5. Sample Size, Power & Thresholds

- review of sample size for *t*-test (known QTL)
- sample size, marker spacing & power
  - Zeng talk at Plant & Animal Genome XI, 2003 (statgen.ncsu.edu/zeng/QTLPower-Presentation.pdf)
- QTL thresholds & genome-wide error rates
- positive false detection rates
- selection bias

5.1 review sample size for *t*-test

- set up test under "null hypothesis" of no QTL
  - two-sided test for difference
  - size arbitrary--usually $\alpha = .05$
- determine power or sample size under alternative
  - power = chance to reject null if alternative true
  - sample size ($n/4$ per *MM,mm*) follows square root idea

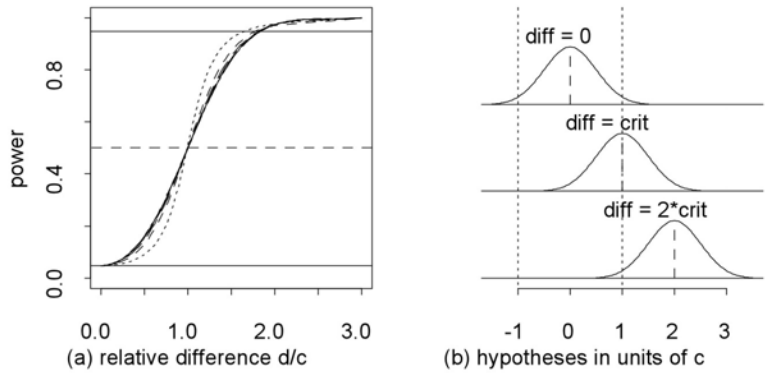
$$t = \frac{\mu_{MM} - \mu_{mm}}{\sigma\sqrt{8/n}} = \frac{d}{SE}$$

$$\text{size/2}: \text{pr}(\mu_{MM} - \mu_{mm} > c \mid \text{no QTL}) = \alpha/2$$

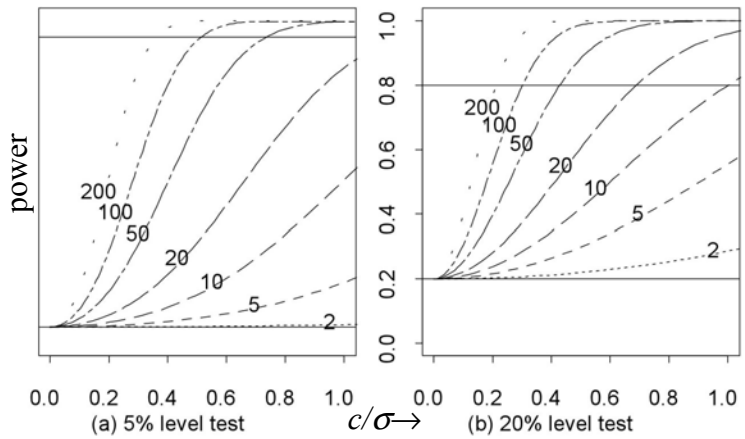
$$\text{power}: \text{pr}(\mu_{MM} - \mu_{mm} > c \mid \text{QTL}) = 1 - \beta$$

$$n = 2[(z_{\alpha/2} + z_{\beta})\sigma/d]^2$$

rough guide: $d = 2c = 4SE$ for power = $1-\alpha/2$



power increase with sample size (numbers in figures are $n/4$)



5.3 QTL thresholds for IM

- IM scans across loci λ in genome
  - evidence for QTL in large LOD(λ)
  - LOD(λ) = 0.217×LR
- set genome-wide LOD threshold
  - protect against one or more false positives
  - roughly adjust size of each individual test
- LOD distribution under null hypothesis
  - at any particular location λ
  - depends on design (1 df for BC, 2 df for F2)

$$\frac{LOD(\lambda)}{0.217} = LR \sim \chi_1^2 \text{ or } \chi_2^2$$

threshold: from point to genome
(developed for BC)

- maximum LR within marker interval
  - χ^2 with d.f. between 1 and 2
  - d.f. close to 1 for small interval
- sparse marker map across genome
  - assume M independent marker intervals
  - use Bonferroni correction to level ($\alpha \rightarrow \alpha/M$)

$$\max_{\lambda \text{ in interval}} LR(\lambda) \sim \chi_\nu^2, 1 < \nu < 2$$

$$\max_{\lambda \text{ in genome}} LR(\lambda) \sim ?$$

LR, LOD and point-wise *p*-values

LR	LOD	*p*-value 1 d.f.	*p*-value 2 d.f.
10	1	0.0319	0.1
31.6	1.5	0.0086	0.0316
100	2	0.0024	0.01
1000	3	0.0002	0.001
10000	4	<0.0001	0.0001

genome-wide threshold: theory

- dense marker map: markers everywhere
- LR test statistics are correlated
  - correlation drops off quickly with distance
  - no correlation for unlinked markers
- theory: Ohrnstein-Uhlenbeck process
  - C = number of chromosomes,
  - G = length of genome in cM
  - t = genome-wide threshold value
  - α_t= corresponding point-wise significance level

