

# Binary trait locus mapping in experimental crosses

Ani Manichaikul and Karl Broman

Department of Biostatistics  
Johns Hopkins University



# Overview

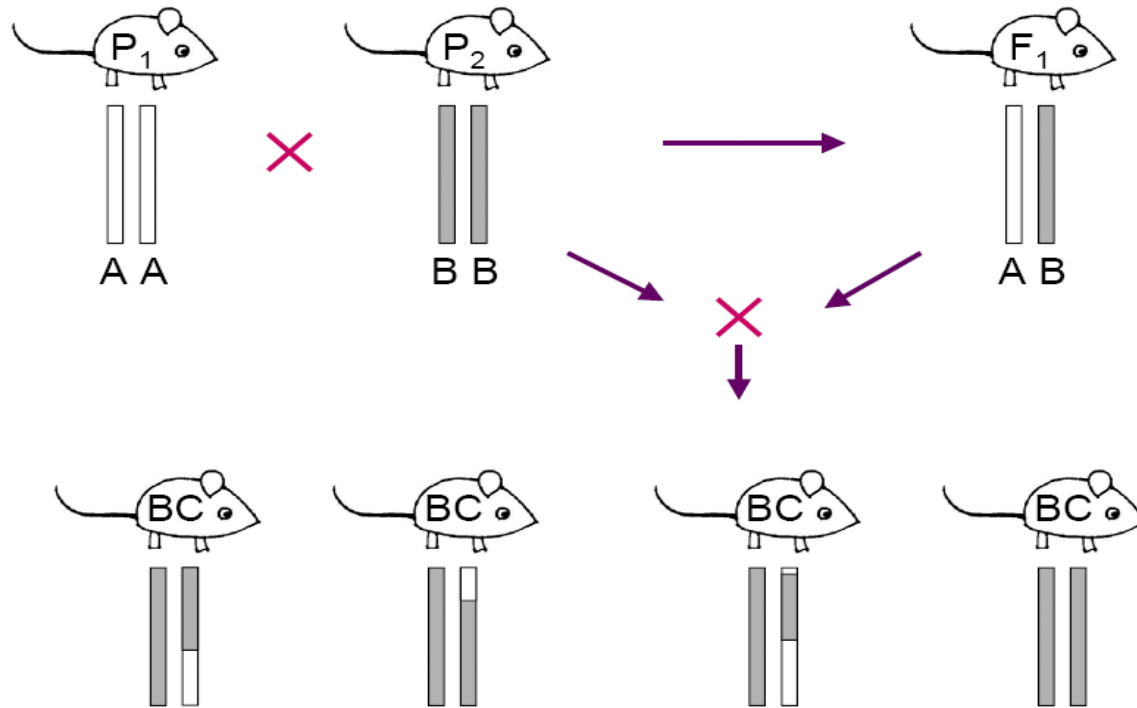
- Inbred lines and experimental crosses
- Mapping genes for binary traits
- Deal with selectively genotyped data
- Avoid spurious linkage due to segregation distortion, if possible



