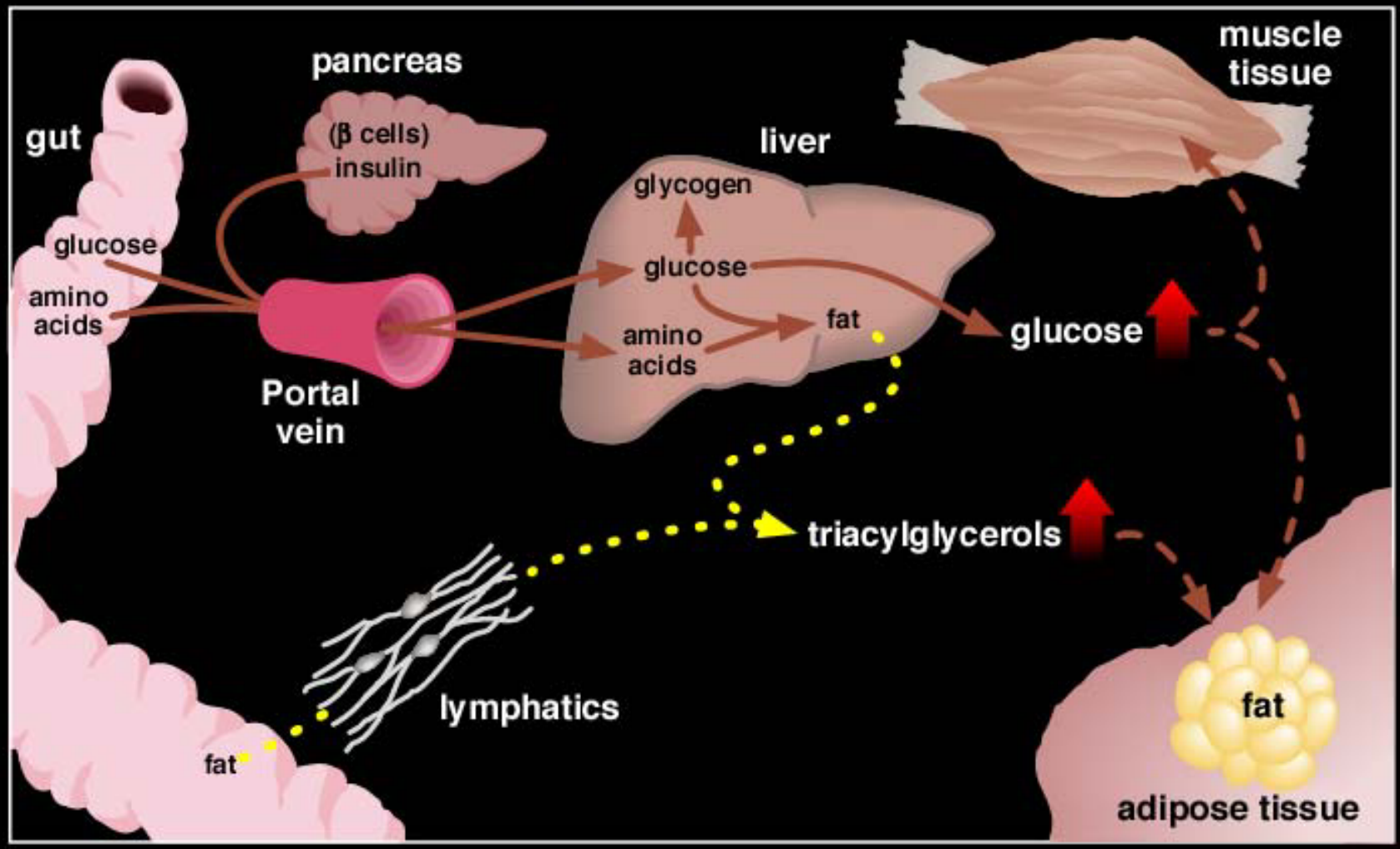
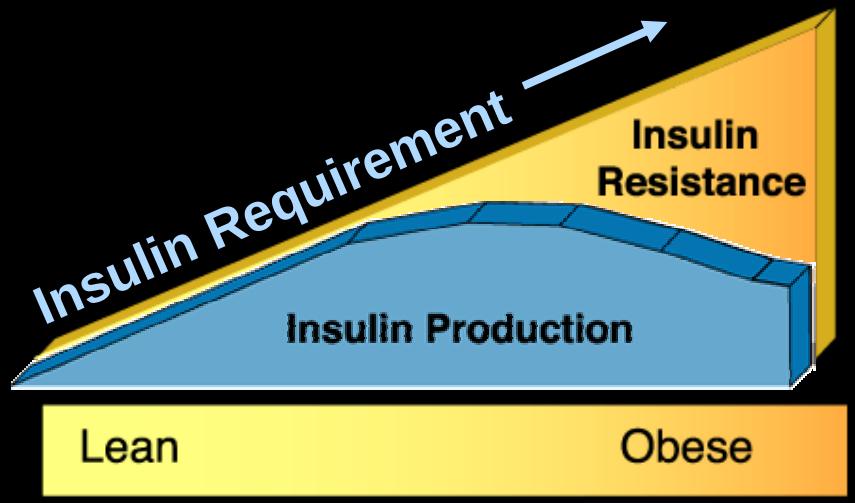
Multiple Traits & Microarrays

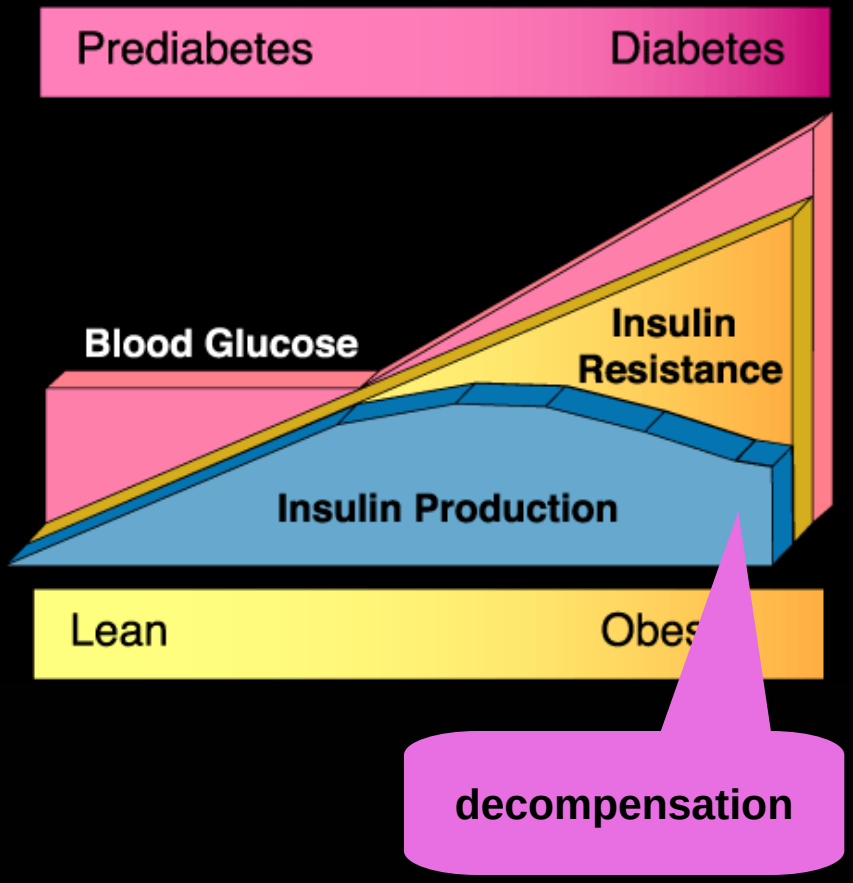
1. why study multiple traits together?

- avoid reductionist approach to biology
  - address physiological/biochemical mechanisms
  - Schmalhausen (1942); Falconer (1952)
- separate close linkage from pleiotropy
  - 1 locus or 2 linked loci?
- identify epistatic interaction or canalization
  - influence of genetic background
- establish QTL x environment interactions
- decompose genetic correlation among traits
- increase power to detect QTL

Type 2 Diabetes Mellitus



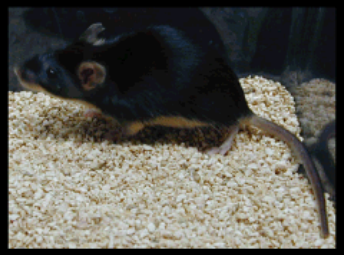




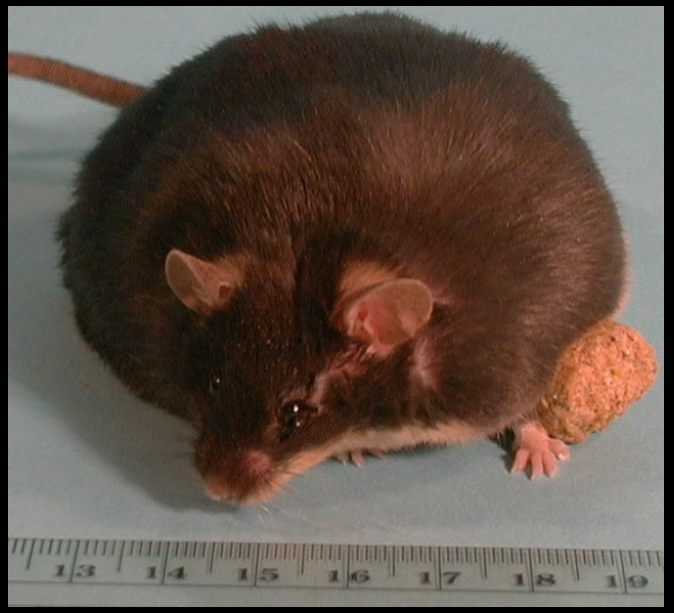
from Unger & Orci *FASEB J.* (2001) 15,312

Insulin Resistant Mice

Bill Dove



BTBR strain



insulin resistance alleles

+

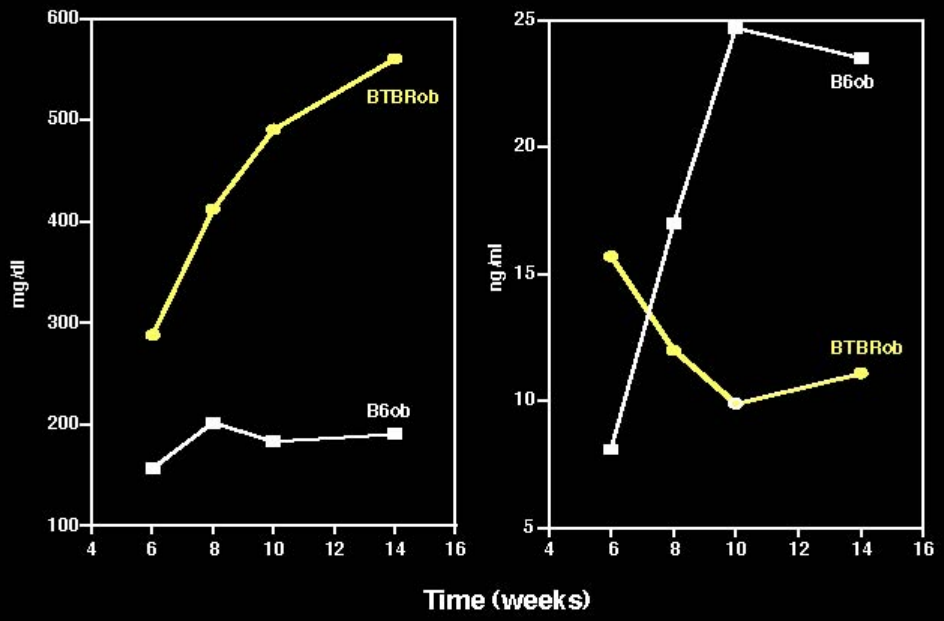
obesity

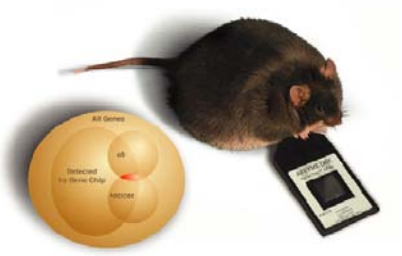
??? → diabetes

(courtesy AD Attie)