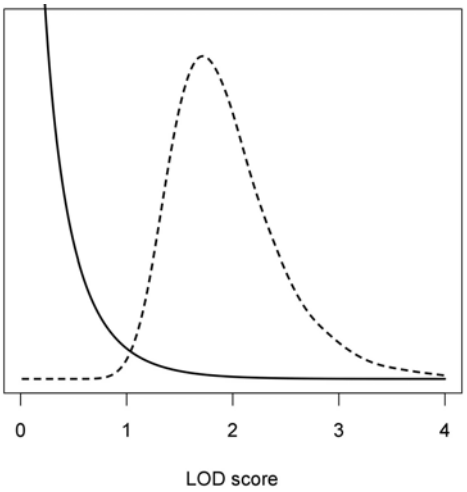
$$\text{pr}(\max_{\lambda \text{ in genome}} LR(\lambda) > t) = \alpha \approx (C + 2Gt)\alpha_t$$

with $\text{pr}(\chi^2_1 > t) = \alpha_t$

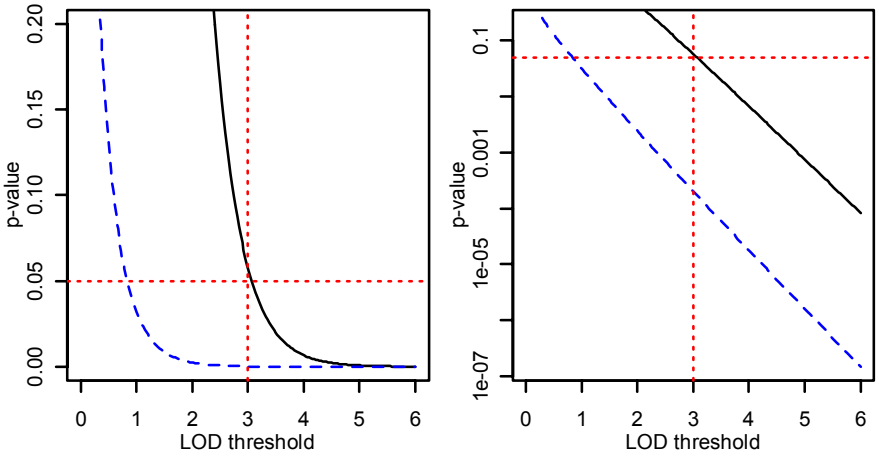
point-wise & genome-wide distributions

- backcross
- one QTL
- one locus (solid)
- dense map (dash)
- pr(LOD>*t*) = area under curve

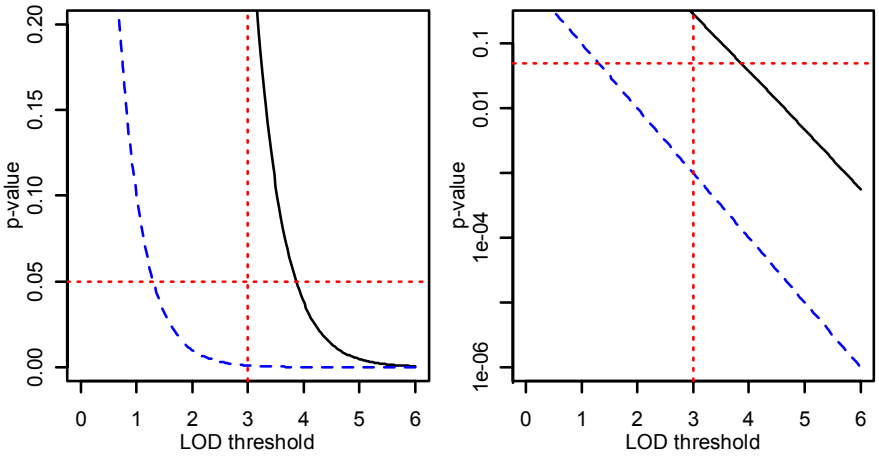
(from Broman 2001 *Lab Animal*)



point-wise and genome-wide *p*-values for backcross



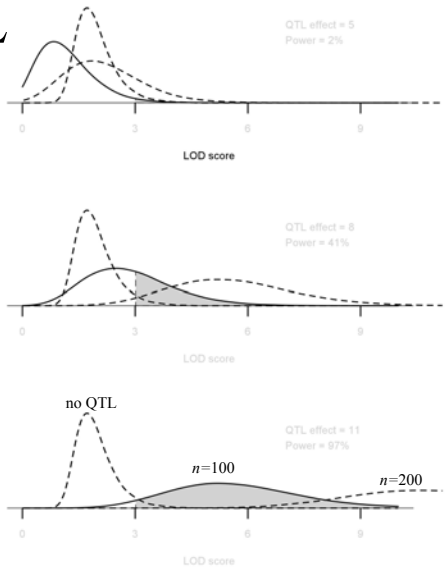
point-wise and genome-wide *p*-values for F2 intercross



LOD with QTL

- backcross
- genome-wide
- 3=threshold
- dash: no QTL
- solid: *n*=100
- dot: *n*=200

(from Broman 2001 *Lab Animal*)