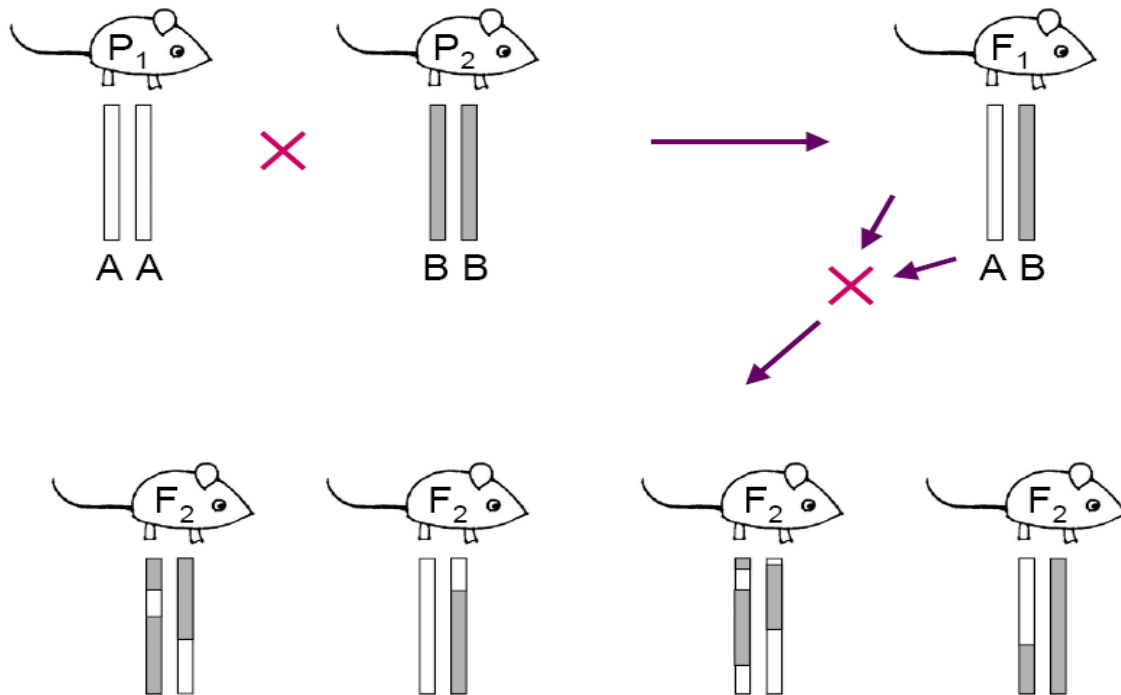
# Motivation

- Develop methods for binary trait locus mapping
- Allow for incomplete genotyping strategies
- Understand relationships between various likelihood approaches

# Backcross



# Intercross



# Binary trait mapping

Analysis of a 2 x 2 table

		Phenotype		Genotype total
		Aff	Unaff	
Genotype	AA	$n_{AA,D}$	$n_{AA,\bar{D}}$	$n_{AA}$
	AB	$n_{AB,D}$	$n_{AB,\bar{D}}$	$n_{AB}$
Phenotype total		$n_D$	$n_{\bar{D}}$	$N$

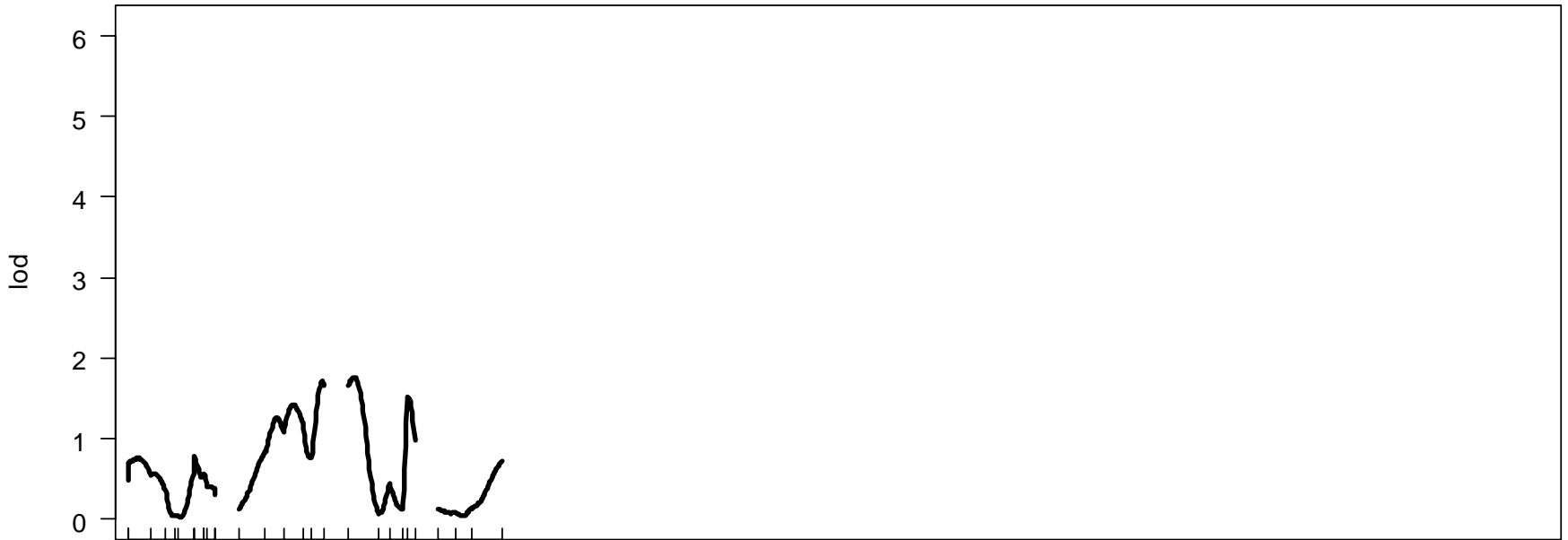
Standard approach:

examine phenotypes conditional on genotypes

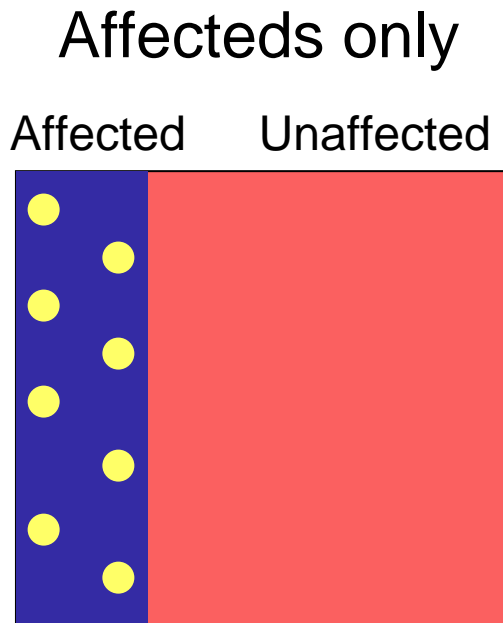
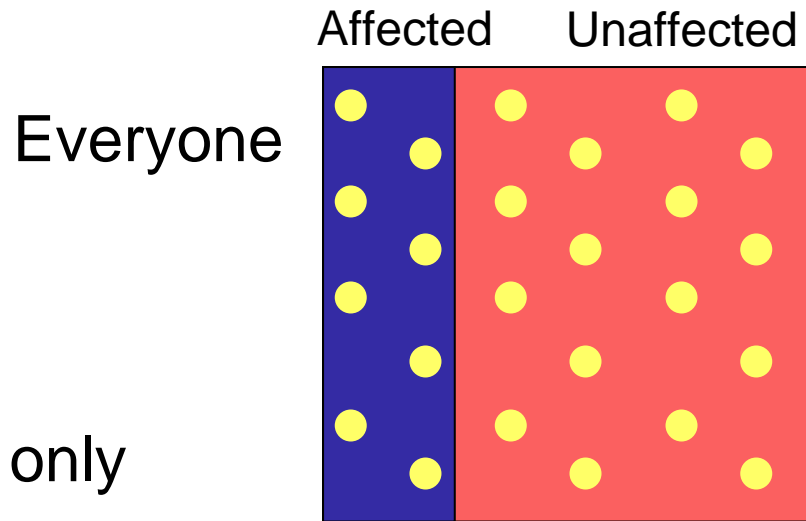
$$LOD_S = \log_{10} \left\{ \frac{\Pr(\text{phenotypes} \mid \text{genotypes})}{\Pr(\text{phenotypes})} \right\}$$

# Binary trait LOD score

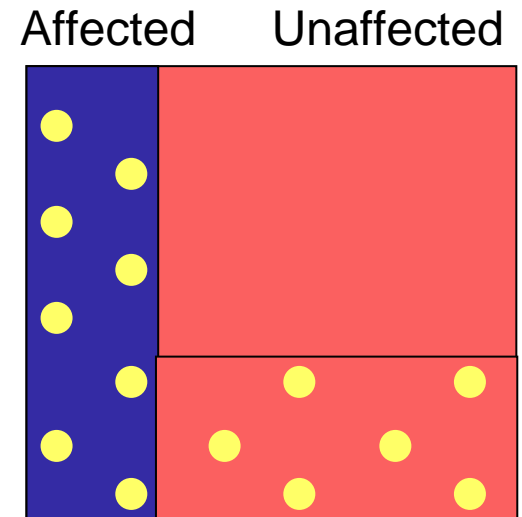
Listeria data: death due to bacterial infection



# Selective genotyping strategies



All affecteds  
and some  
unaffecteds



# Analysis for partial genotyping

- Recall: 
$$LOD_S = \log_{10} \left\{ \frac{\Pr(\text{phenotypes} \mid \text{genotypes})}{\Pr(\text{phenotypes})} \right\}$$
- Standard approach only works assuming complete genotype data
- For selective genotyping, we cannot estimate  $\Pr(\text{phenotypes} \mid \text{genotypes})$  without accounting for missing data

# Reverse approach

- Assume no segregation distortion:

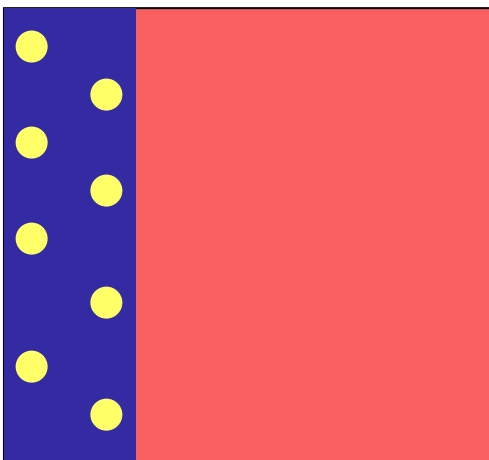
- $\Pr(AA) = \Pr(AB) = \frac{1}{2}$

- Consider:

$$LOD_{R, \frac{1}{2}} = \log_{10} \left\{ \frac{\Pr(\text{genotypes} | \text{phenotypes})}{(1/2)^N} \right\}$$

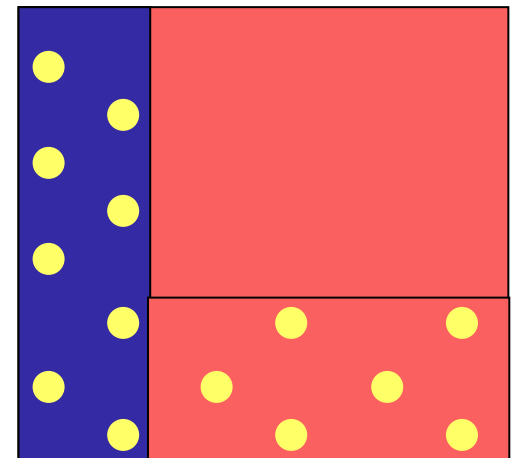
Affecteds only

Affected      Unaffected



All affecteds and some unaffecteds

Affected      Unaffected



# Relationship

- Algebra yields

$$LOD_{R, \frac{1}{2}} = LOD_S + LOD_{\text{segregation distortion}}$$

- Under selective genotyping, evidence for segregation distortion would be relevant
- For complete genotype data, evidence segregation distortion could lead to false positives

# Revisit standard approach

- Assume no segregation distortion
- Can include observations with no genotypes
  - Let  $\pi_0 = \Pr(\text{disease} | AA)$  and  $\pi_1 = \Pr(\text{disease} | AB)$
  - Then incorporate ungenotyped individuals into the likelihood as:

$$\Pr(\text{affected}_i | \text{no genotype}) = \frac{\pi_0 + \pi_1}{2}$$

- The LOD score is written as

$$LOD_{SM, \frac{1}{2}} = \log_{10} \left\{ \frac{\Pr(\text{pheno} | \text{geno}; \text{no seg. distortion})}{\Pr(\text{pheno})} \right\}$$

# Estimate marginal marker probabilities

- Instead of assuming  $\Pr(AA) = \Pr(AB) = \frac{1}{2}$ , could plug in estimates
- Define  $p_{AA} = \Pr(AA | \text{aff}) \cdot \Pr(\text{aff}) + \Pr(AA | \text{unaff}) \cdot \Pr(\text{unaff})$
- Obtain estimates from full likelihood:  
$$\Pr(\text{geno}, \text{pheno}) = \Pr(\text{geno} | \text{pheno}) \cdot \Pr(\text{pheno})$$
- Motivation:
  - Protect against genotyping error
  - Avoid irrelevant evidence
  - Account for bias introduced by selective genotyping
  - What we do if we have complete data

# Reverse approach using marginal marker probabilities

In reverse approach, plug  $\hat{p}_{AA}$  into the denominator instead of requiring assumption of no segregation distortion

$$LOD_{R, \hat{p}_{AA}} = \log_{10} \left\{ \frac{\Pr(\text{genotypes} | \text{phenotypes})}{\Pr(\text{genotypes}; \hat{p}_{AA})} \right\}$$

# Standard approach using marginal marker probabilities

- Ungenotyped individuals can be incorporated into the standard likelihood as

$$\Pr(\text{affected}_i \mid \text{no genotype}) = \pi_0 \cdot \hat{p}_{AA} + \pi_1 \cdot (1 - \hat{p}_{AA})$$