glucose

insulin





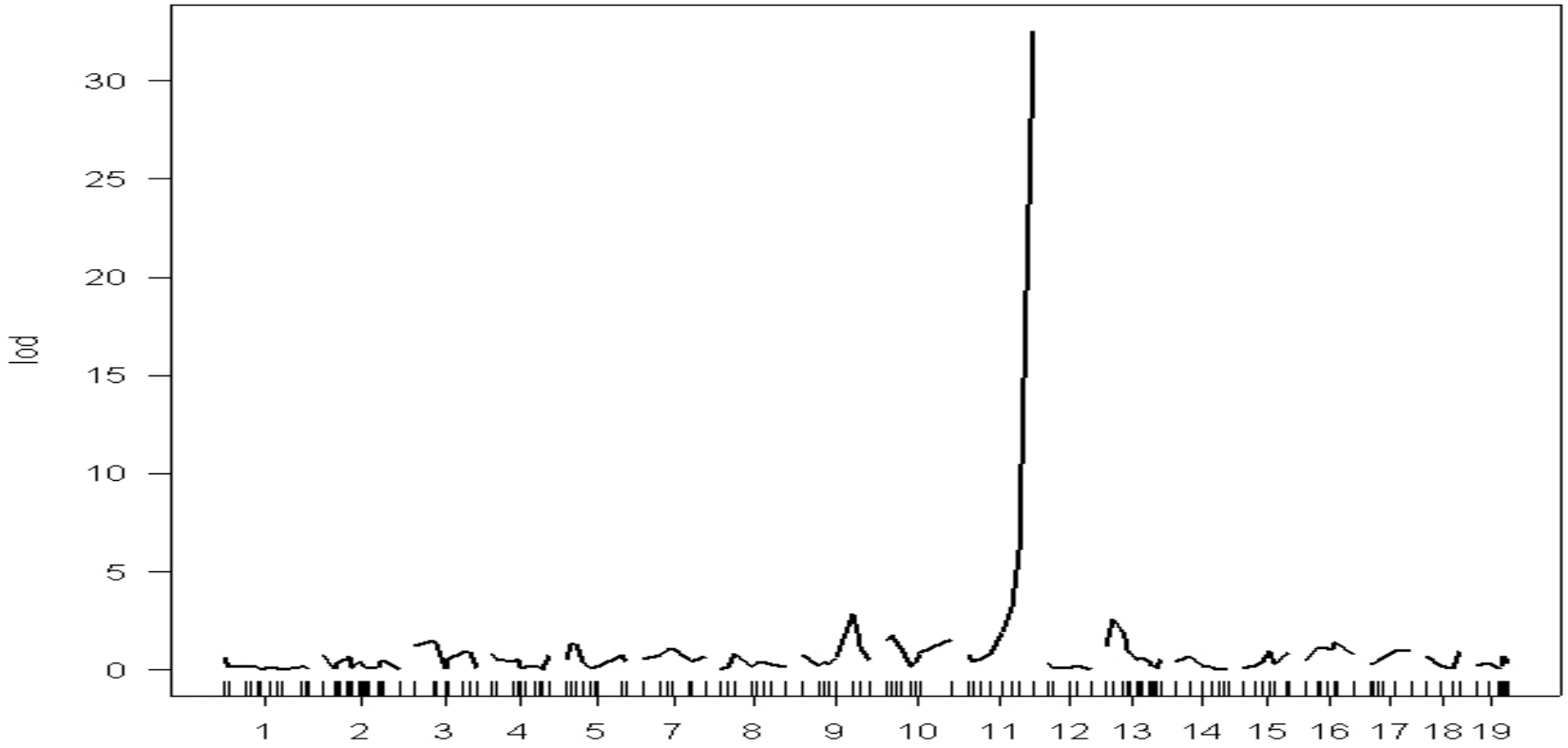
studying diabetes in an F2

- segregating cross of inbred lines
  - B6.ob x BTBR.ob → F1 → F2
  - selected mice with ob/ob alleles at leptin gene (chr 6)
  - measured and mapped body weight, insulin, glucose at various ages (Stoehr et al. 2000 *Diabetes*)
  - sacrificed at 14 weeks, tissues preserved
- gene expression data
  - Affymetrix microarrays on parental strains, F1
    - (Nadler et al. 2000 *PNAS*; Ntambi et al. 2002 *PNAS*)
  - RT-PCR for a few mRNA on 108 F2 mice liver tissues
    - (Lan et al. 2003 *Diabetes;* Lan et al. 2003 *Genetics*)
  - Affymetrix microarrays on 60 F2 mice liver tissues
    - design (Jin et al. 2004 *Genetics* tent. accept)
    - analysis (work in prep.)

why map gene expression as a quantitative trait?

- *cis*- or *trans*-action?
  - does gene control its own expression?
  - or is it influenced by one or more other genomic regions?
  - evidence for both modes (Brem et al. 2002 Science)
- simultaneously measure all mRNA in a tissue
  - ~5,000 mRNA active per cell on average
  - ~30,000 genes in genome
  - use genetic recombination as natural experiment
- mechanics of gene expression mapping
  - measure gene expression in intercross (F2) population
  - map expression as quantitative trait (QTL)
  - adjust for multiple testing

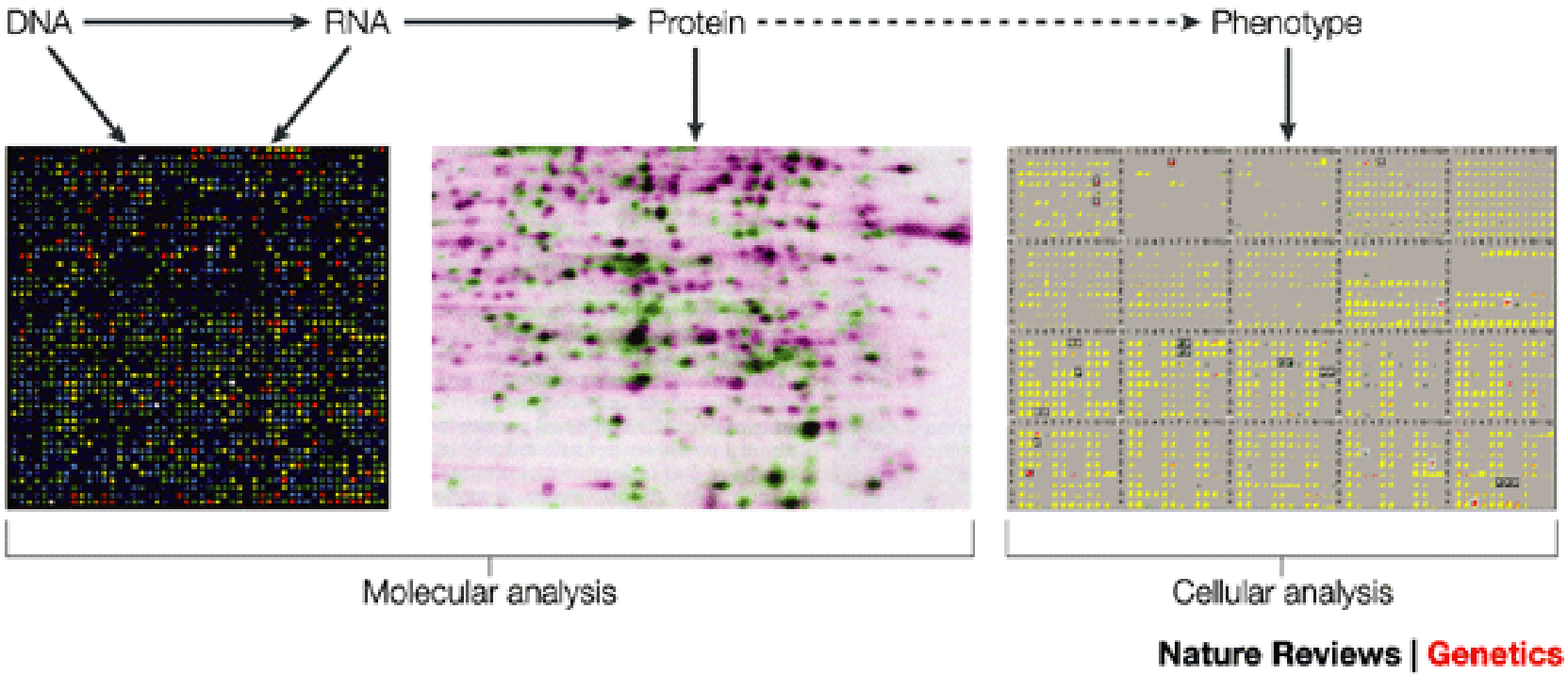
LOD map for PDI: *cis*-regulation (Lan et al. 2003)



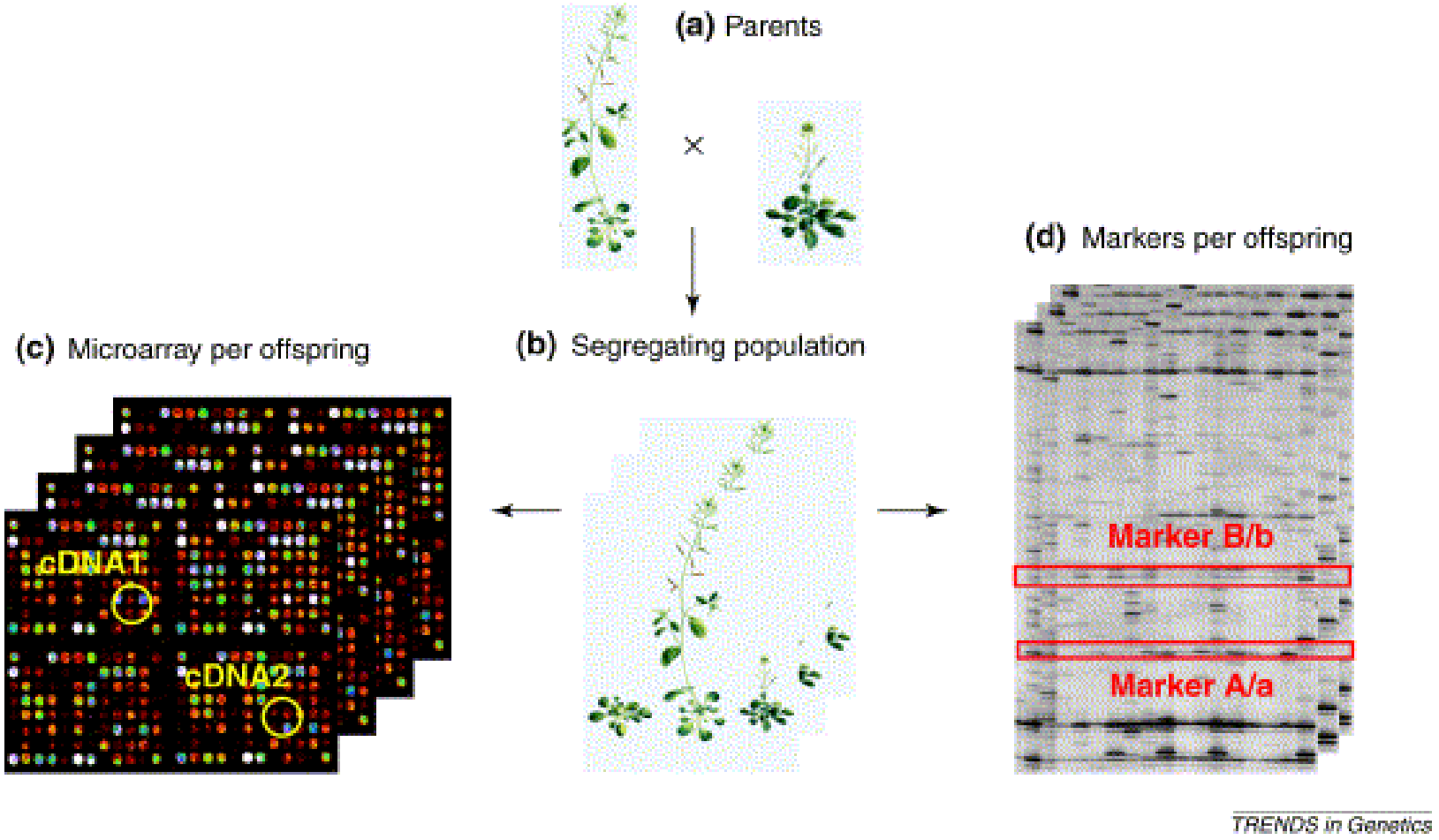
mapping microarray data

- single gene expression as trait (single QTL)
  - Dumas et al. (2000 *J Hypertens*)
- overview, wish lists
  - Jansen, Nap (2001 *Trends Gen*); Cheung, Spielman (2002); Doerge (2002 *Nat Rev Gen*); Bochner (2003 *Nat Rev Gen*)
- microarray scan via 1 QTL interval mapping
  - Brem et al. (2002 *Science*); Schadt et al. (2003 *Nature*); Yvert et al. (2003 *Nat Gen*)
  - found putative *cis*- and *trans*- acting genes
- multivariate and multiple QTL approach
  - Lan et al. (2003 *Genetics*)