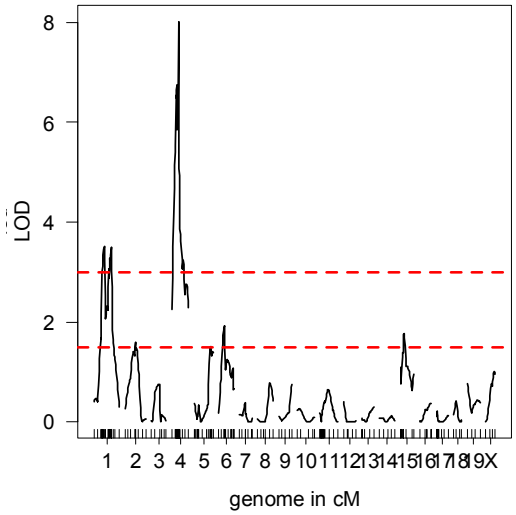
5.4 false detection rates and thresholds

- threshold: test QTL across genome
  - size = pr(LOD(λ) > threshold | no QTL at λ)
  - guards against a single false detection
    - very conservative on genome-wide basis
  - difficult to extend to multiple QTL
- positive false discovery rate (Storey 2001)
  - pFDR = pr(no QTL at λ | LOD(λ) > threshold)
  - post-hoc estimate of proportion of false positives
  - easily extended to multiple QTL in Bayesian context

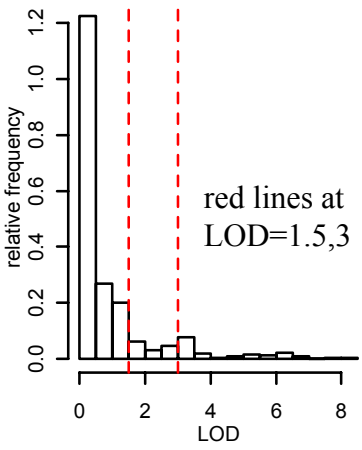
pFDR for interval mapping

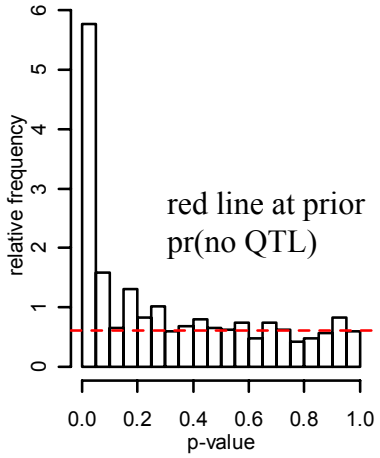
- $\text{pFDR} = \dfrac{\text{pr(no QTL)*size}}{\text{pr(no QTL)*size+pr(QTL)*power}}$
- decide on threshold (e.g. genome-wide)
- power = pr(LOD(λ) > threshold | QTL at λ)
  - proportion of genome with LOD > threshold
- size = pr(LOD(λ) > threshold | no QTL at λ)
  - point-wise *p*-value of threshold
- prior = pr(no QTL at λ)
  - proportion of genome with LOD < "small value"
  - divided by point-wise *p*-value for "small value"

