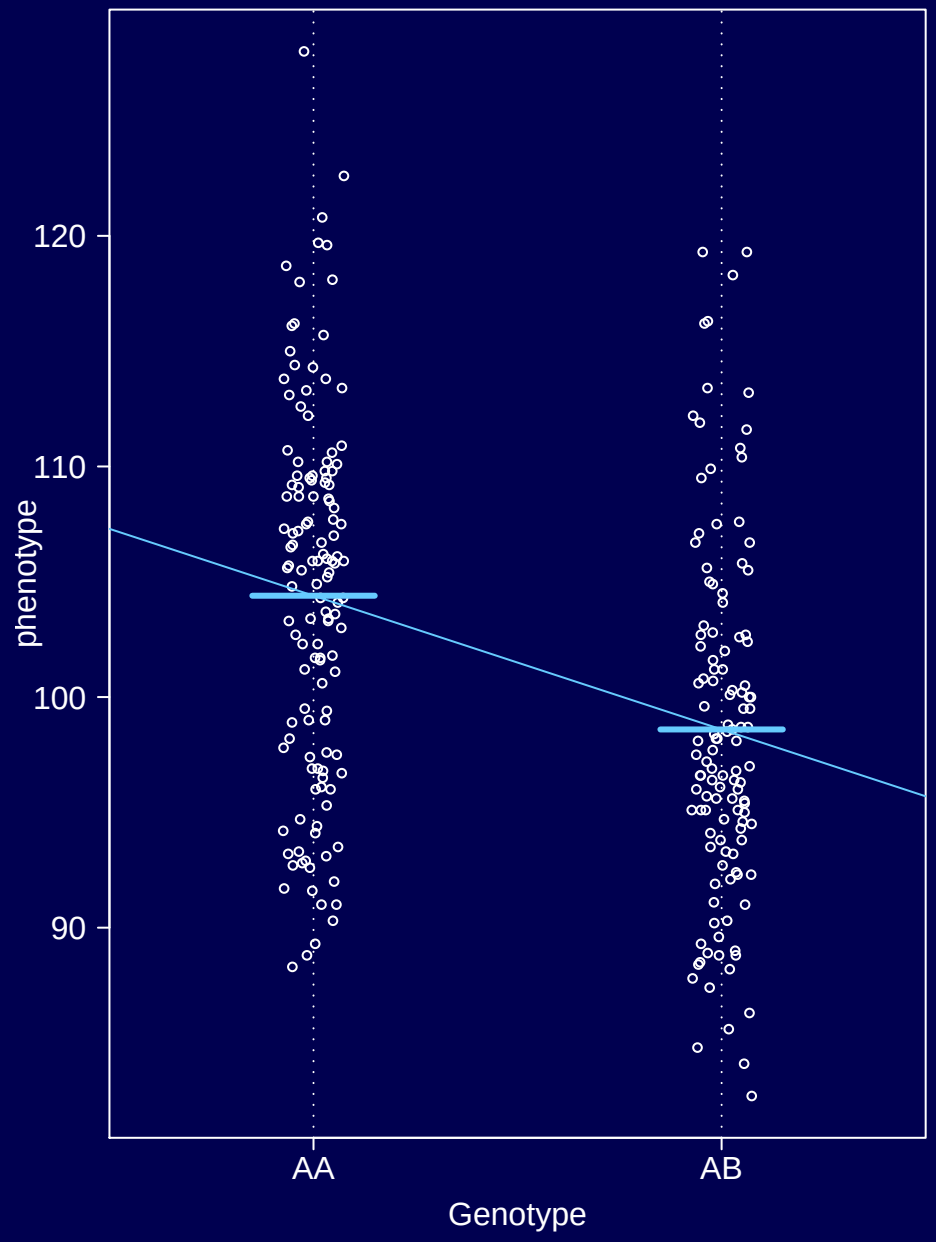
- Two markers separated by 20 cM, with the QTL closer to the left marker.
- The figure at right shows the distributions of the phenotype conditional on the genotypes at the two markers.
- The dashed curves correspond to the components of the mixtures.