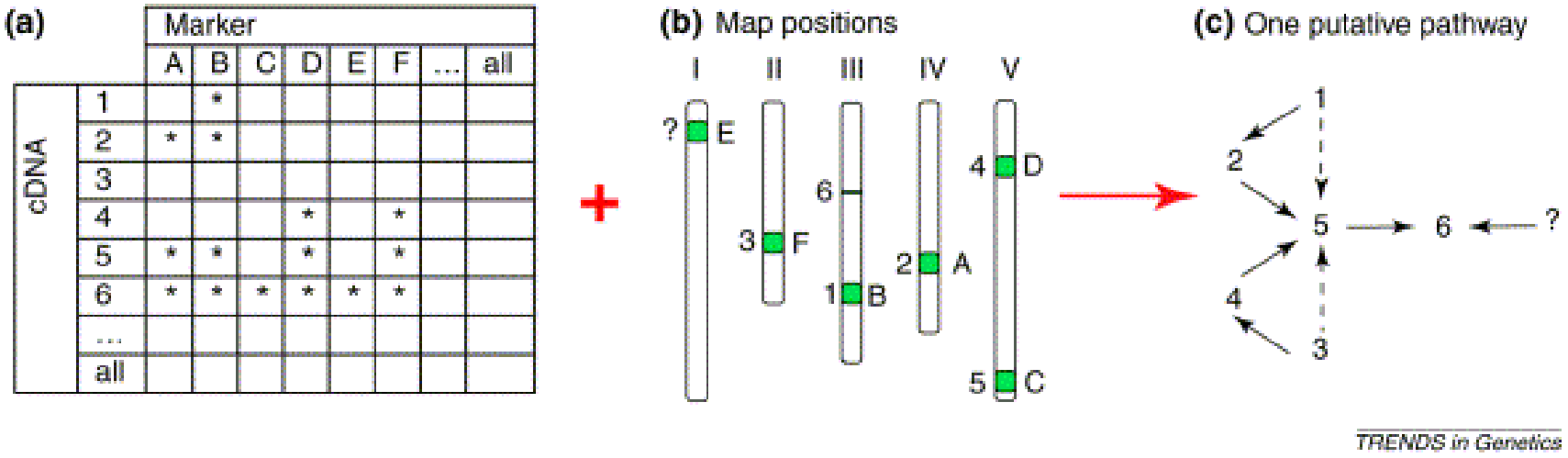
central dogma via microarrays (Bochner 2003)



idea of mapping microarrays (Jansen Nap 2001)



goal: unravel biochemical pathways (Jansen Nap 2001)

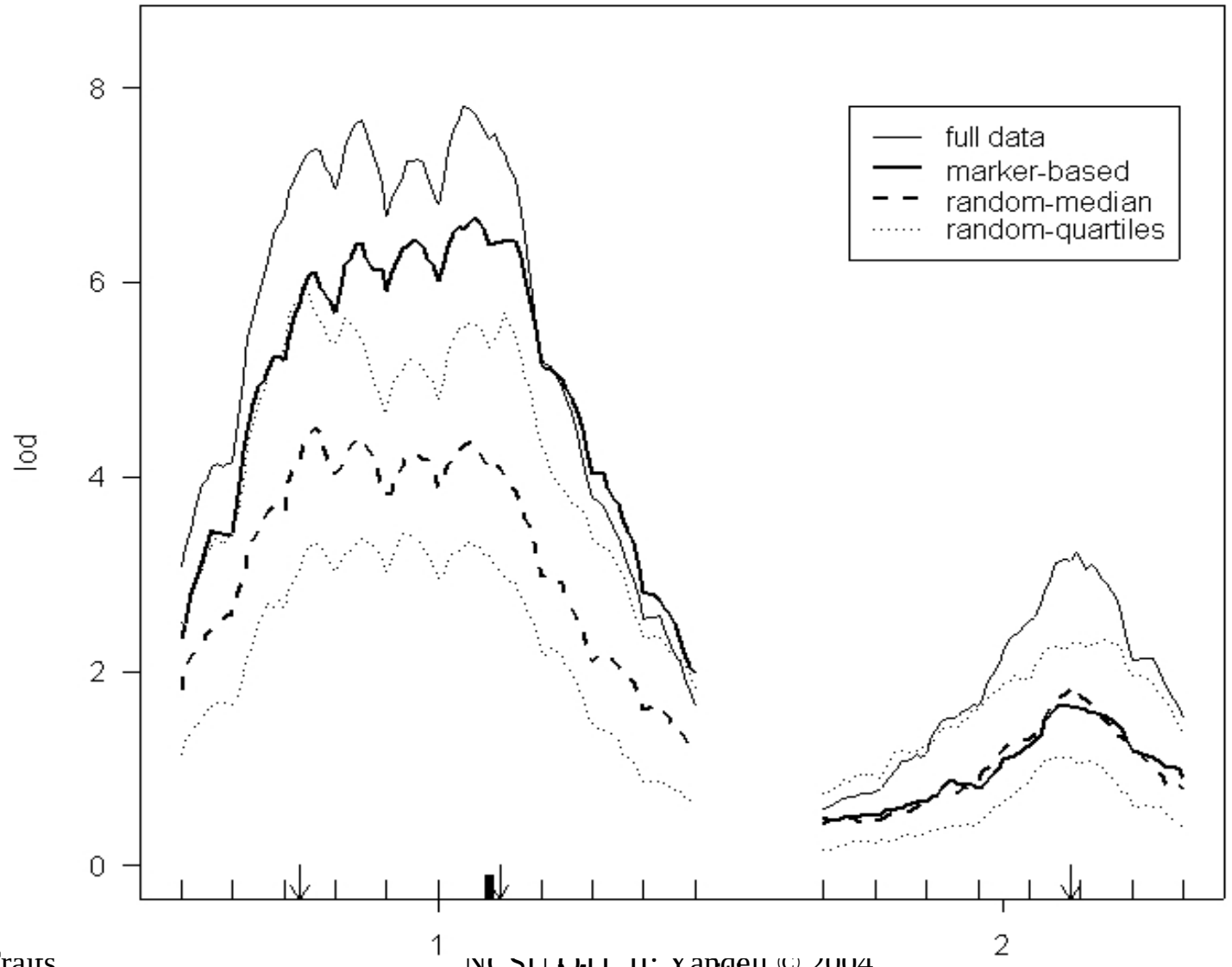



GeneChip
AFFYMETRIX

2. design issues for expensive phenotypes (thanks to CF “Amy” Jin)

- microarray analysis ~ $1000 per mouse
  - can only afford to assay 60 of 108 in panel
  - wish to not lose much power to detect QTL
- selective phenotyping
  - genotype all individuals in panel
  - select subset for phenotyping
  - previous studies can provide guide

simulated LOD profiles with 3 QTL

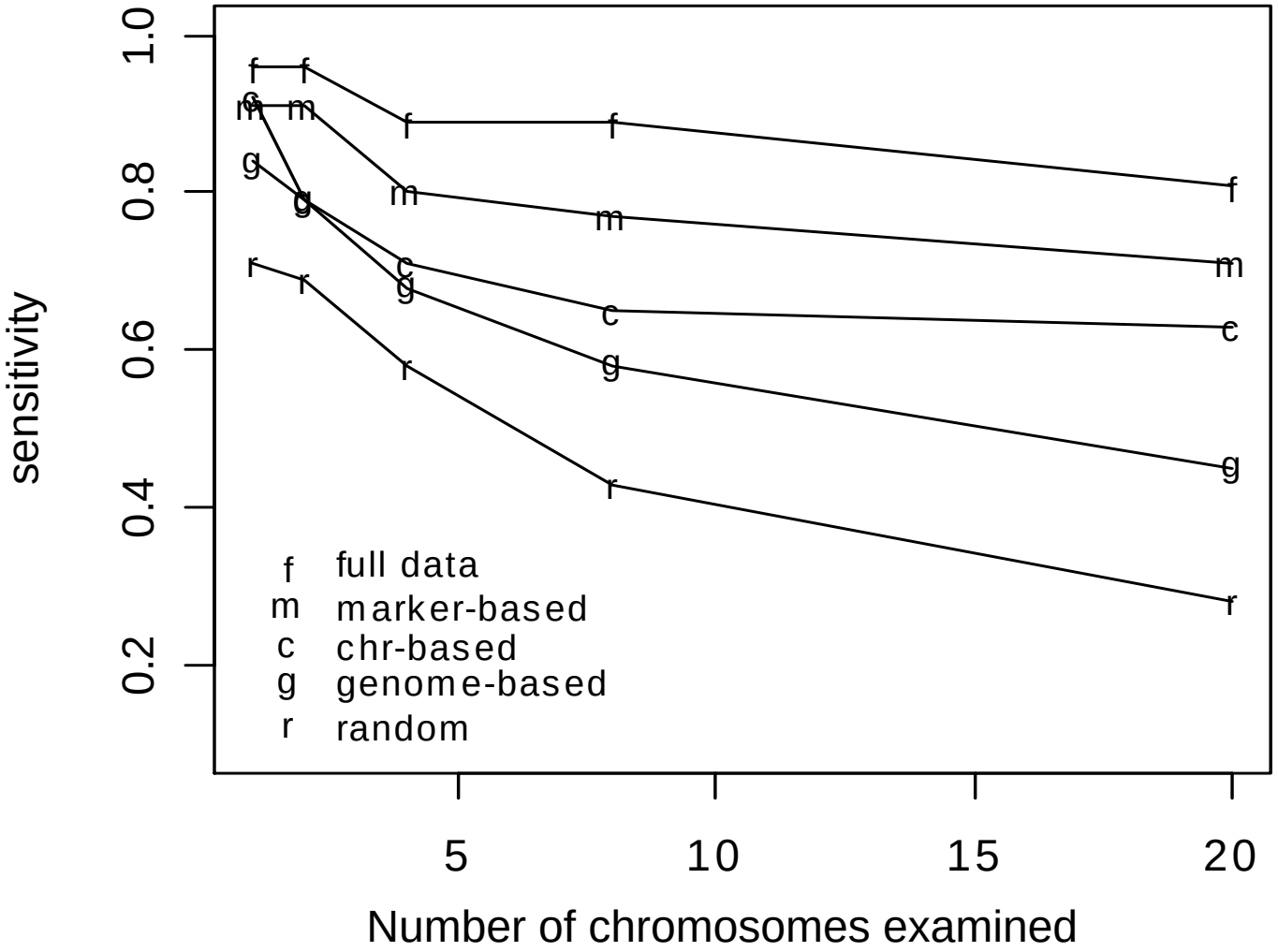


selective phenotyping

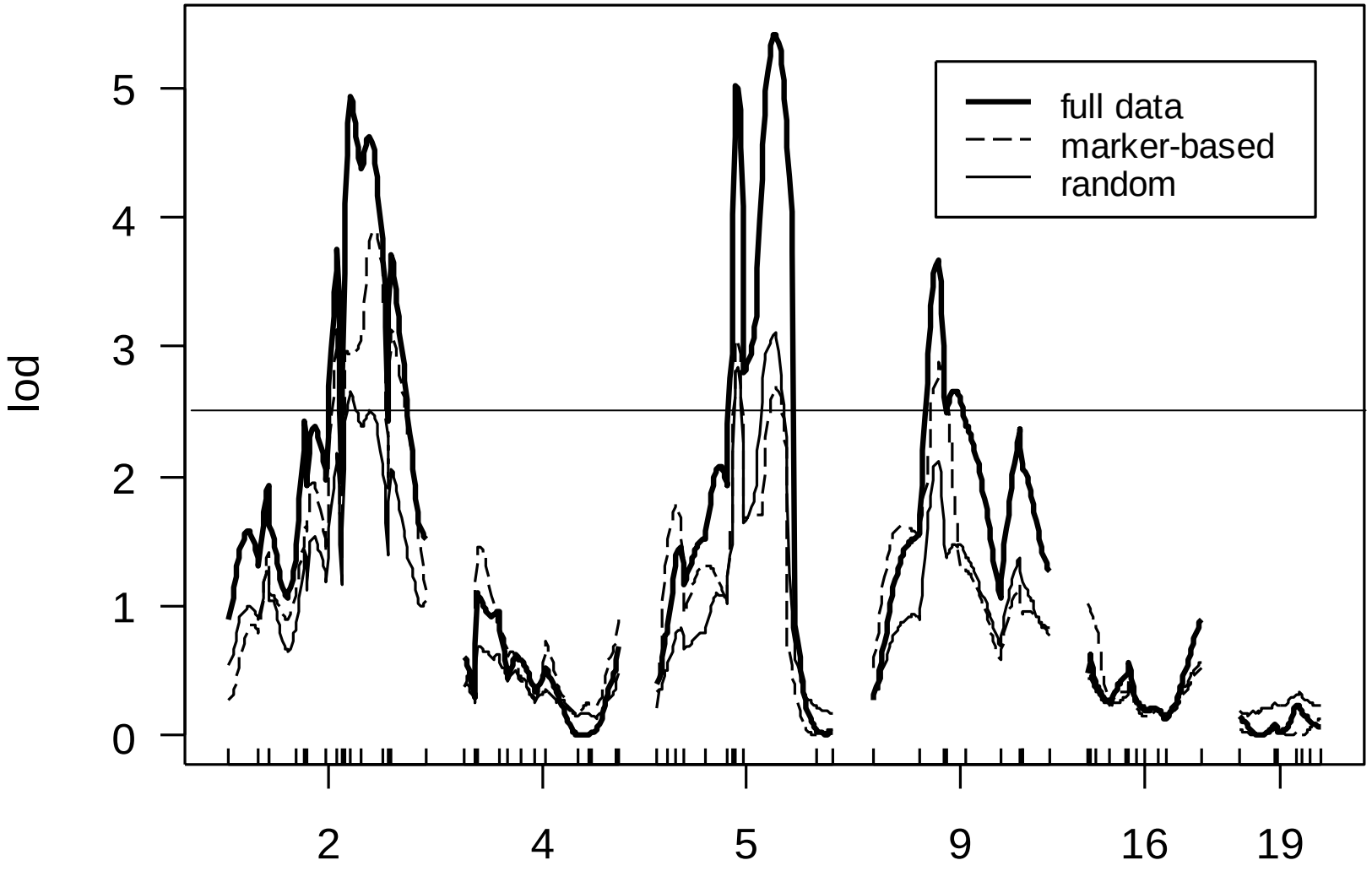
- genotype all individuals in panel
  - whole genome or selected genomic regions?
  - maintain high power in selected regions
  - sensitivity similar to random sample in other regions
- select subset for phenotyping
  - select individuals with large genetic distance
  - use experimental design concepts (Jin et al. 2004)
- previous studies: key regions of chr 2,4,5,9,16,19
  - QTL for important physiological traits