IM for hypertense mice

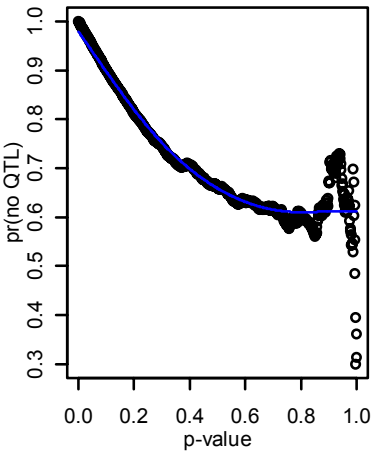


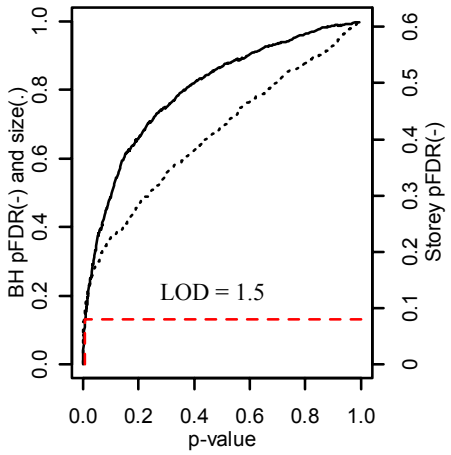
histograms of LOD and *p*-value





pFDR for hyperactive mice

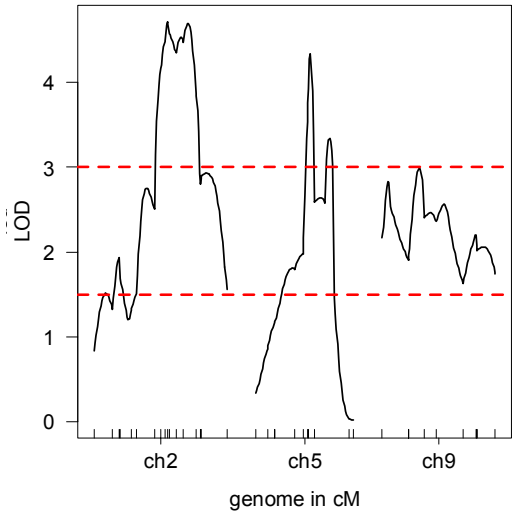




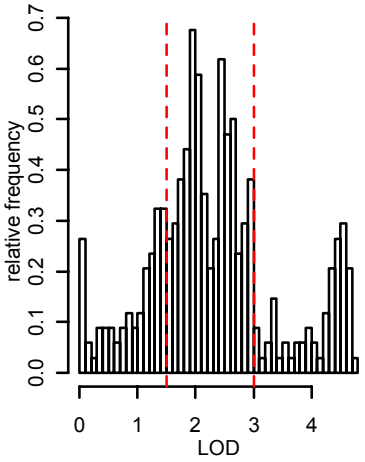
false detection rates and thresholds

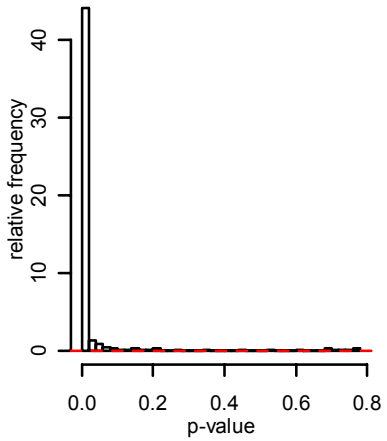
- multiple comparisons: test QTL across genome
  - size = pr(LOD(λ) > threshold | no QTL at λ)
  - threshold guards against a single false detection
    - very conservative on genome-wide basis
  - difficult to extend to multiple QTL
- positive false discovery rate (Storey 2001)
  - pFDR = pr(no QTL at λ | LOD(λ) > threshold)
  - Bayesian posterior HPD region based on threshold
    - $\Lambda = \{\lambda \mid \text{LOD}(\lambda) > \text{threshold}\} \approx \{\lambda \mid \text{pr}(\lambda \mid Y,X,m) \text{ large}\}$
  - extends naturally to multiple QTL