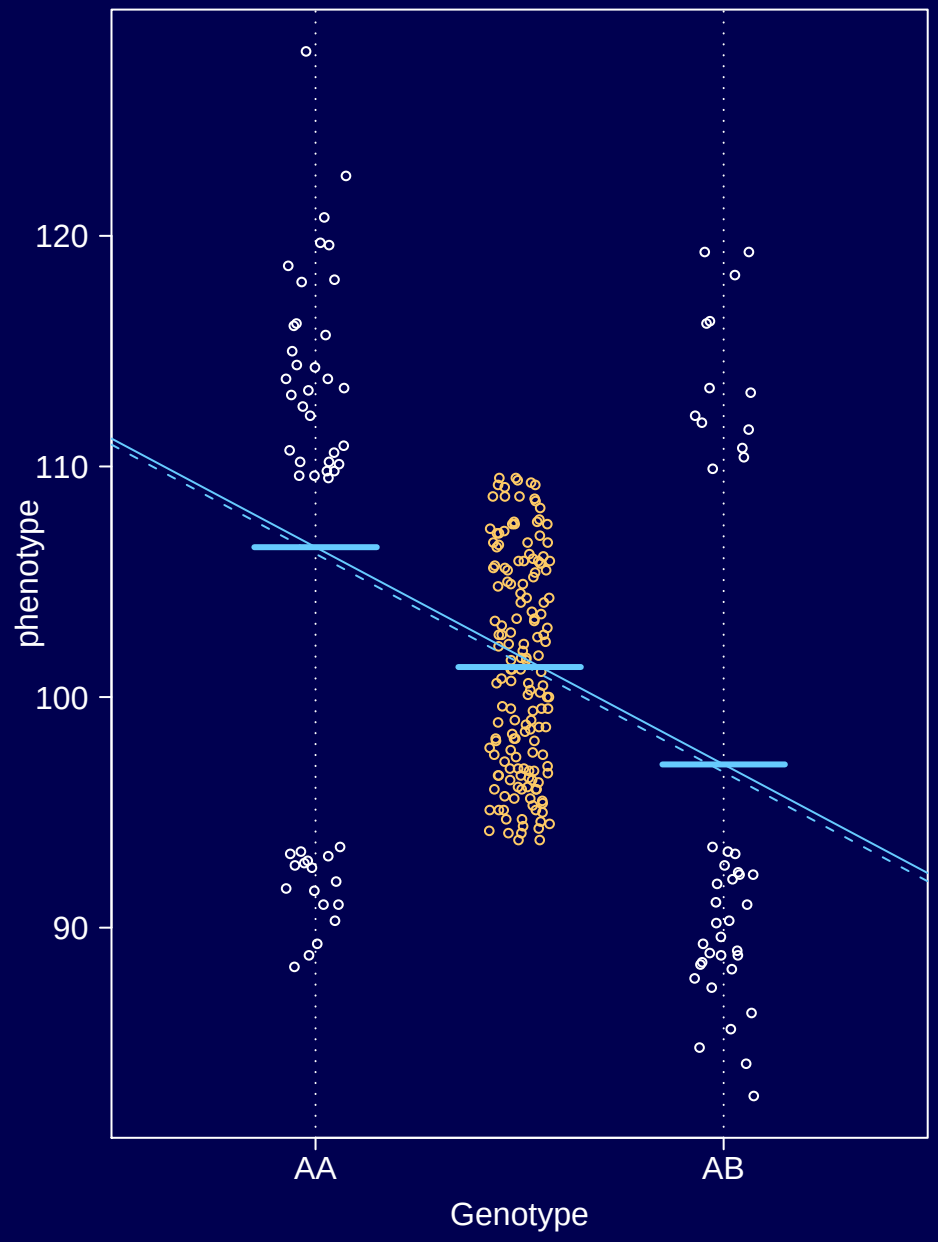


Haley-Knott results

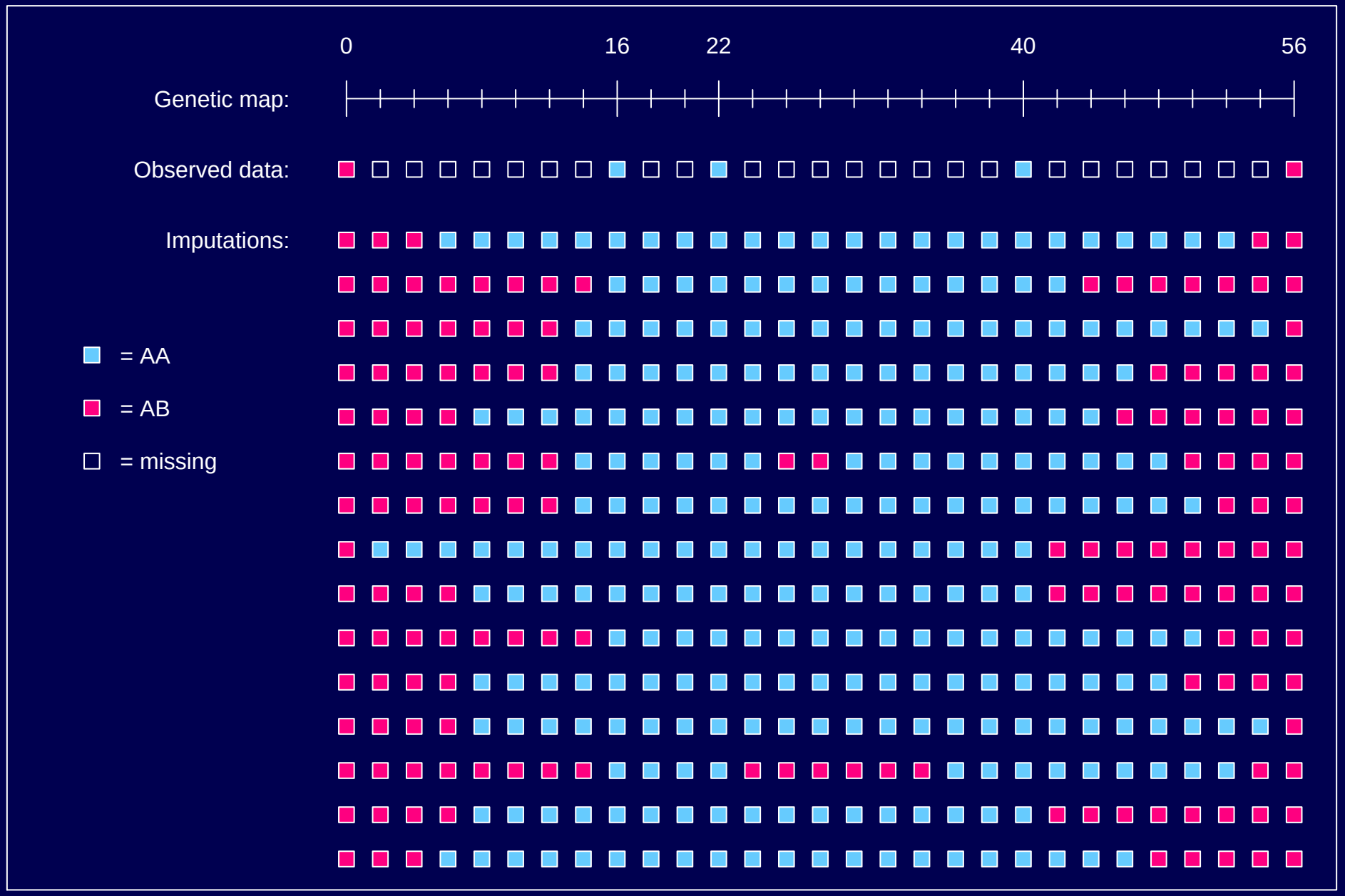


H-K with selective genotyping

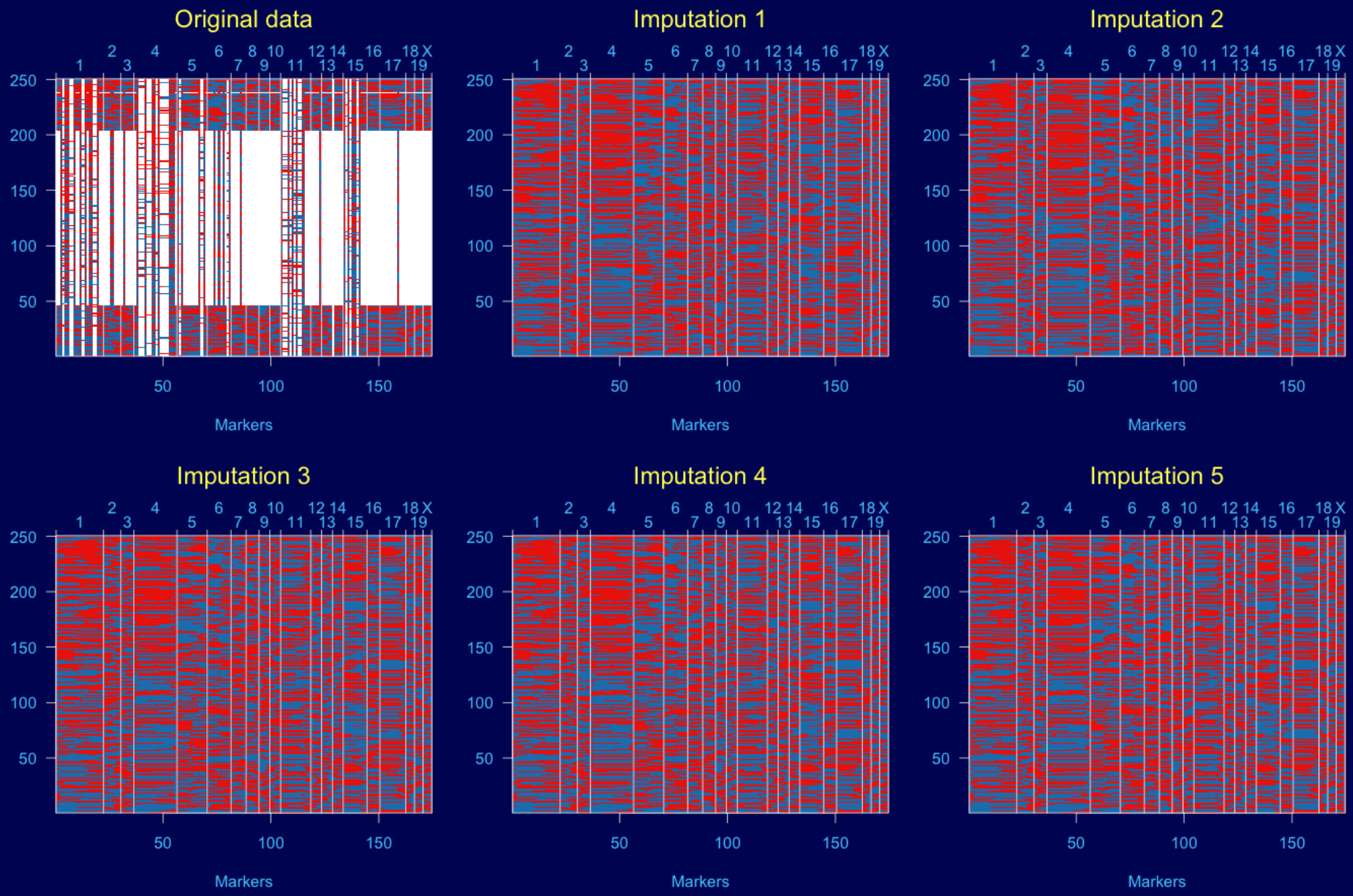




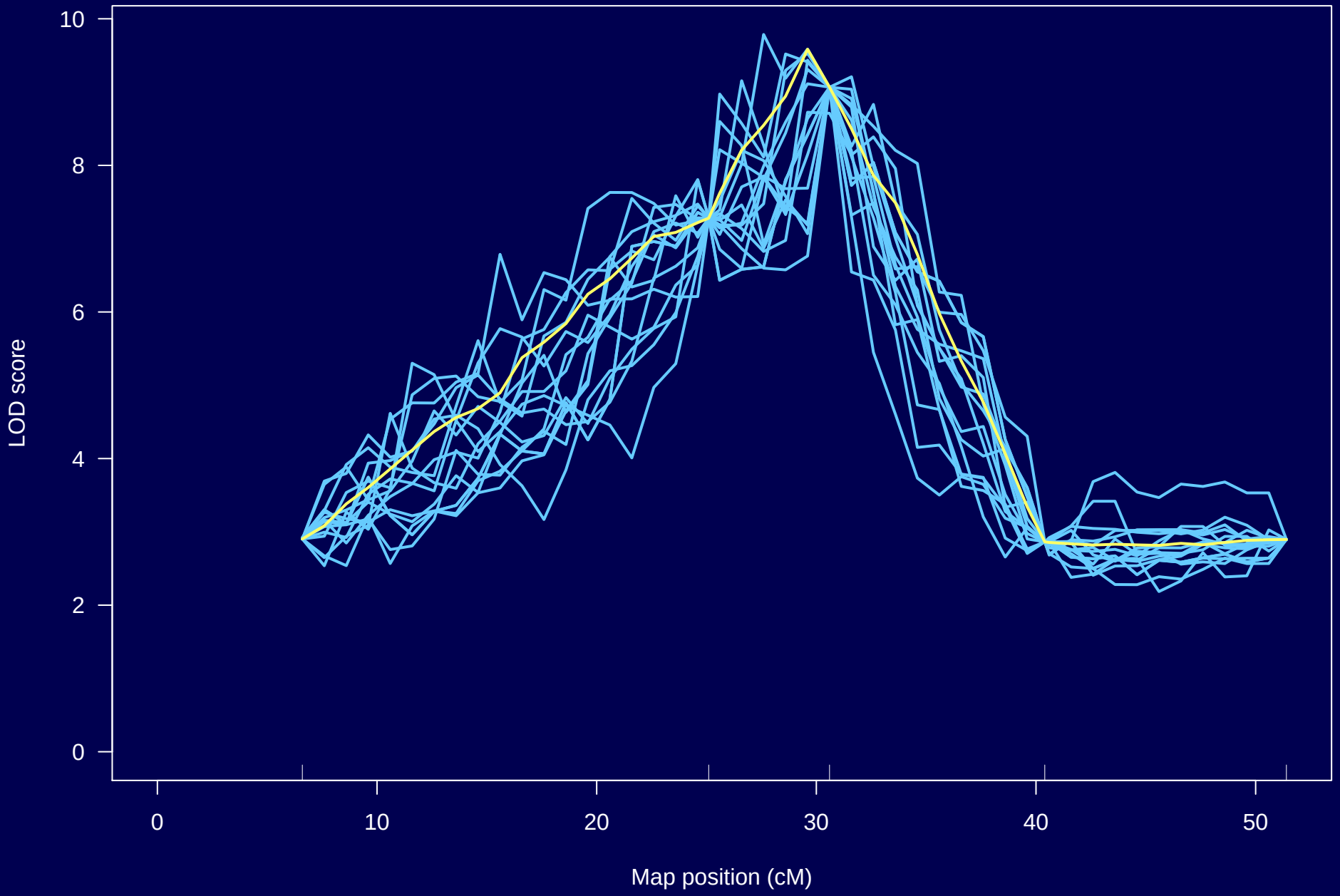