comparison of different selection methods

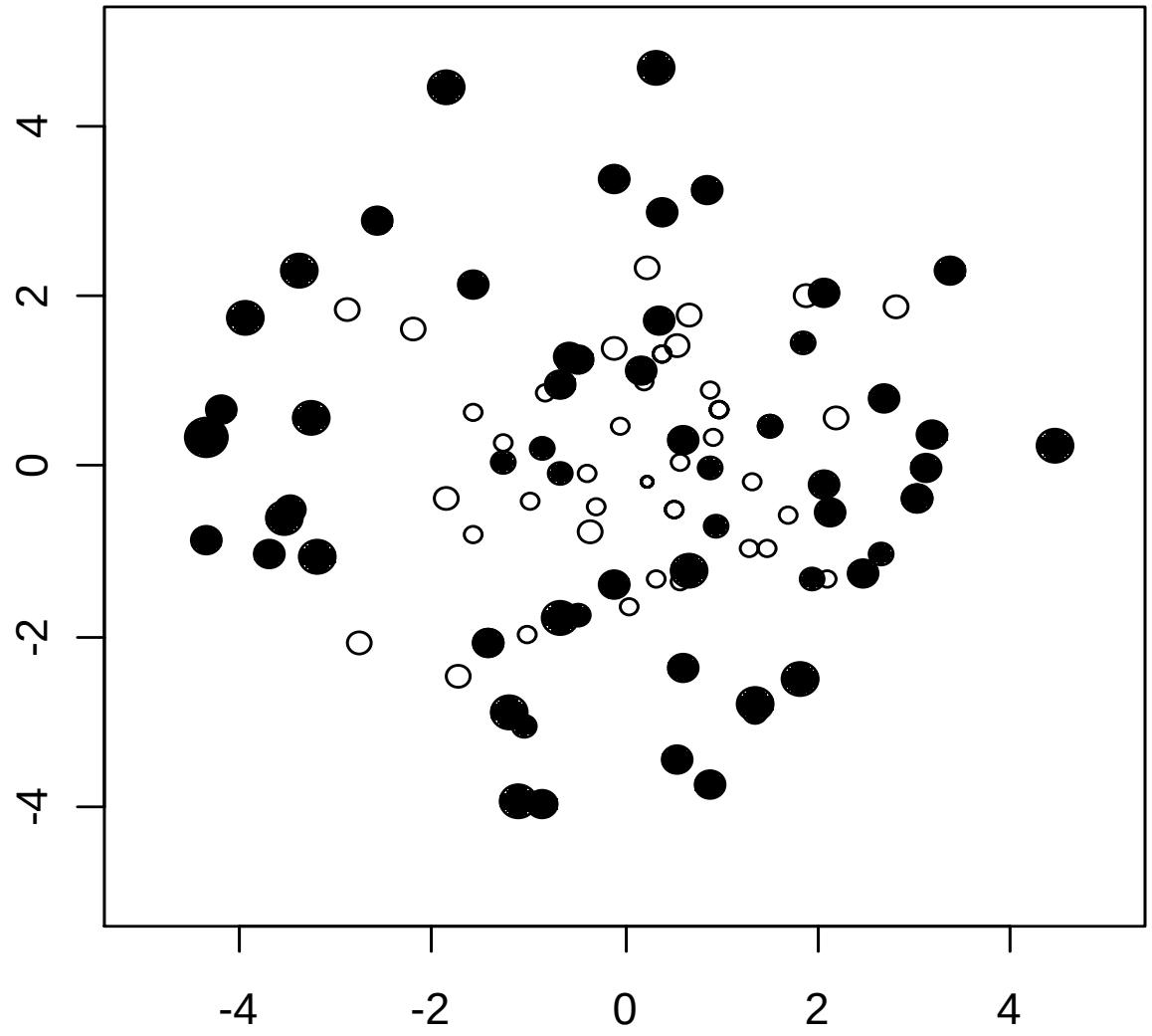
sensitivity = pr(detect QTL | QTL is real)



LOD profile of SCD trait



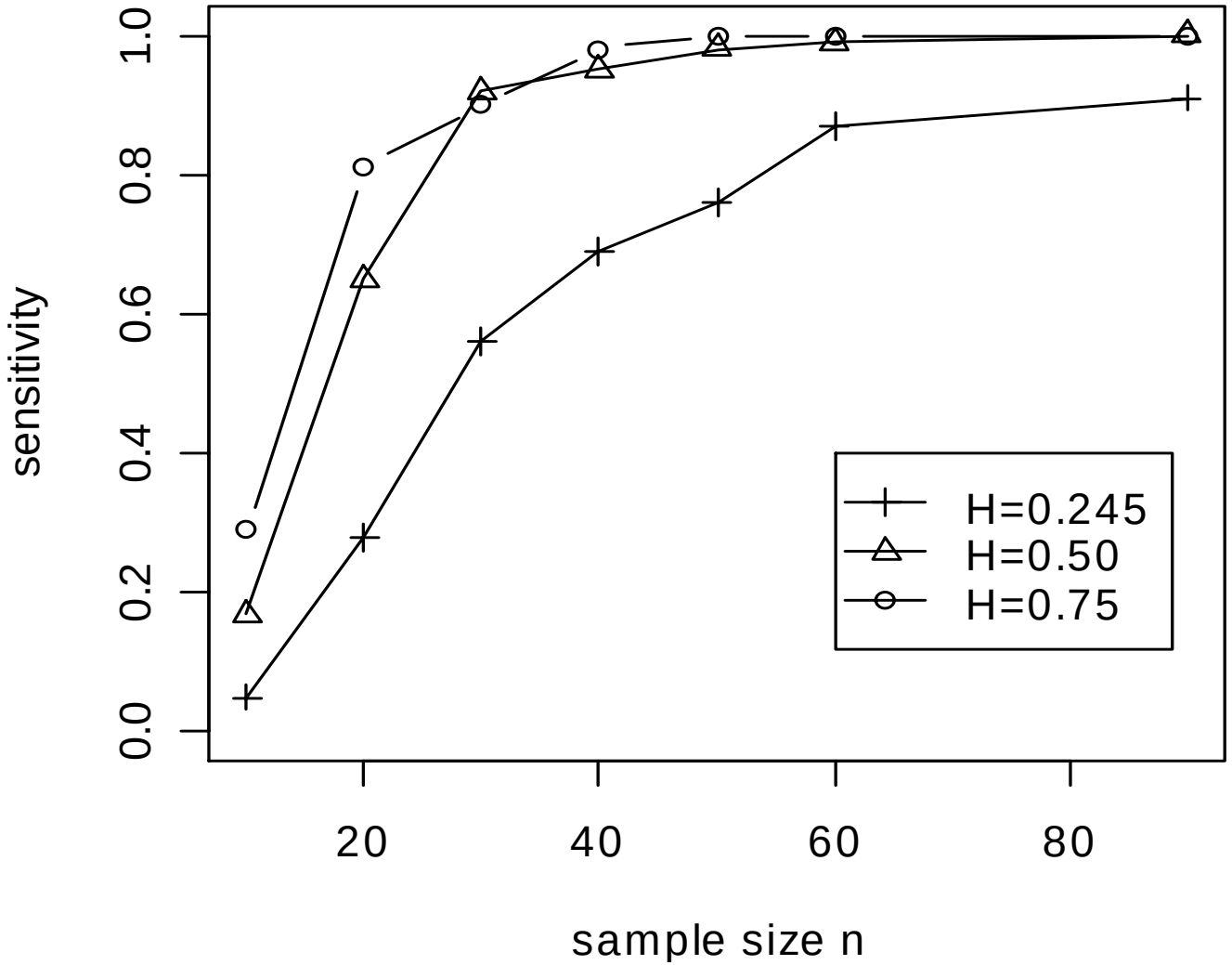
multidimensional scaling of mice selection (close points have similar genotypes)



is this relevant to large QTL studies?

- why not phenotype entire mapping panel?
  - selectively phenotype subset of 50-67%
  - may capture most effects
  - with little loss of power
- two-stage selective phenotyping?
  - genotype & phenotype subset of 100-300
    - could selectively phenotype using whole genome
  - QTL map to identify key genomic regions
  - selectively phenotype subset using key regions

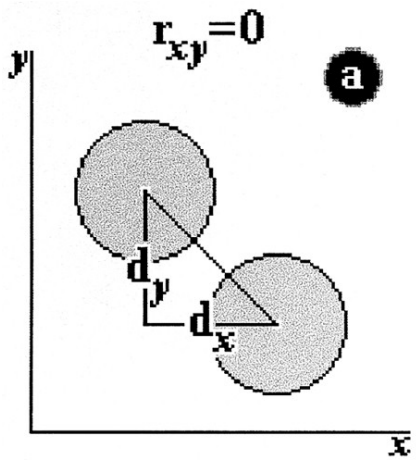
sensitivity = pr(detect QTL | QTL is real)
depends on heritability and proportion sampled (of N=100)



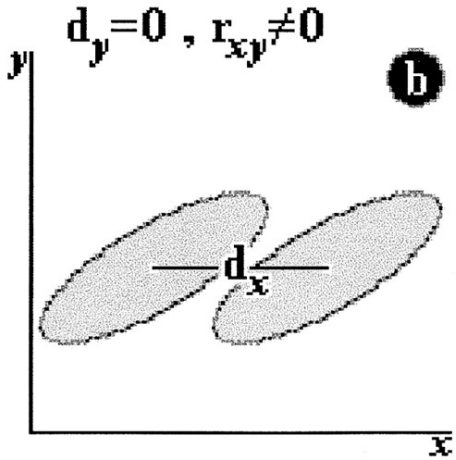
3. why are traits correlated?