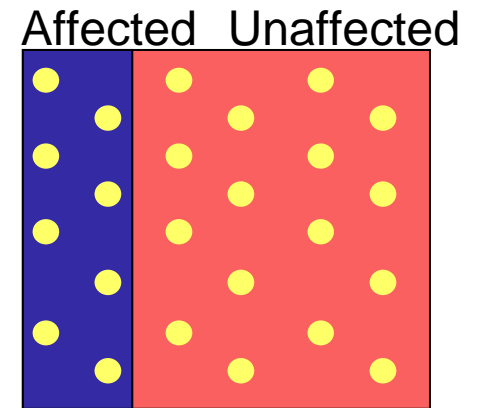
- The corresponding LOD score is written as

$$LOD_{SM, \hat{p}_{AA}} = \log_{10} \left\{ \frac{\Pr(\text{pheno} \mid \text{geno}; \hat{p}_{AA})}{\Pr(\text{pheno})} \right\}$$

# Summary of methods

	No segregation distortion	Estimate marginal genotype probabilities
Standard approach (condition on genotypes)	None	$LOD_S$
Reverse approach (condition on phenotypes)	$LOD_{R, \frac{1}{2}}$	$LOD_{R, \hat{p}_{AA}}$
Standard approach incorporating missing genotypes	$LOD_{SM, \frac{1}{2}}$	$LOD_{SM, \hat{p}_{AA}}$

# Relationships: Complete data



No segregation distortion

Estimate marginal  
genotype probabilities

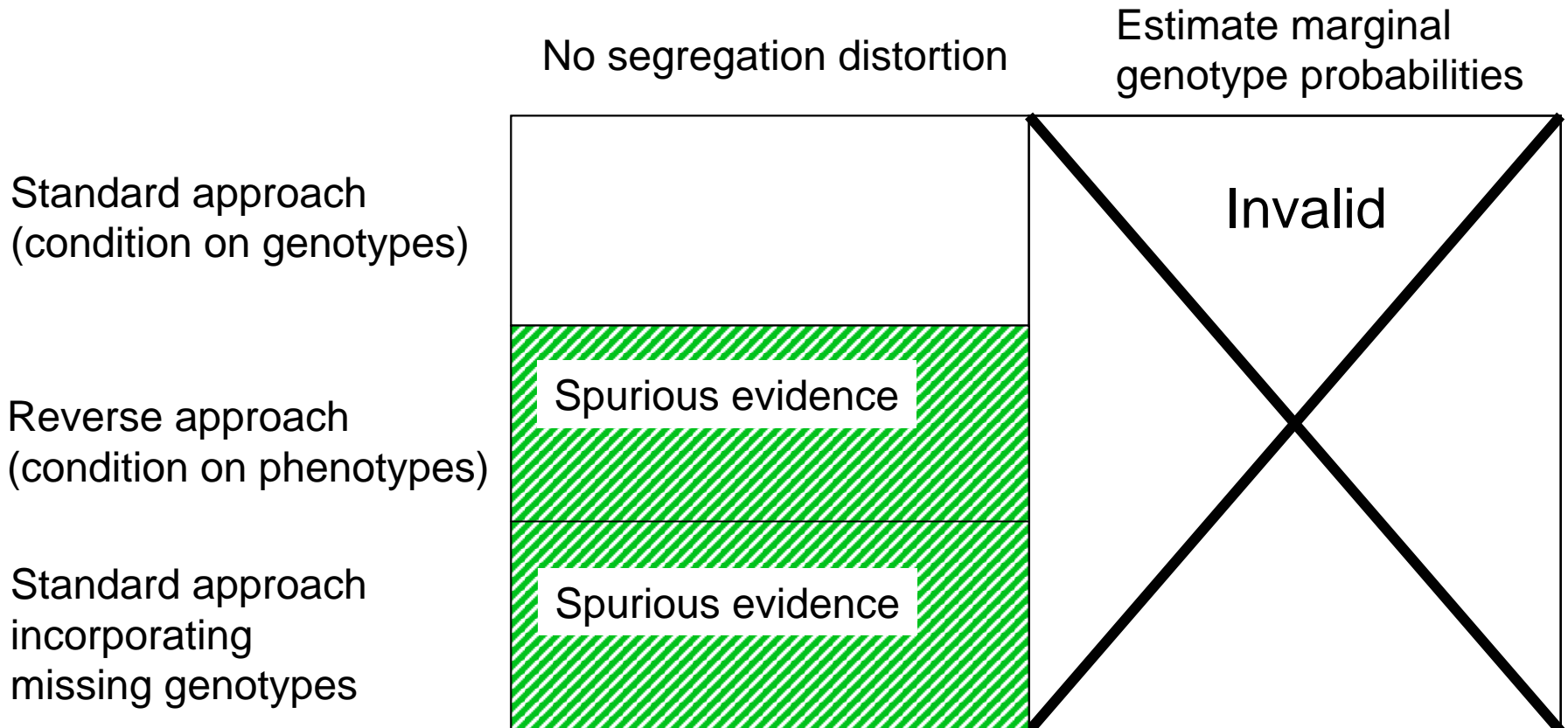
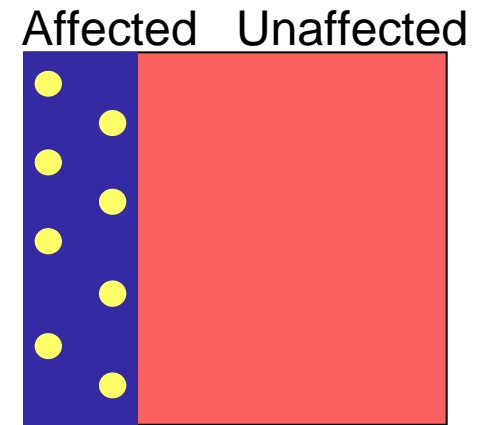
Standard approach  
(condition on genotypes)