Multiple imputation



Multiple imputations



Imputation LOD curves



Summary comparison

Approach	Speed	Extensibility	Stability	Missing data	Parallelization
HK	++	+	+	−	++
EM	+	−	−	+	−
Imputation	−	+	+	+	+

Covariates

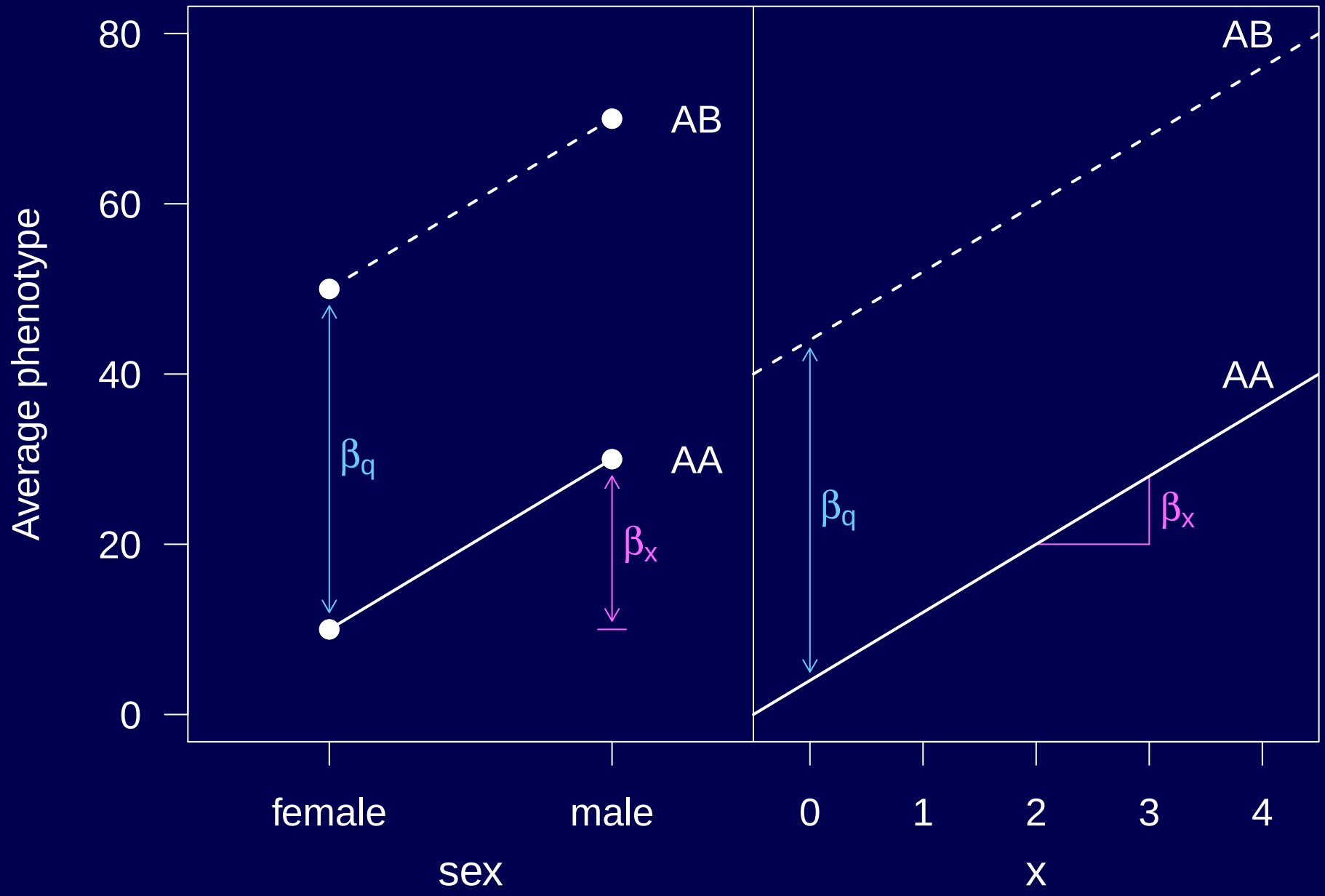
- Examples: treatment, sex, age, weight
- Control residual variation $\rightarrow$ increase power
- Look for QTL $\times$ covariate interactions