- environmental correlation
  - non-genetic, controllable by design
  - historical correlation (learned behavior)
  - physiological correlation (same body)
- genetic correlation
  - pleiotropy
    - one gene, many functions
    - common biochemical pathway, splicing variants
  - close linkage
    - two tightly linked genes
    - genotypes *Q* are collinear

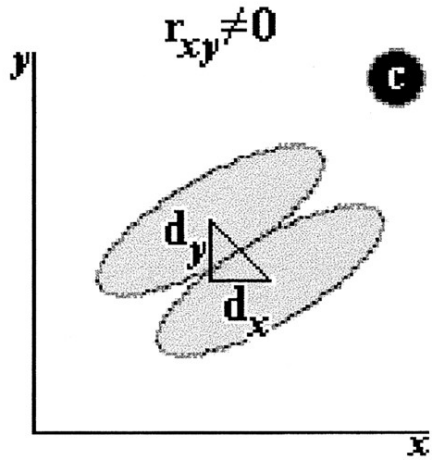
interplay of pleiotropy & correlation



pleiotropy only



correlation only



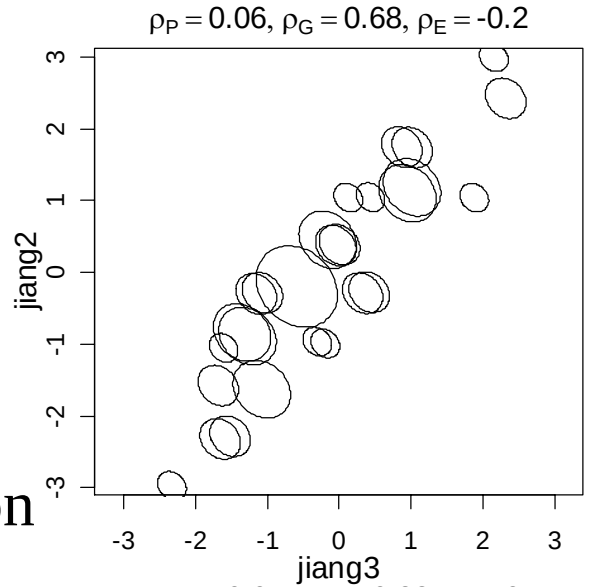
both

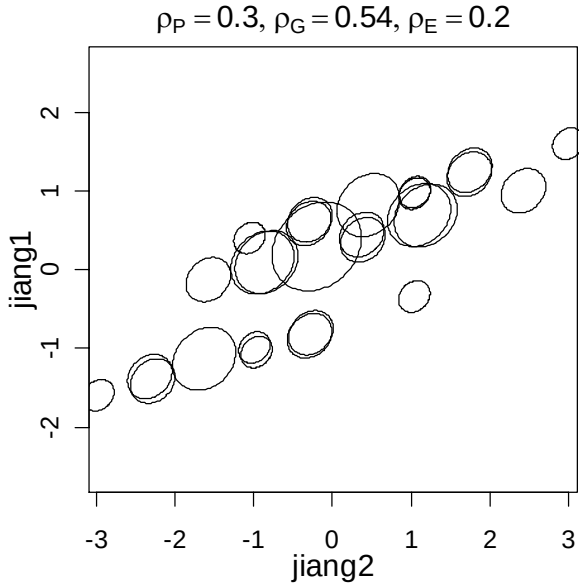
Korol et al. (2001)

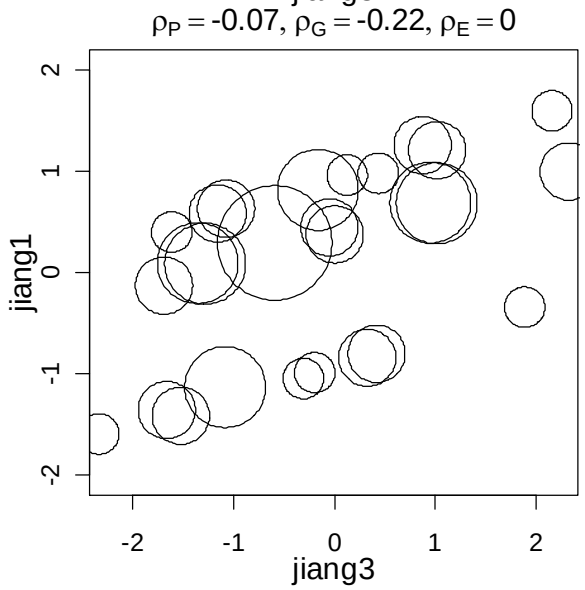
3 correlated traits (Jiang Zeng 1995)

ellipses centered on genotypic value
width for nominal frequency
main axis angle environmental correlation
3 QTL, F2
27 genotypes

note signs of genetic and environmental correlation

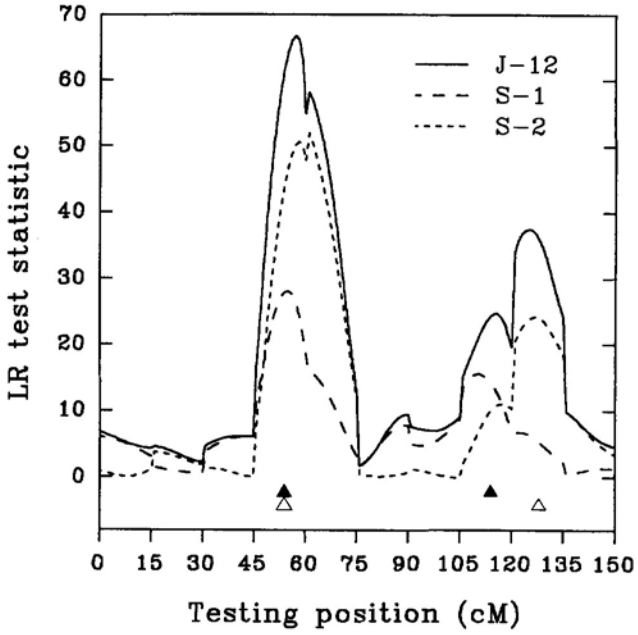






pleiotropy or close linkage?

2 traits, 2 qtl/trait
pleiotropy @ 54cM
linkage @ 114,128cM
Jiang Zeng (1995)



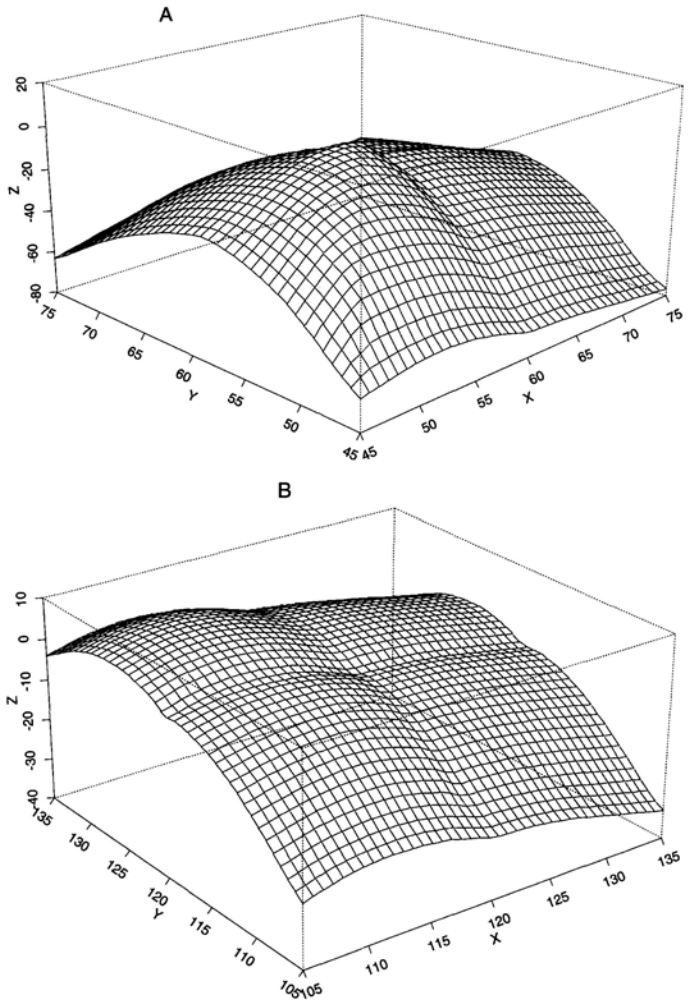
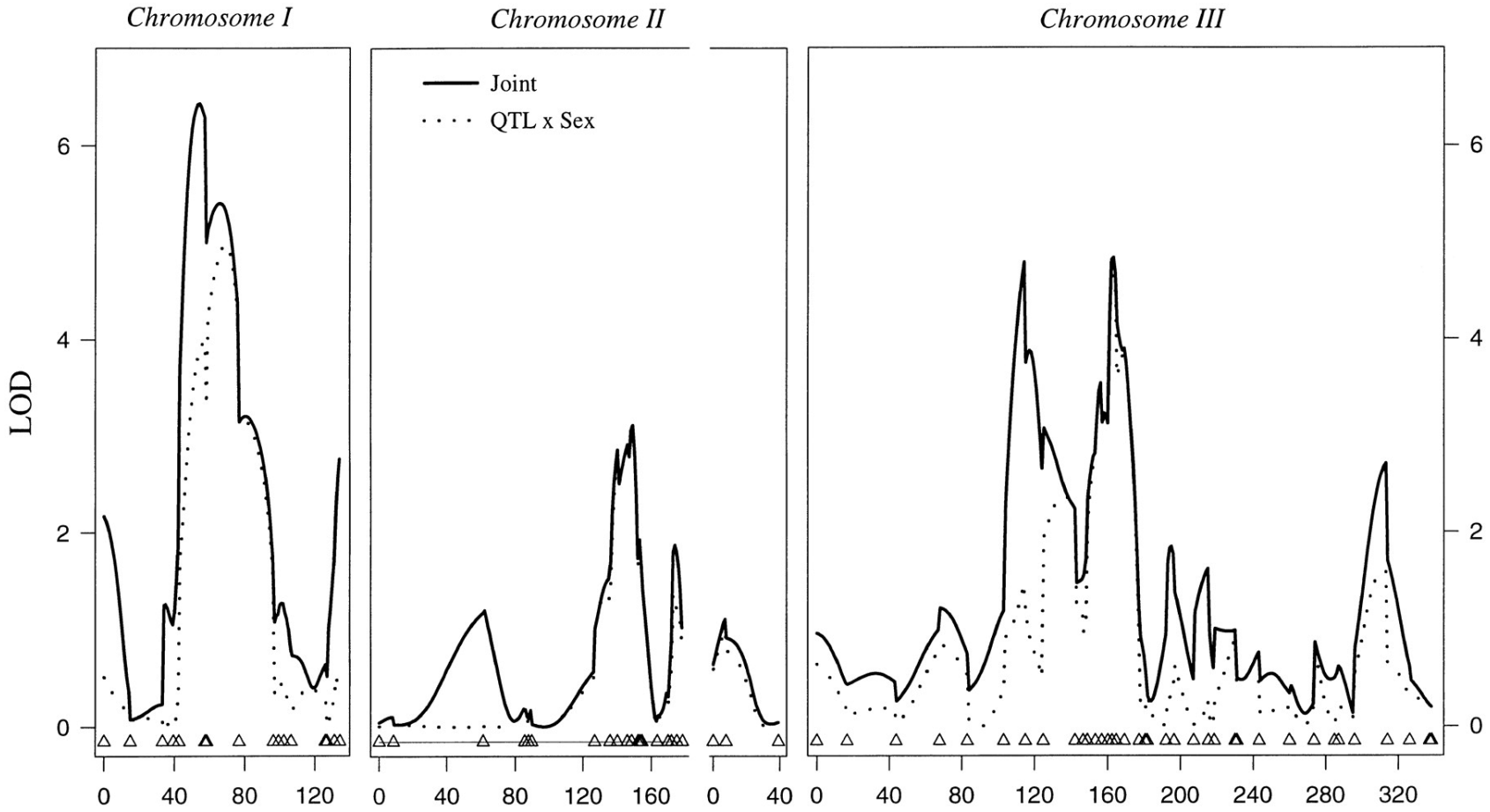


FIGURE 2.—Two-dimensional log-likelihood surfaces (expressed as deviations from the maximum of the log-likelihoods on the diagonal) for the test of pleiotropy *vs.* close linkage are presented for two regions: the region between 45 and 75 cM of Figure 1 (A) and the region between 105 and 135 cM (B). *X* is the testing position for a QTL affecting trait 1 and *Y* is the testing position for a QTL affecting trait 2. On the diagonal of *X-Y* plane, two QTL are located in the same position and statistically are treated as one pleiotropic QTL. *Z* is the likelihood ratio test statistic scaled to zero at the maximum point of the diagonal.