SCD expression in B6/BTBR

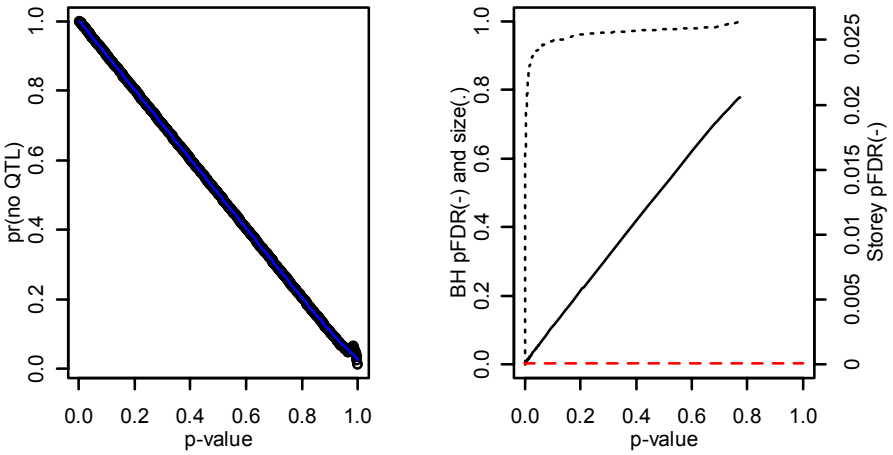


histograms of LOD and *p* for SCD





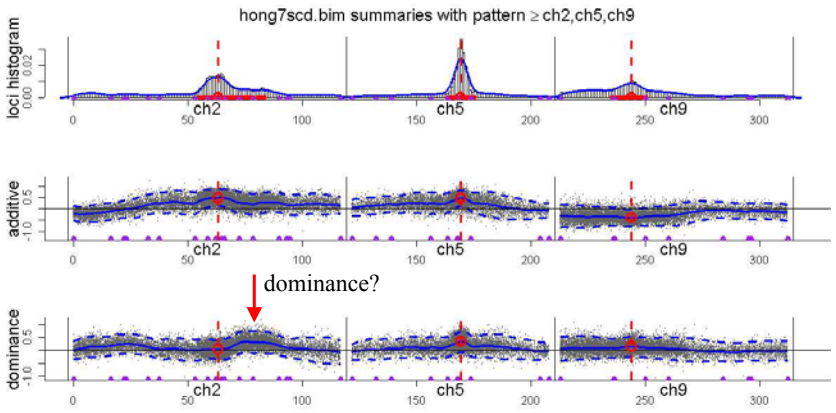
pFDR estimation curves for SCD



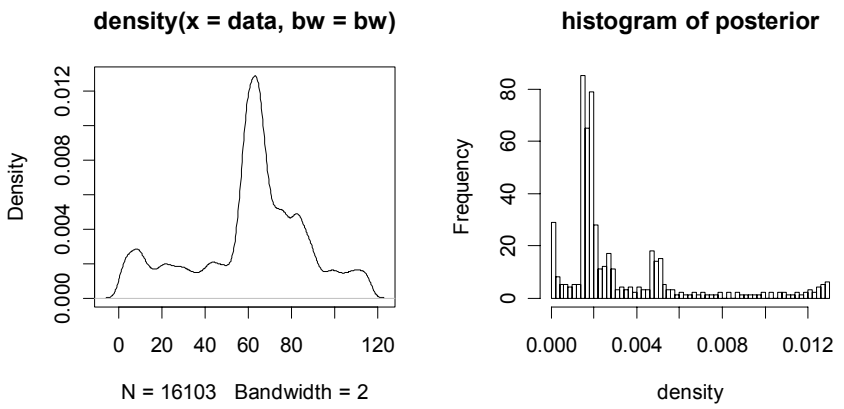
pFDR and QTL posterior

- positive false detection rate
  - pFDR = pr(no QTL at λ | *Y,X,* λ in Λ)
  - pFDR = $\dfrac{\text{pr}(H=0)*\text{size}}{\text{pr}(m=0)*\text{size}+\text{pr}(m>0)*\text{power}}$
  - power = posterior = pr(QTL in Λ | *Y,X,* $m>0$)
  - size = (length of Λ) / (length of genome)
- extends to other model comparisons
  - $m = 1$ vs. $m = 2$ or more QTL
  - pattern = ch1,ch2,ch3 vs. pattern > 2*ch1,ch2,ch3