Additive covariate

$$\mathsf{H}_0 : y = \mu + \beta_x x + \epsilon$$
$$\mathsf{H}_a : y = \mu + \beta_x x + \beta_q q + \epsilon$$

- If covariate has strong effect on the phenotype, accounting for it can give improved power to detect QTL.
- In permutations, keep phenotype and covariate together
- Use care when the covariate is another phenotype

Additive covariate

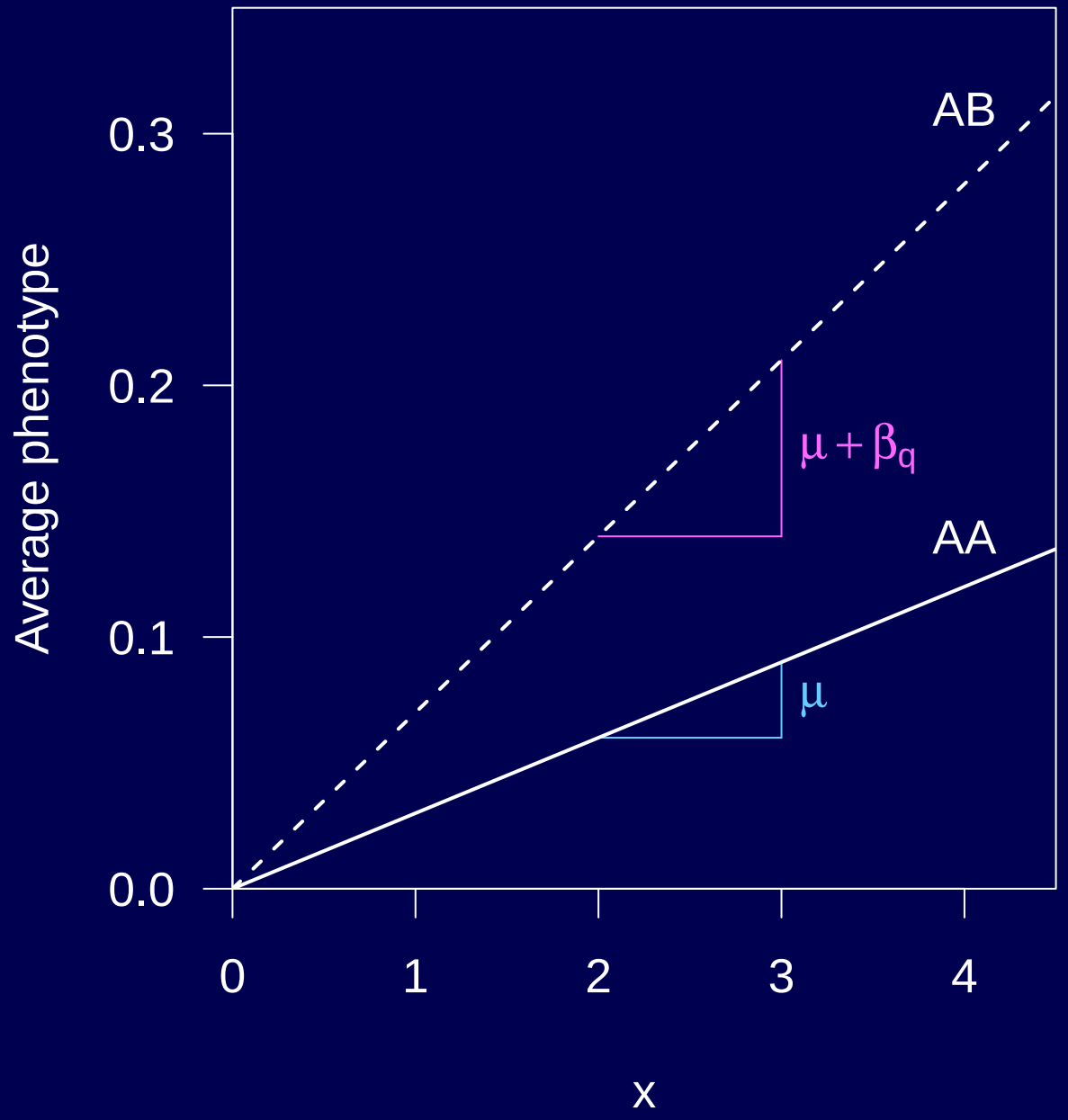


Adjust then scan?

- Consider adjusted phenotype $y' = y/x$
- The QTL model is $(y/x) = \mu + \beta_q q + \epsilon$
- Equivalently

$$y = \begin{cases} \mu\, x + \epsilon' & \text{if } q = 0 \\ (\mu + \beta_q)x + \epsilon' & \text{if } q = 1 \end{cases}$$

Adjust then scan?