QTL x sex interaction (Vieira et al. 2000)

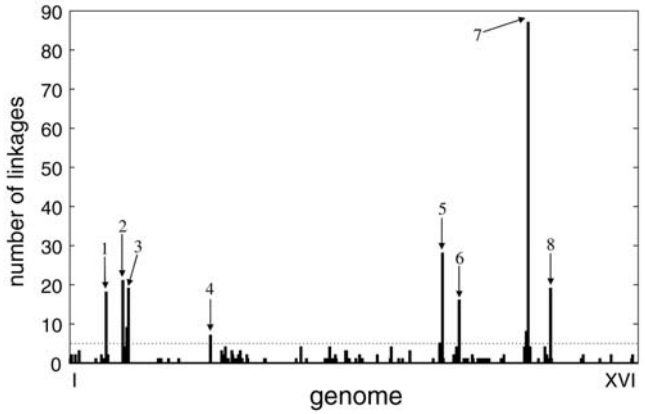


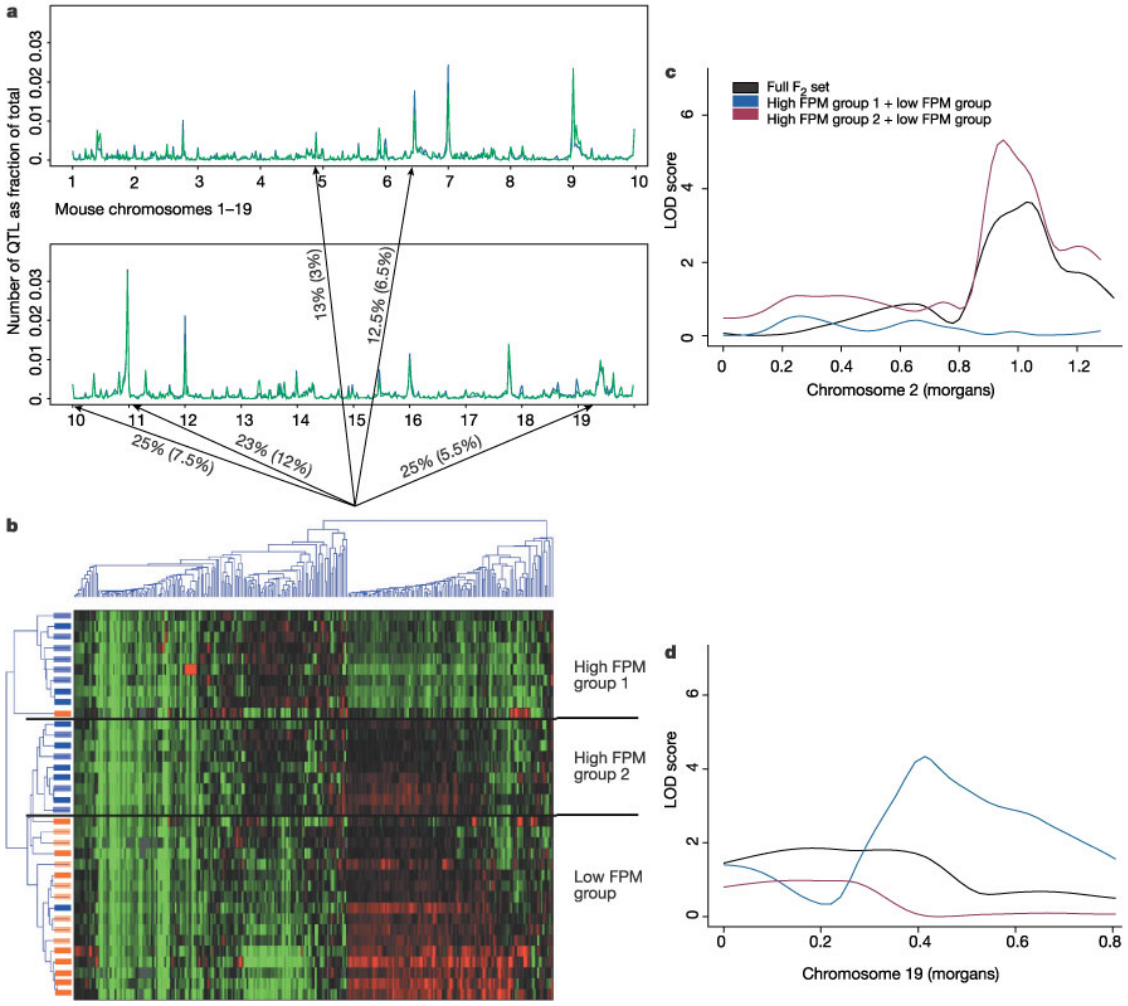
4. high throughput dilemma

- want to focus on gene expression network
  - ideally capture pathway in a few dimensions
  - allow for complicated genetic architecture
- may have multiple controlling loci
  - could affect many genes in coordinated fashion
  - could show evidence of epistasis
  - quick assessment via interval mapping may be misleading
- try mapping principle components as super-traits
  - capture key multivariate features of multiple traits
  - elicit biochemical pathways (Henderson et al. Hoeschele 2001; Ong Page 2002)

coordinated expression in mouse genome (Schadt et al. 2003)

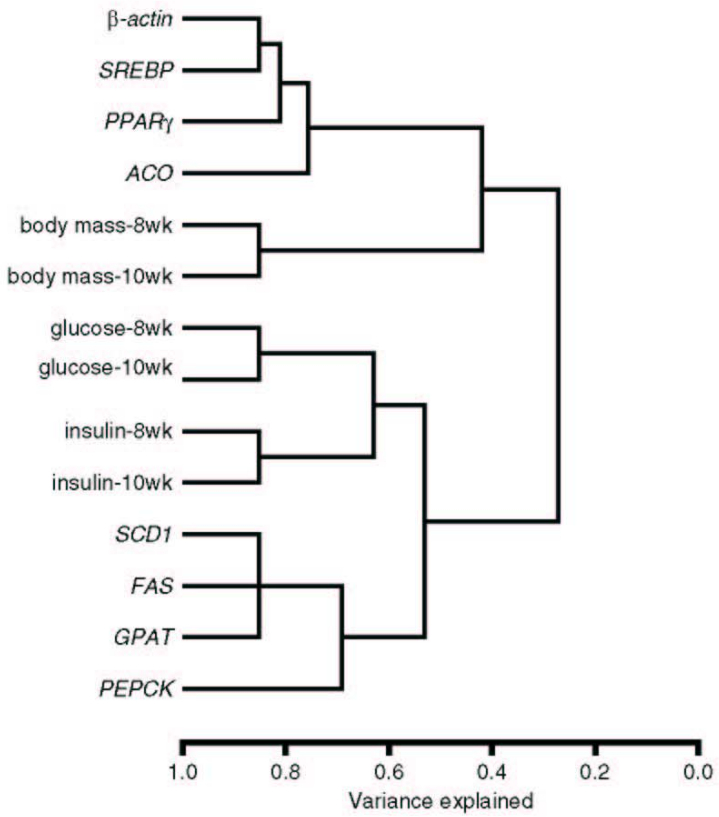
expression pleiotropy in yeast genome (Brem et al. 2002)





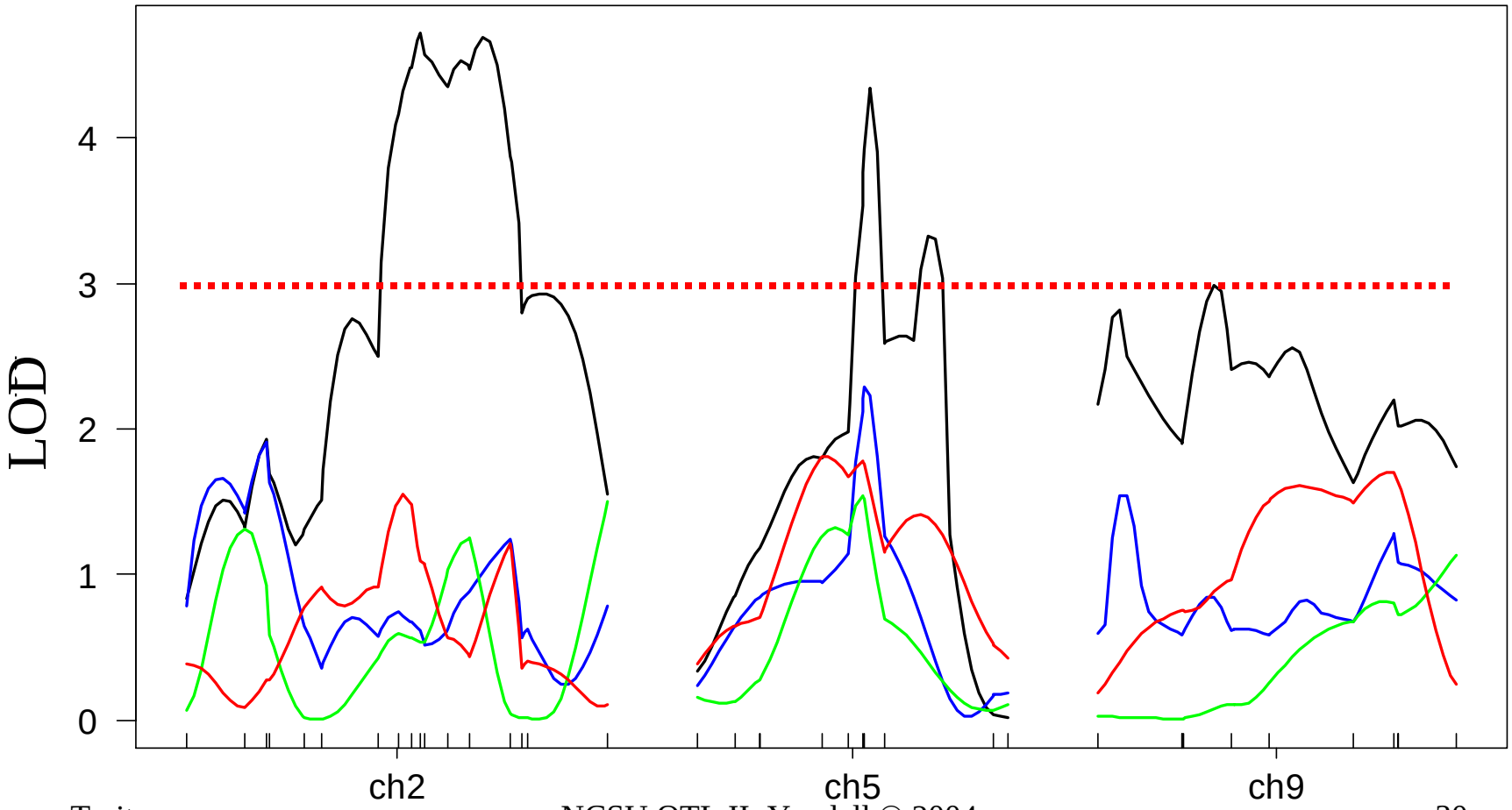
high throughput:
which genes are the key players?

Lan et al., mapping mRNA, Figure 2



- clustering of expression
  seed by insulin, glucose
- advantage:
  subset relevant to trait
- disadvantage:
  still many genes to study

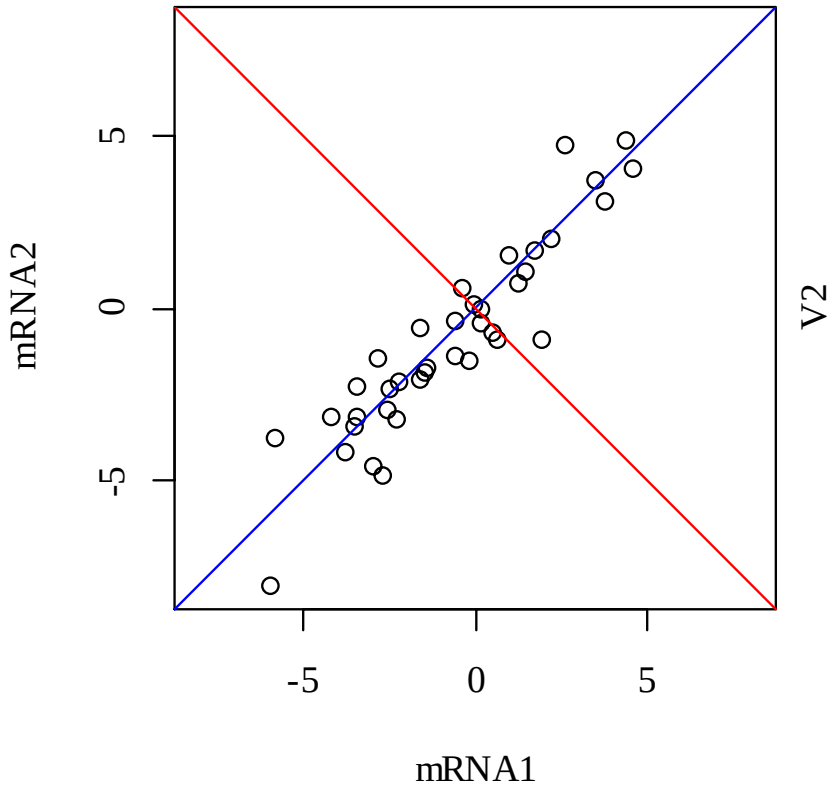
SCD1, FAS,GPAT, PEPCK: *trans*-regulation by multiple QTL?