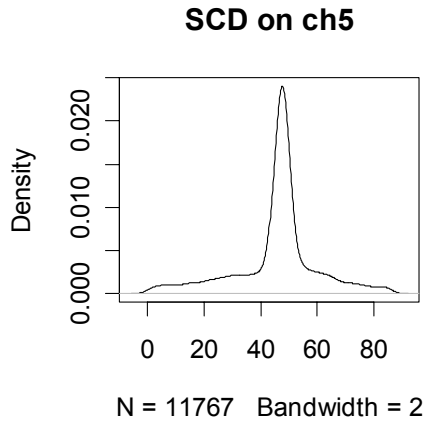
trans-acting QTL for SCD1

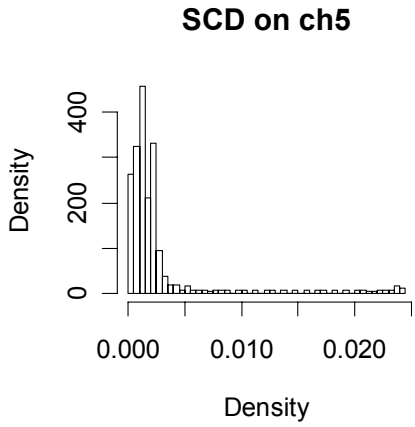


SCD posterior for chr 2

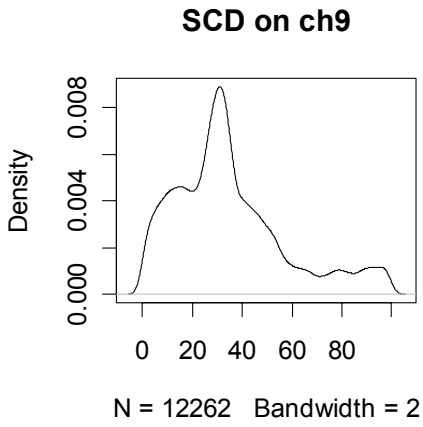


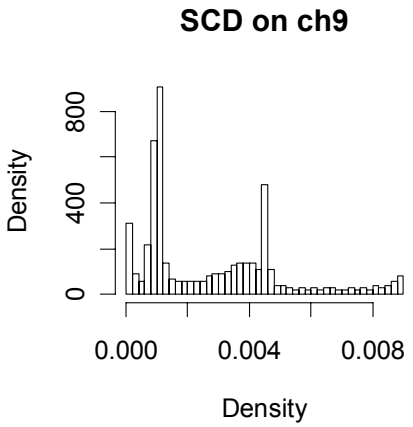
SCD posterior for chr 5



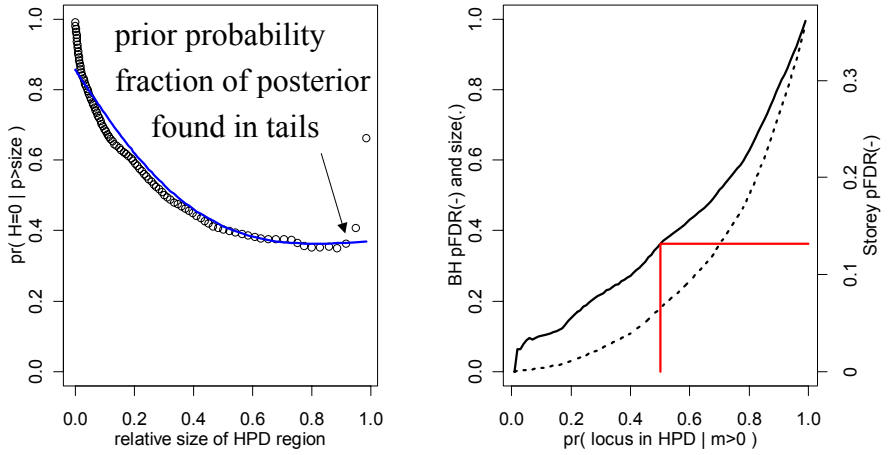


SCD posterior for chr 9



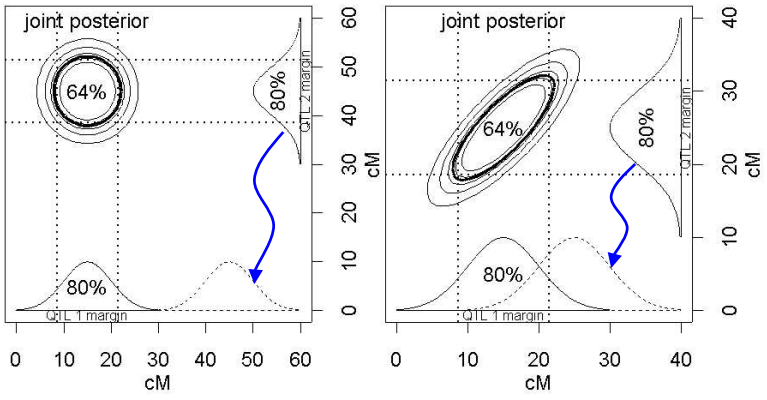


pFDR for SCD1 analysis



1-D and 2-D marginals
pr(QTL at λ | *Y,X, m*)

unlinked loci linked loci



5.5 selection bias

- locus bias: locus picked via maximum LOD(λ)
  - danger of ghost QTL (Haley Knott 1992)
  - CIM or Haley-Knott estimate can be biased
- effect bias: estimated effects θ often too large
  - Utz, Melchinger Schön (2000); Broman (2001)
- consider moderate sized effect
  - chance of missing QTL: undetected QTL not reported
  - estimated effect is conditional on detection
    - estimates of effects and genetic variance can be inflated
    - confidence intervals for locus can be quite large

selection bias

white=missed

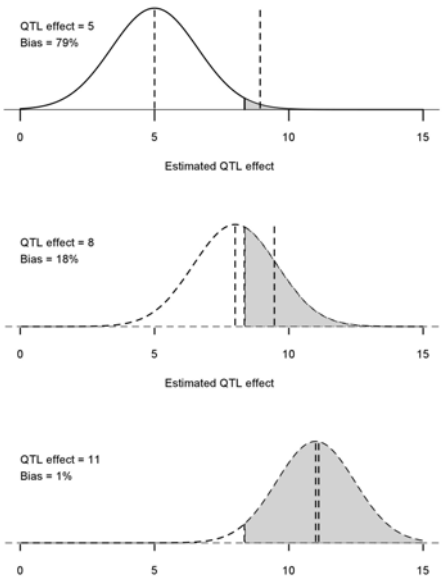
grey = observed

above 8 (say)

left dash = true mean

right dash = obs. mean

(from Broman 2001 *Lab Animal*)



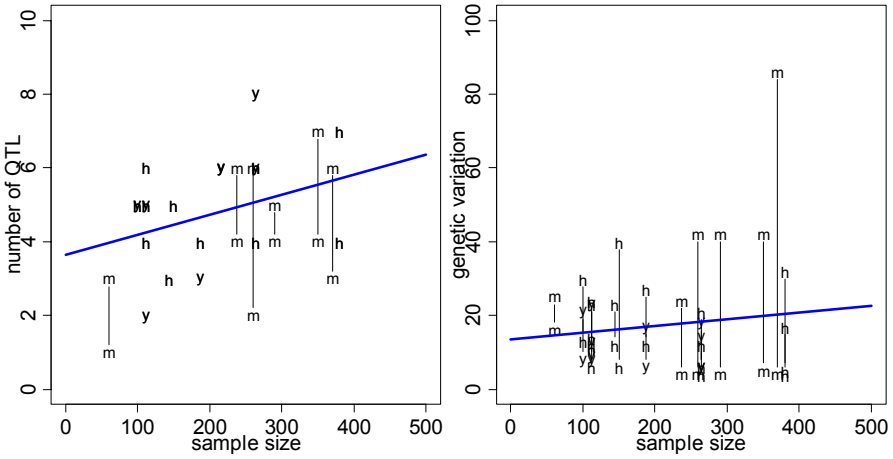
independent validation

- separate study on same parent lines
  - Beavis (1994) simulation study
  - Beavis et al. (1994 *Crop Sci*) field study
  - Melchinger et al. (1998 *Genetics*) field study
- resampling evaluation
  - Beavis (1994) simulation study
  - Visscher et al. (1996) bootstrap
  - Utz et al. (2000 *Genetics*) cross validation`

many QTL of small effect?