Reverse approach  
(condition on phenotypes)

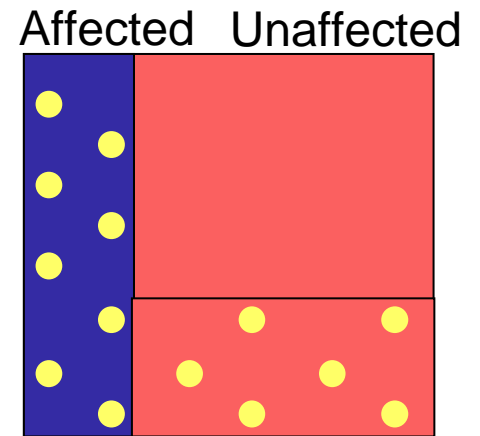
Standard approach  
incorporating  
missing genotypes

Spurious evidence		

# Relationships: Affecteds only



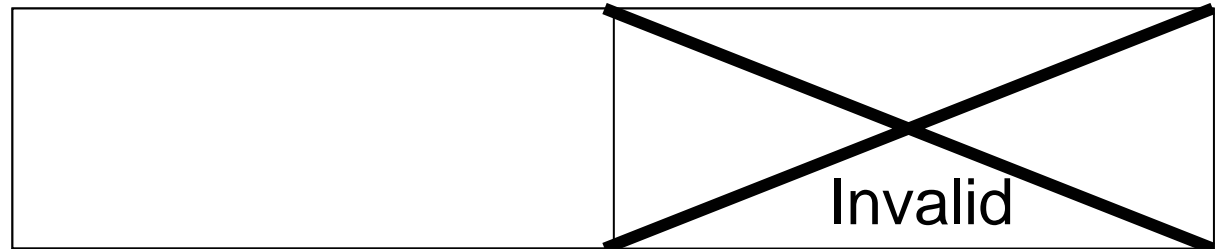
# All affecteds and some unaffecteds



No segregation distortion

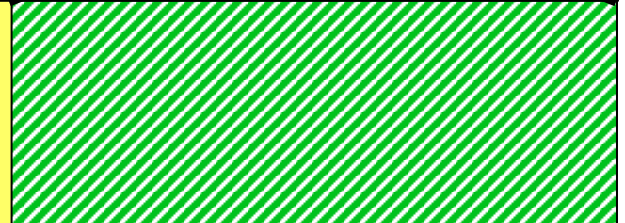
Estimate marginal genotype probabilities

Standard approach  
(condition on genotypes)



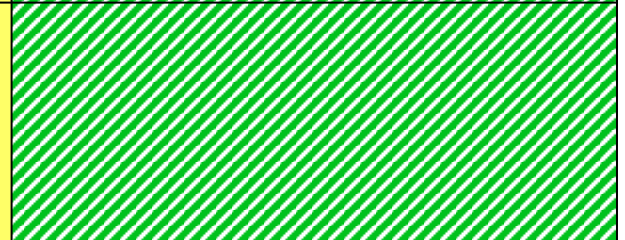
Reverse approach  
(condition on phenotypes)