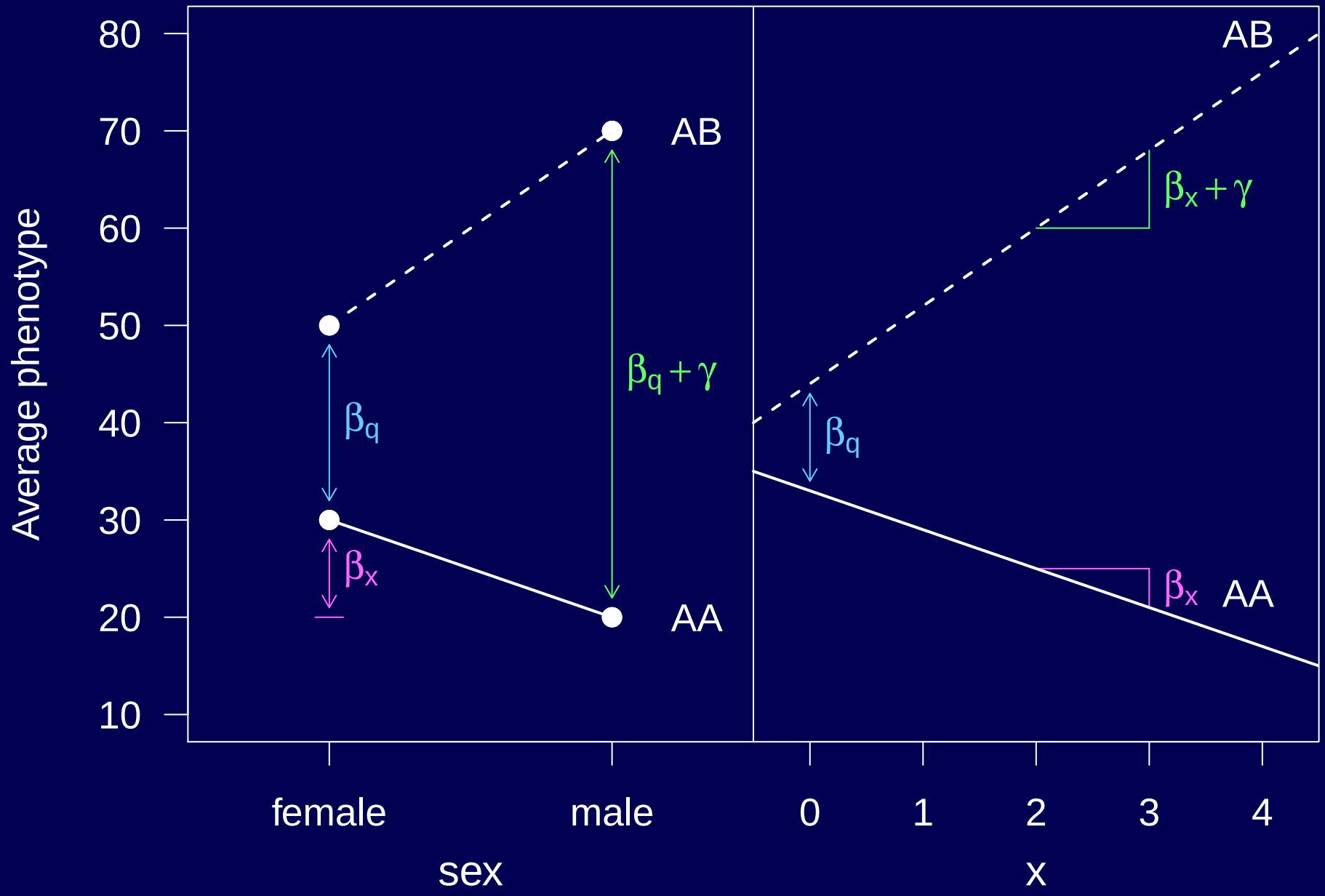
Interactive covariate

$$\begin{aligned} \mathsf{H}_0 : y &= \mu + \beta_x x + \epsilon \\ \mathsf{H}_a : y &= \mu + \beta_x x + \beta_q q + \epsilon \\ \mathsf{H}_i : y &= \mu + \beta_x x + \beta_q q + \gamma x q + \epsilon \end{aligned}$$

Can consider 3 LOD scores:

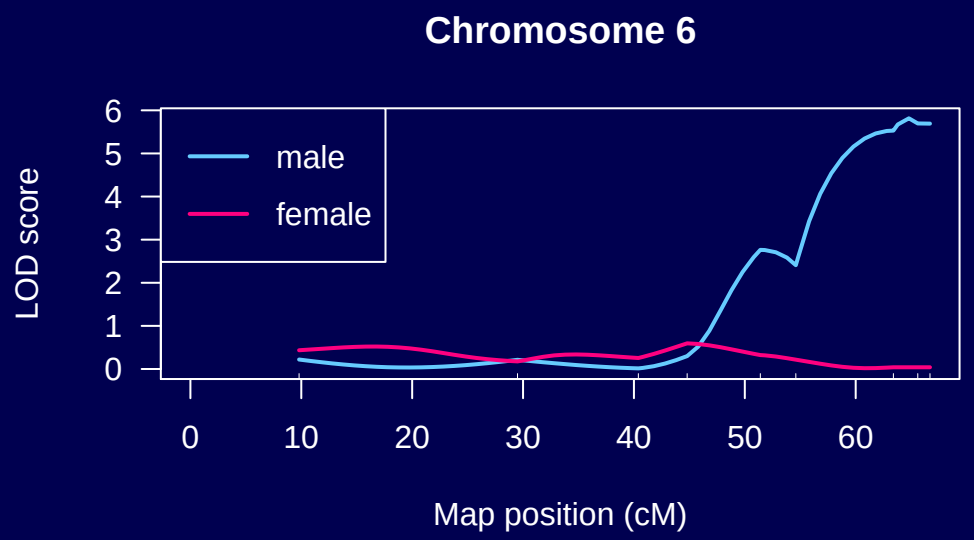
- LOD_a comparing H_a and H_0
- LOD_f comparing H_i and H_0
- LOD_i comparing H_i and H_a