- repeated field studies yield different QTL
  - environment effect?
  - genetic differences: parents, independent cross?
  - chance variation?
- simulation studies (Beavis 1994, 1998)
  - 10 QTL of same size
  - only chance variation
  - repeated trials yield different QTL again

Beavis (1998) QTL Analyses
estimated number & magnitudes



cross-validation idea

- sample from genetic cross (e.g. F2 intercross)
- divide into several "subsamples" (e.g. k=5)
  - 344 F2 → 69(S1),69(S2),69(S3),69(S4),68(S5)
- ES: estimate with some (4) subsamples
  - predict QTL loci λ and effects θ using ≈CIM
  - predict heritability h^2 :
- TS: test with other (1) subsamples
  - use loci from ES to predict effects θ
  - proportion of genotypic variance explained
    - regress predicted effects on phenotype → adjusted R^2
    - proportion = adjusted R^2 / predicted h^2

$$h^2 = \frac{\sigma_g^2}{\frac{\sigma^2}{re} + \frac{\sigma_{ge}^2}{e} + \sigma_g^2}$$

Utz et al. (2000) design

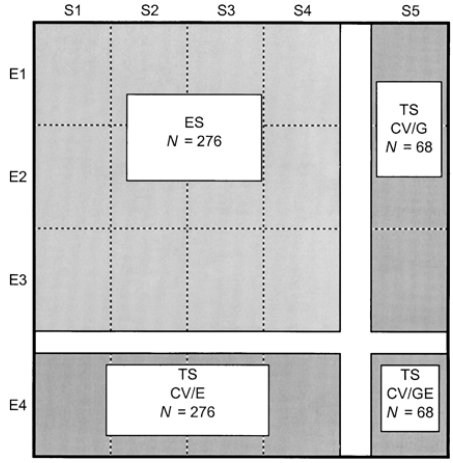
ES: estimation set
- u = 3 environments
- k–1 = 4 subsamples

TS: test sets for CV

CV/G: genotypic sampling
- same environments
- 1 subsample

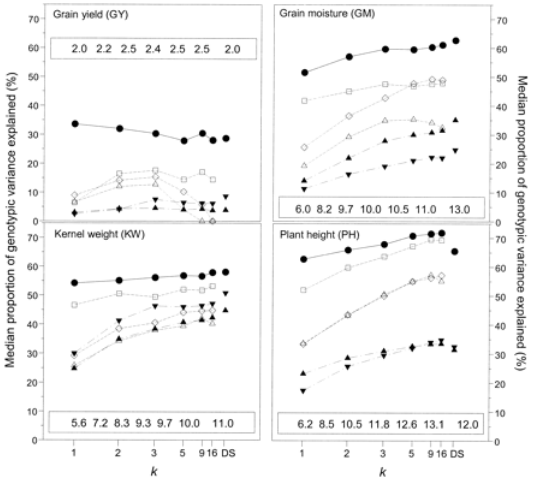
CV/E: environmental sampling
- e = 1 environment
- same subsamples

CV/GE: "unbiased"
- new environment, subsample

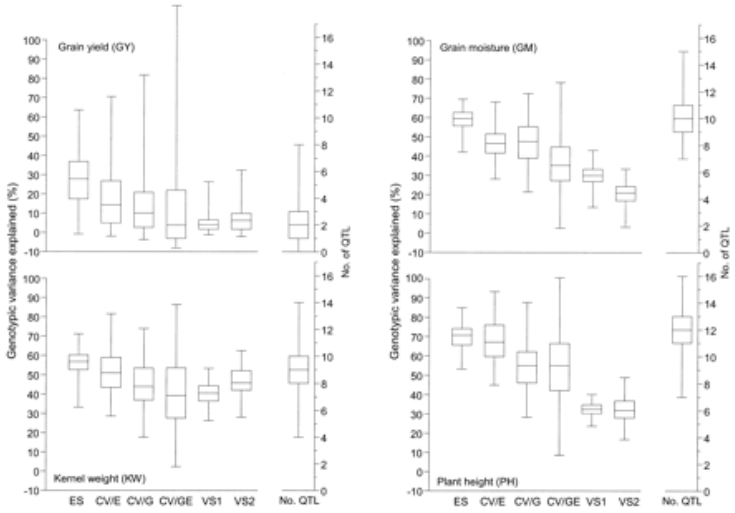


Utz et al. (2000) resampling

- estimation ES(•)
- cross validation
  - CV/E (box)
  - CV/G (diamond)
  - CV/GE (open tri)
- independent validation
  - VS1 (up triangle)
  - VS2 (down triangle)
- subsamples (k)
  - to entire data set (DS)