Spurious evidence

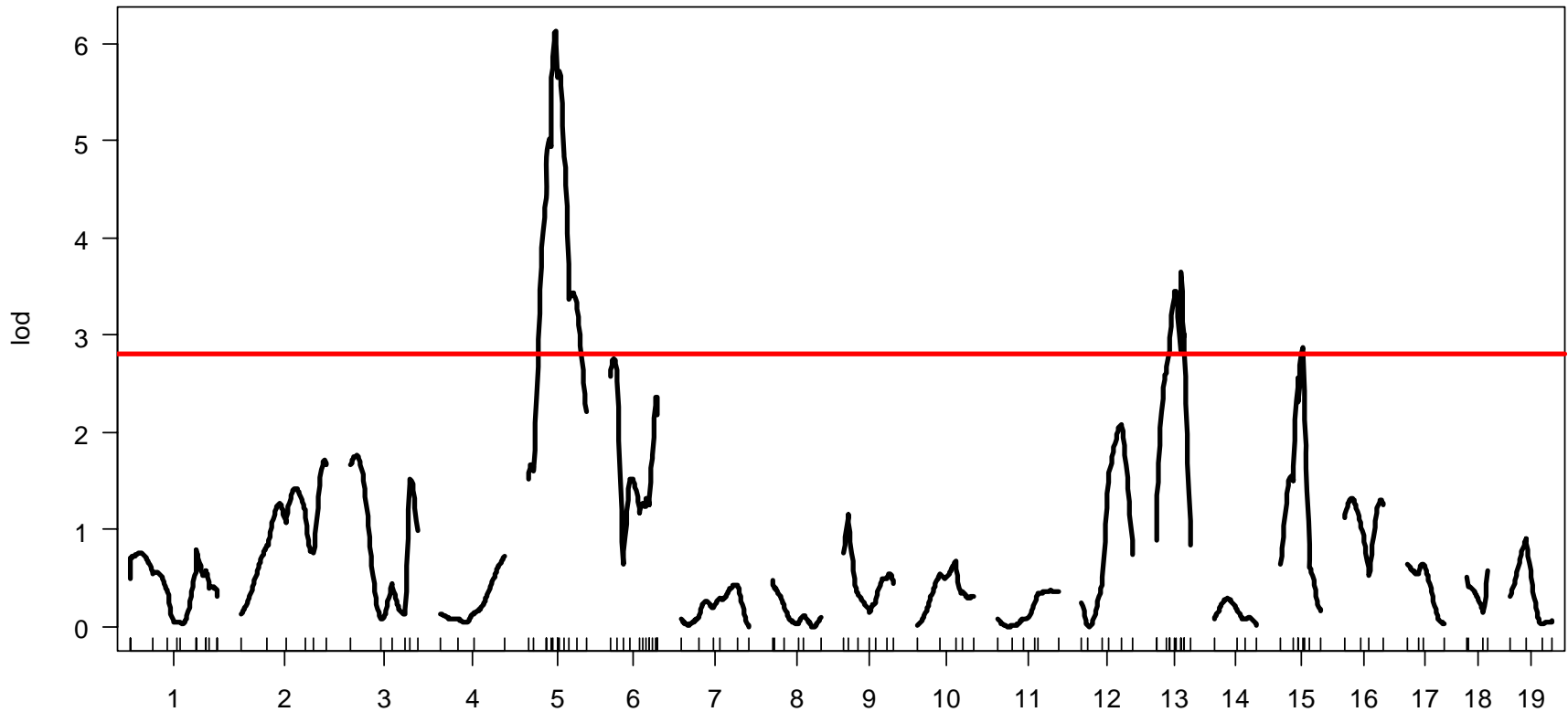


Standard approach  
incorporating  
missing genotypes

Spurious evidence



# Genome-wide thresholds





# Genome-wide thresholds

## ■ Permutation

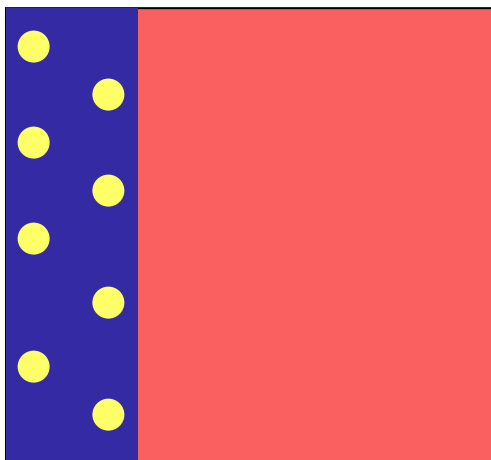
- Typical approach
- Conditions on observed genotype and phenotype distributions
- Works for complete data setting (or missing at random)

# Genome-wide thresholds

What's wrong with permutation test for selective genotyping case?