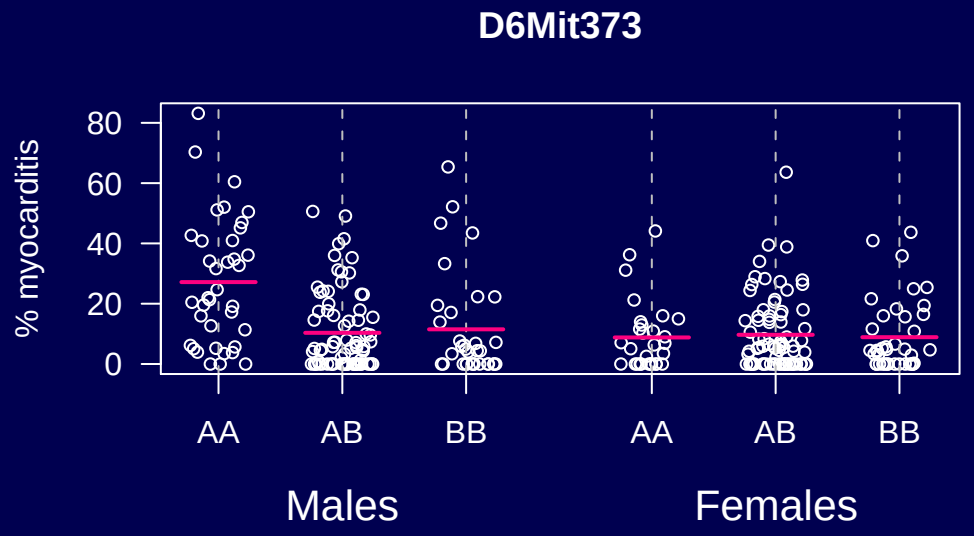
Interactive covariate



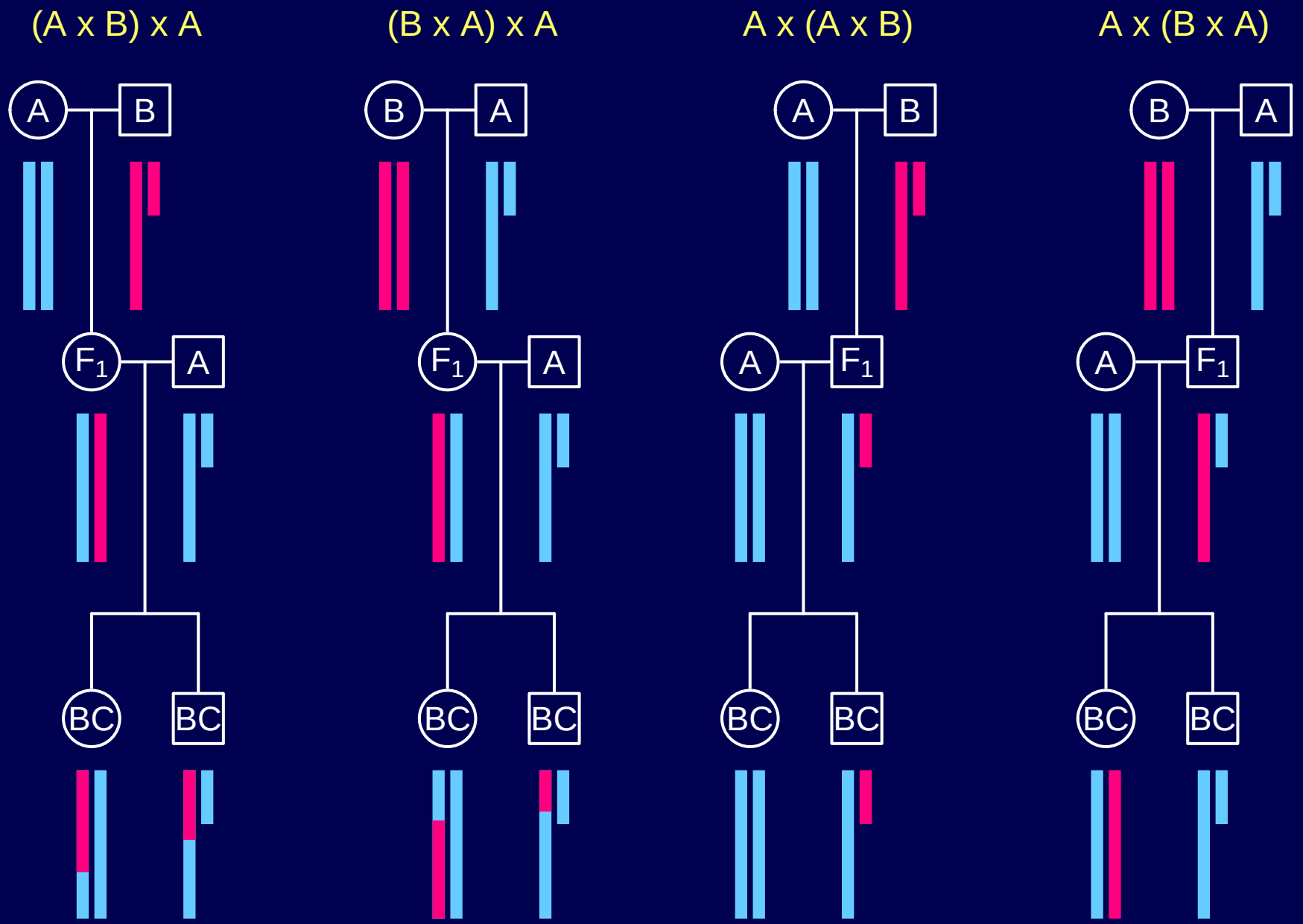
Split on sex?