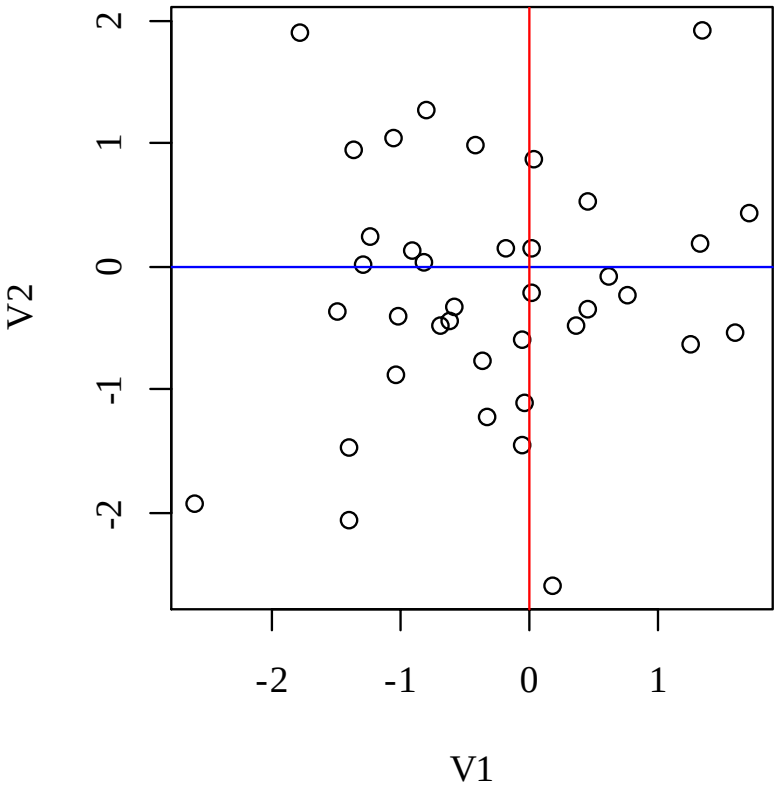


from gene expression to super-genes

- PC or SVD decomposition of multiple traits
  - Y = t traits × n individuals
  - decompose as $Y = UDW^{T}$
    - U,W = ortho-normal transforms (eigen-vectors)
    - D = diagonal matrix with singular values
- transform problem to principal components
  - W_1 and W_2 uncorrelated "super-traits"
- interval map each PC separately
  - $W_1 = \mu^*_1 + G^*_{1Q} + e^*_1$
- may only need to map a few PCs

PC simply rotates & rescales to find major axes of variation



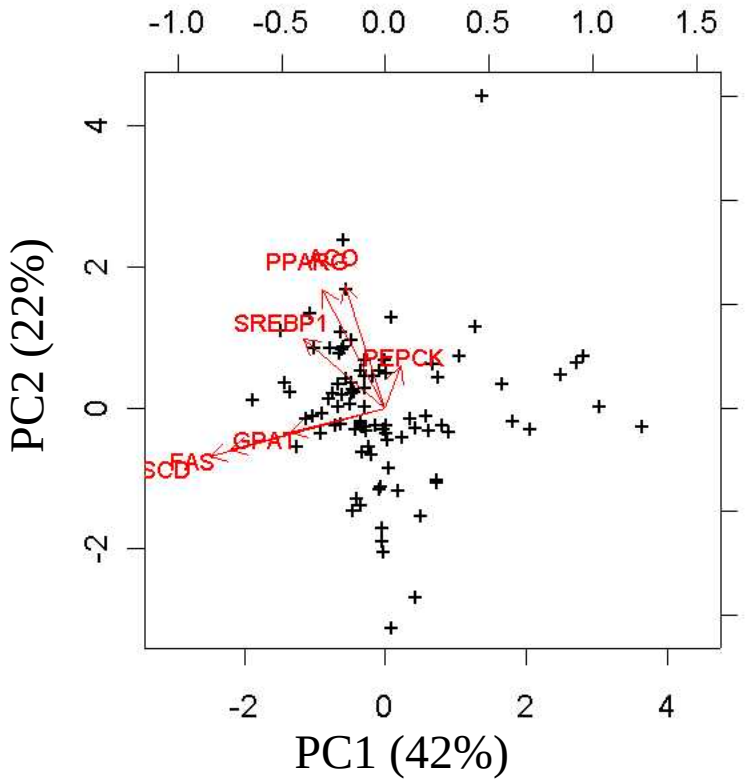


QTL via Principal Components

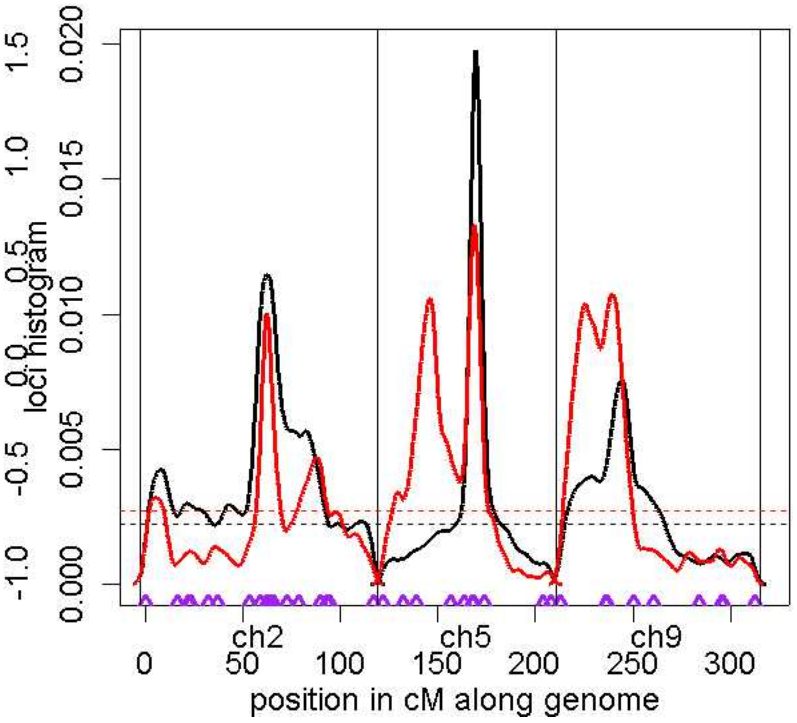
- *Drosophila* gonad shape
  - Liu et al. (1996); Zeng et al. (2000)
- other refs of interest
  - Weller et al. (1996); Mangin et al. (1998); Olson et al. (1999); Mahler et al. (2002)
- problems
  - PC may have no relation to genetics!
  - residuals from QTL correlated across PCs
  - PC is descriptive summary, not interpretive

multivariate screen for gene expressing mapping

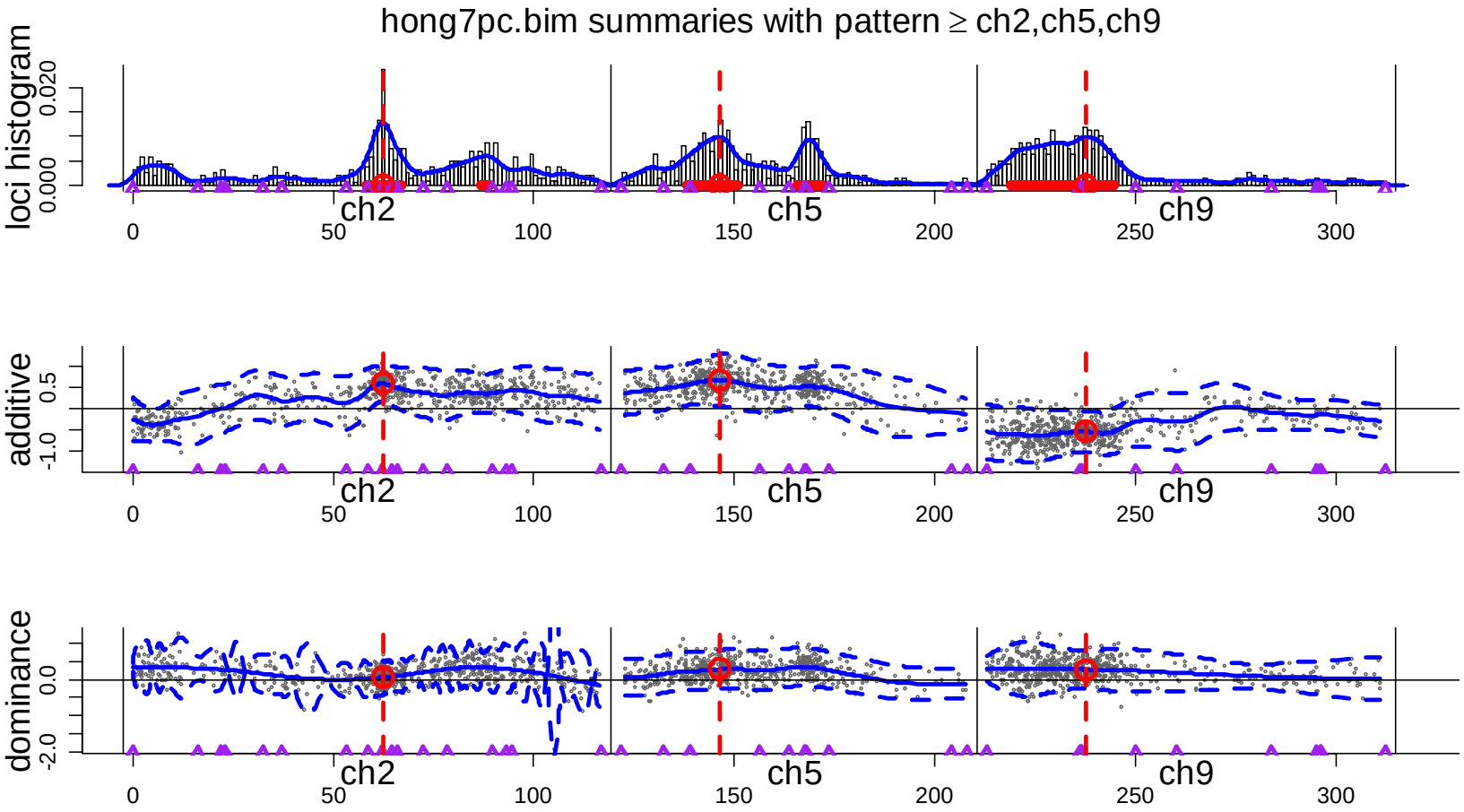
principal components



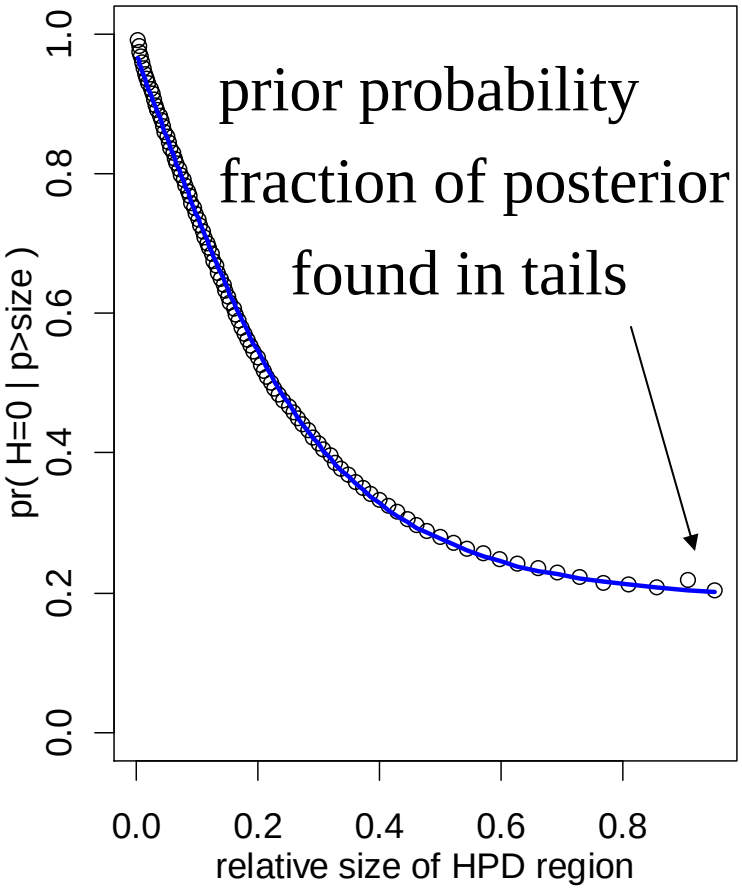
PC1(red) and SCD(black)