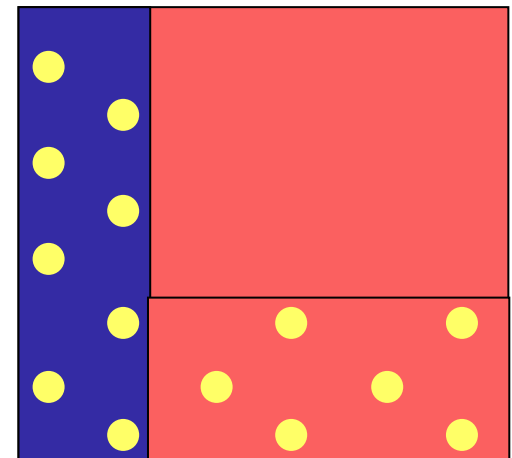
Affecteds only

Affected      Unaffected



All affecteds and some unaffecteds

Affected      Unaffected



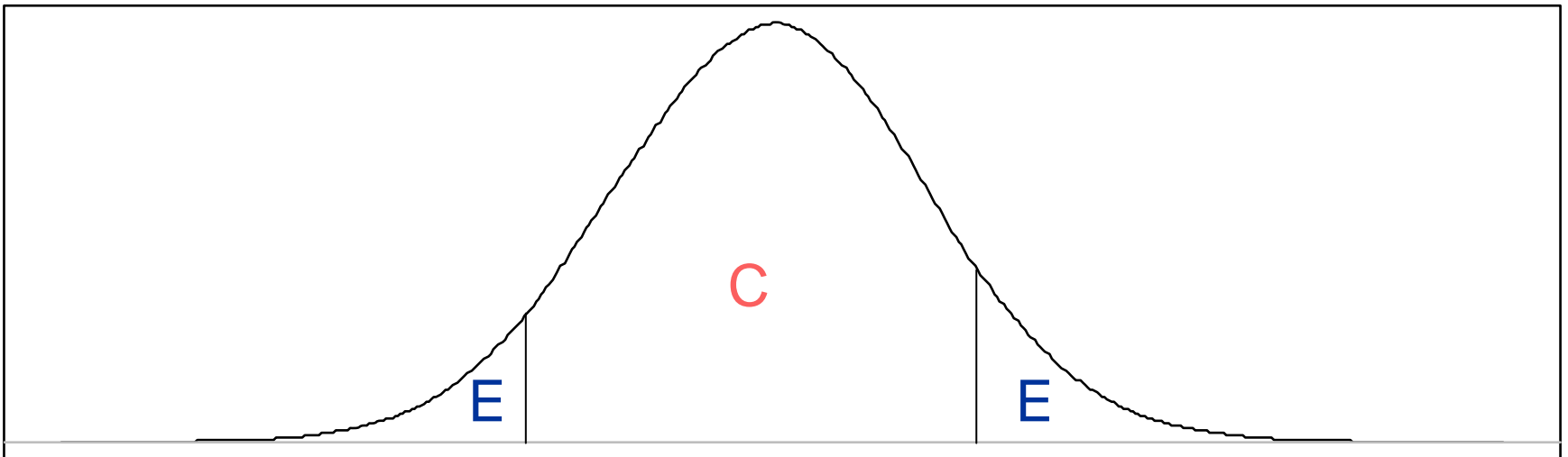


# Simulation test

- Rather than permuting genotypes, replace observed genotypes according to the null model
- Simulate genotypes according to
  - $\Pr(AA) = \Pr(AB) = \frac{1}{2}$
  - or, use estimates of marginal genotype probabilities
- Approach still conditions on observed phenotype distribution

# Continuous trait extension

- Focus genotyping on individuals with extreme phenotypes
- Could incorporate missing genotypes into the standard approach by estimating  $\Pr(AA)$  averaged over phenotype groups





# Summary

- Standard approach works for complete data and missing at random
- Assume  $P(AA) = P(AB) = \frac{1}{2}$  if genotyping affecteds only
- For selective genotyping, estimate marginal genotype probabilities from observed genotypes