- Informative, understandable
- But tempting to falsely conclude "sex-specific QTL"
- Absence of evidence is not *evidence of absence*.
- Use explicit test of QTL $\times$ sex interaction

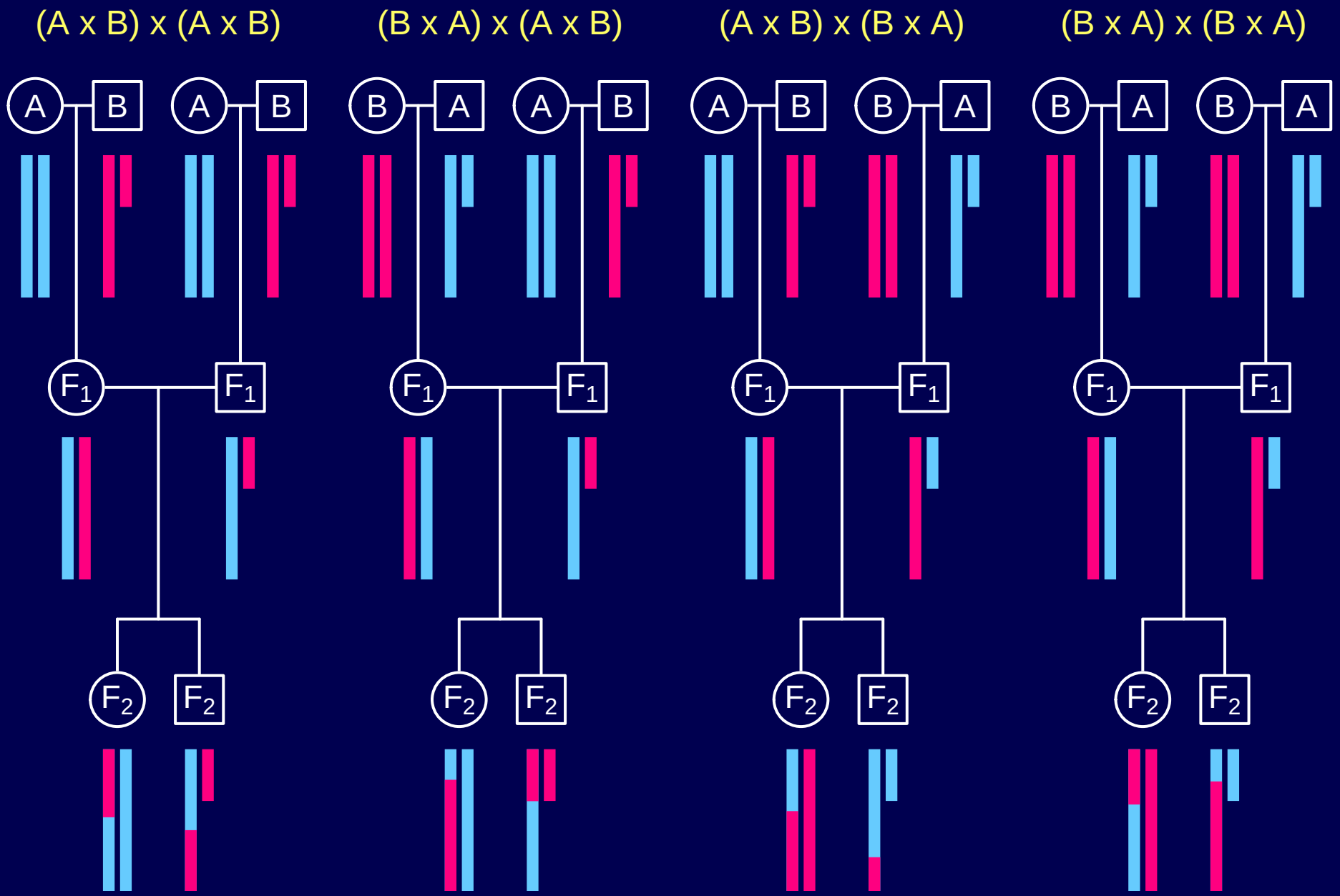




X chr in backcross



X chr in intercross



Example

Intercross: both dir, both sexes

- ♀ forward AA or AB
- ♀ reverse AB or BB
- ♂ forward AY or BY
- ♂ reverse AY or BY

Principles

- Sex- or cross-direction-difference in the phenotype shouldn't lead to spurious linkage on the X chromosome
- Simple as possible
- Null nested within alternative

Approach

Cross	Direction	Sexes	Contrasts	H_0	df
BC		both	AA:AB:AY:BY	♀:♂	2
BC		♀	AA:AB	grand mean	1
BC		♂	AY:BY	grand mean	1
F_2	both	both	AA:ABf:ABr:BB:AY:BY	$♀_f$:$♀_r$:♂	3
F_2	both	♀	AA:ABf:ABr:BB	$♀_f$:$♀_r$	2
F_2	both	♂	AY:BY	grand mean	1
F_2	one	both	AA:AB:AY:BY	♀:♂	2
F_2	one	♀	AA:AB	grand mean	1
F_2	one	♂	AY:BY	grand mean	1

Chromosome-specific thresholds

Let α_i = false positive rate for chromosome i.

We need $1 - \alpha = \prod(1 - \alpha_i)$

For example, $\alpha_1 = \alpha$ and $\alpha_j = 0$ for $j \neq 1$

The usual method: constant LOD threshold
(i.e., constant power)

My approach: $\alpha_i \propto L_i$ where L_i = length of chr i

Similar and more convenient: $\alpha_i = (1 - \alpha)^{L_i/L}$

Data diagnostics

- Plot phenotypes
- Look for sample duplicates
- Look for excessive missing data
- Investigate segregation distortion
- Verify genetic maps/marker positions
- Look for genotyping errors
- Look at counts of crossovers

References

- Beavis WD (1994). The power and deceit of QTL experiments: Lessons from comparative QTL studies. In DB Wilkinson, (ed) 49th Ann Corn Sorghum Res Conf, pp 252–268. Amer Seed Trade Asso, Washington, DC.

  Discusses selection bias in estimated QTL effects.

- Broman KW (2003) Mapping quantitative trait loci in the case of a spike in the phenotype distribution. Genetics 163:1169–1175

  Two-part model; also discusses binary traits and non-parametric QTL mapping.

- Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315–324

  Haley-Knott regression

- Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. Genetics 159: 371–387

  Multiple imputation

- Solberg LC, et al. (2004) Sex- and line-specific lineage inheritance of depression-like behavior in the rat. Mamm Genome 15:648–662

  Additive and interactive covariates.

- Broman KW et al (2006) The X chromosome in quantitative trait locus mapping. Genetics 174:2151–2158