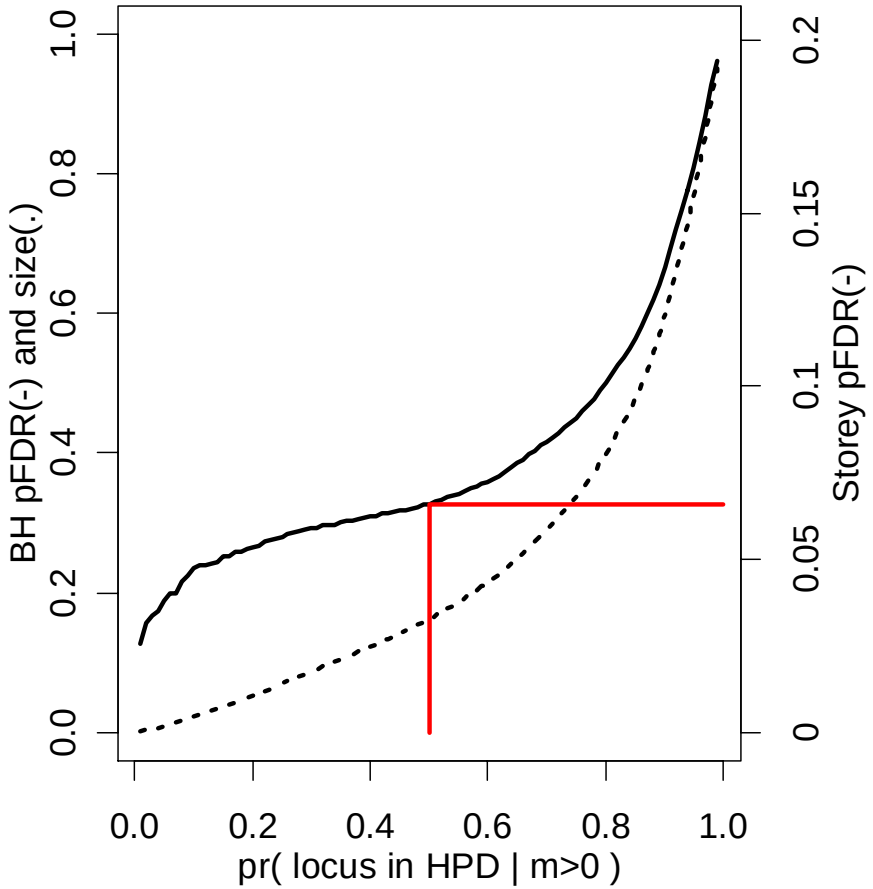


mapping first diabetes PC as a trait



pFDR for PC1 analysis

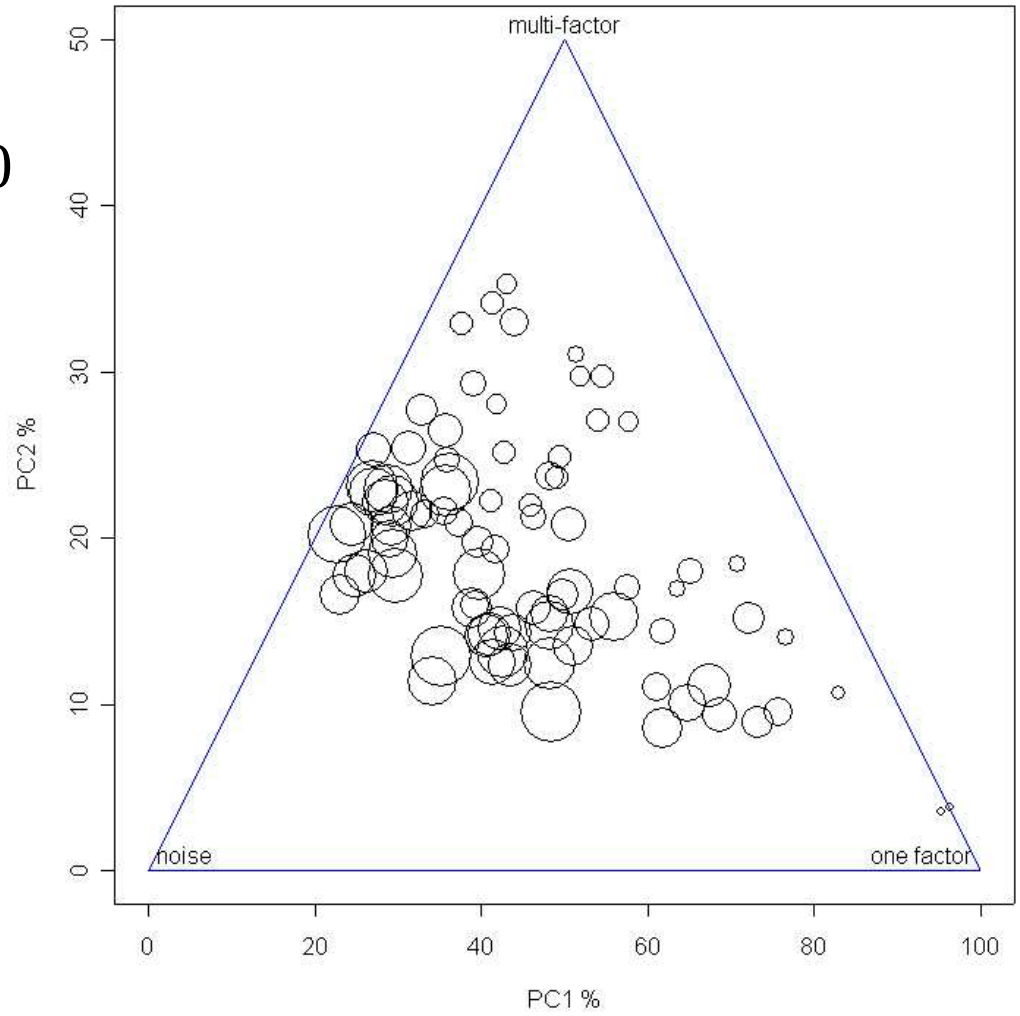




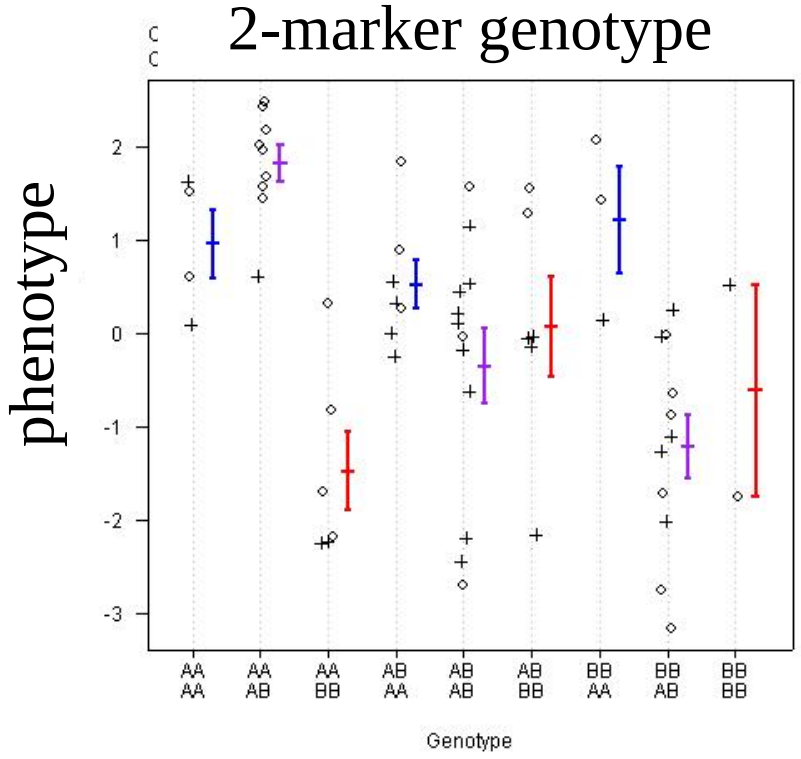
PC across microarray functional groups

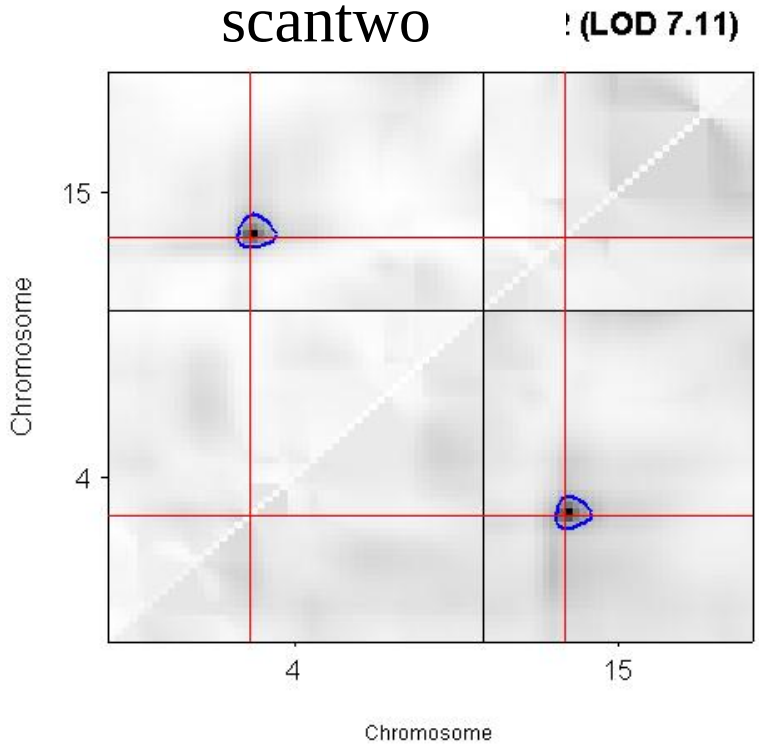
1500+ mRNA of 30,000
85 functional groups
60 mice
2-35 mRNA / group
which are interesting?

examine PC1, PC2
size = # unique mRNA



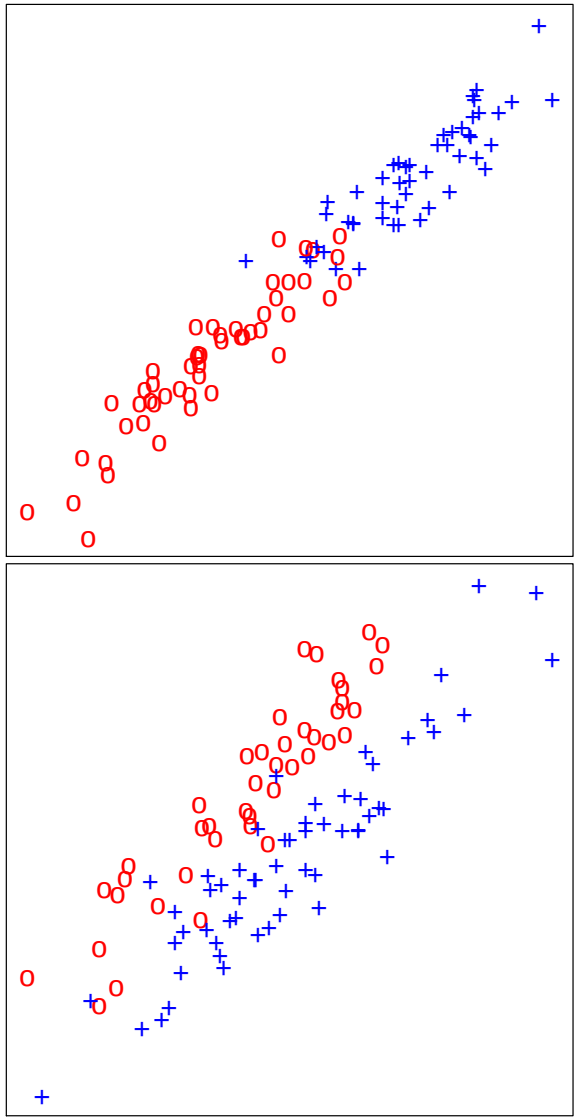
PC-guided search of mRNA
(red lines at main QTL for PC1)





improvements on PC?

- what is our goal?
  - reduce dimensionality
  - focus on QTL
- PC reduces dimensionality
  - but may not relate to genetics
- canonical discriminant analysis
  - rotate to improve discrimination
  - redo at each putative QTL
  - Gilbert and le Roy (2003,2004)



how to map multiple traits?

- WinQTL/QTL Cartographer: IM & CIM
  - Jiang Zeng (1995); statgen.ncsu.edu/qtlcart
- MultiQTL: 1-2 QTL with PC on residuals
  - Korol et al. (2001); www.multiqtl.com
- 1-2 QTL with DA across traits
  - Gilbert and le Roy (2003, 2004)
- QTL Express: Haley-Knott regression
  - Knott Haley (2000); qtl.cap.ed.ac.uk
- SOLAR: outbred pedigrees
  - Almasy Blangero (1997); Williams et al. (1999)