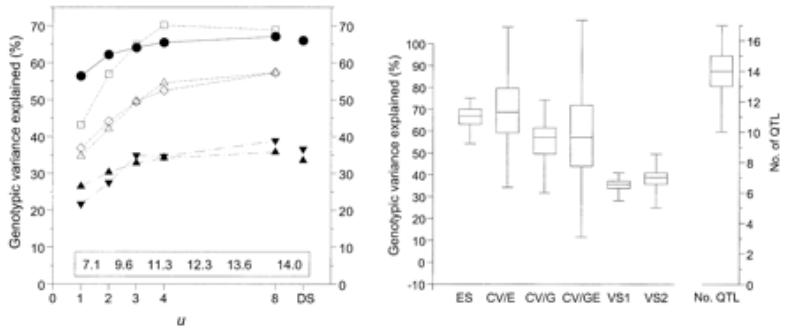


explained genotypic variance

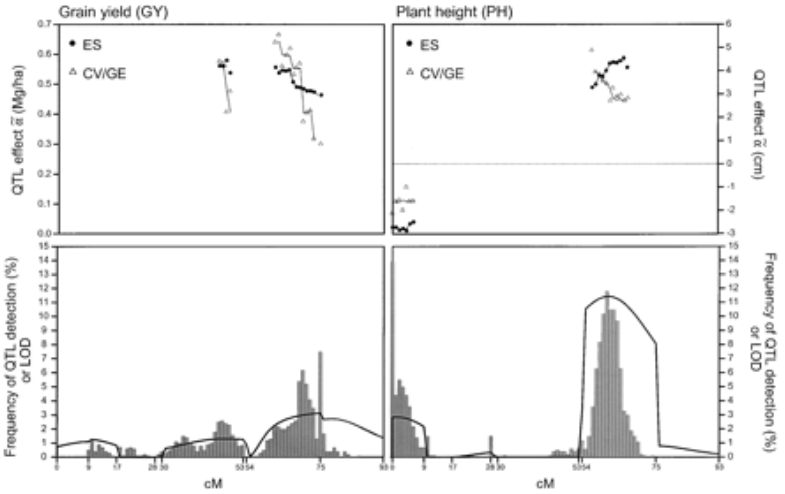


explained variance for plant height

- cross validation: 9 environments
  - u environments for estimation, $e = 9 - u$ for test
- independent: VS1,VS2 = 4,5 environments



1376 CV replicates
estimated effects and LOD by locus



similar tests? a cautionary tale

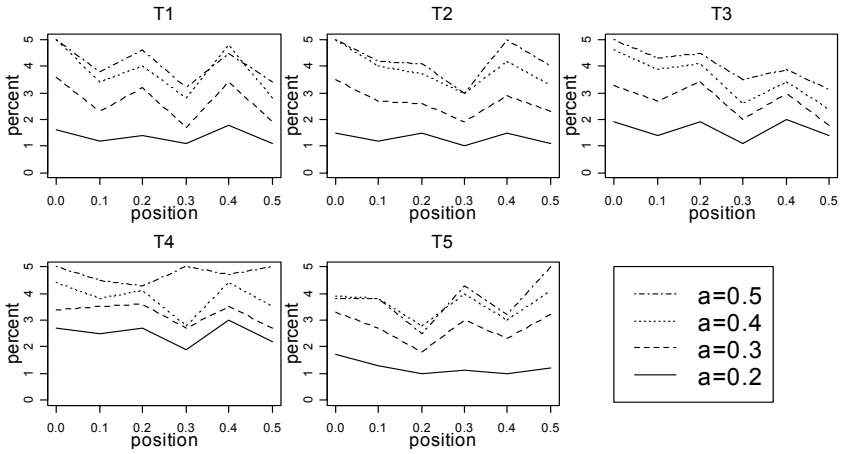
- Goffinet Mangin (1998 *Genetics*)
  - Haley-Knott regression
  - test 1 vs 2 QTL on same chromosome
    - size = pr(detect first and second QTL | only one QTL)
    - power = pr(detect first and second QTL | two QTL)
  - simulations: power to detect more than 1
  - results: tests are not similar
    - Type I error depends on effects, loci
    - 2 QTL model better than 1 QTL
- problems with this approach
  - may be artifact of using Haley-Knott regression
  - size and power conditional on detection of first QTL

tests for second QTL

1. first fixed QTL
   - first QTL position known (estimated)
2. fixed first-QTL interval
   - first QTL interval known
3. multiple QTL
   - CIM (or MQM) model: cofactors
4. two QTL
   - LR test with unconstrained positions
5. likelihood shape
   - correct for bias of QTL at other markers

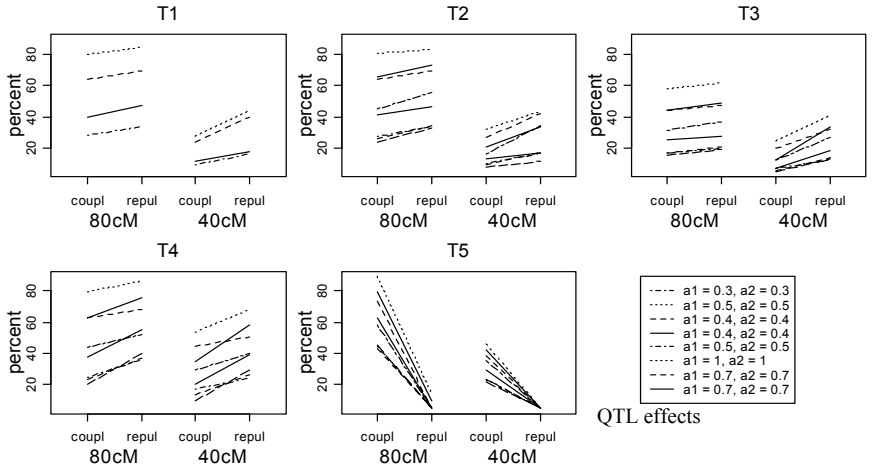
size depends on effect of QTL 1

(Type I error with 5% critical value; Table 1, Goffinet & Mangin 1998)



power depends on effect & distance to linked QTL

(Table 3-4, Goffinet & Mangin 1998)



QTL effects

problem: regression approximation

- paradox goes away if we use likelihood
- recall problems with Haley-Knott:
  - wrong variance
  - mixture over QTL genotypes ignored
- ML yields
  - correct size under null of no QTL
  - high power under alternatives
- Kao (2000 *Genetics*)
  - simulations with h^2 = 0.1,0.3,0.5