Multiple Traits & Microarrays

- why map multiple traits together?
  - central dogma via microarrays
  - diabetes case study
- why are traits correlated?
  - close linkage or pleiotropy?
- how to handle high throughput?
  - dimension reduction: multivariate stats
  - principal components on phenotypes

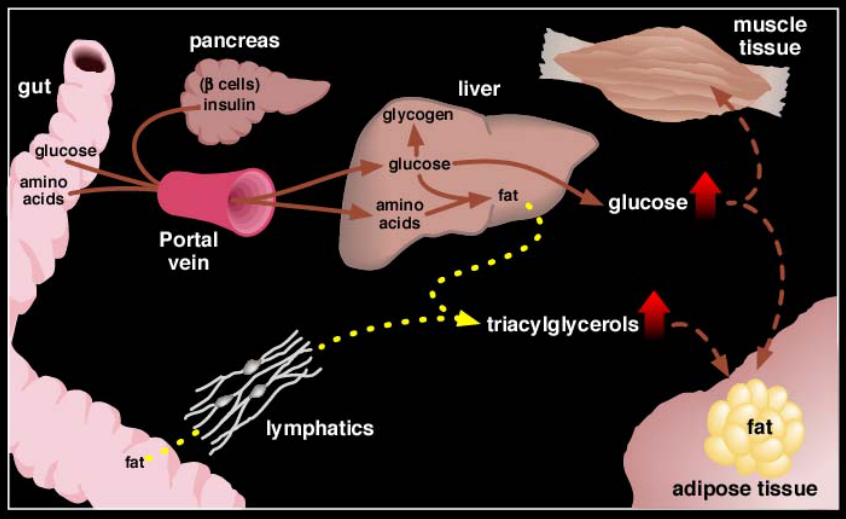
1 why study multiple traits together?

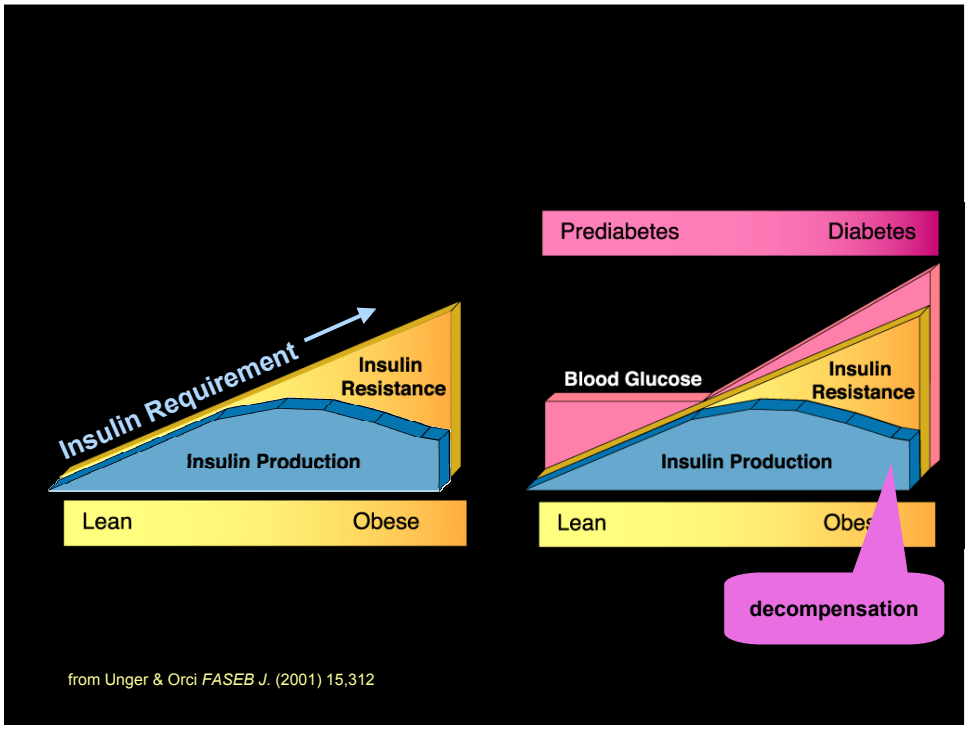
- avoid reductionist approach to biology
  - address physiological/biochemical mechanisms
  - Schmalhausen (1942); Falconer (1952)
- separate close linkage from pleiotropy
  - 1 locus or 2 linked loci?
- identify epistatic interaction or canalization
  - influence of genetic background
- establish QTL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

how to map multiple traits?

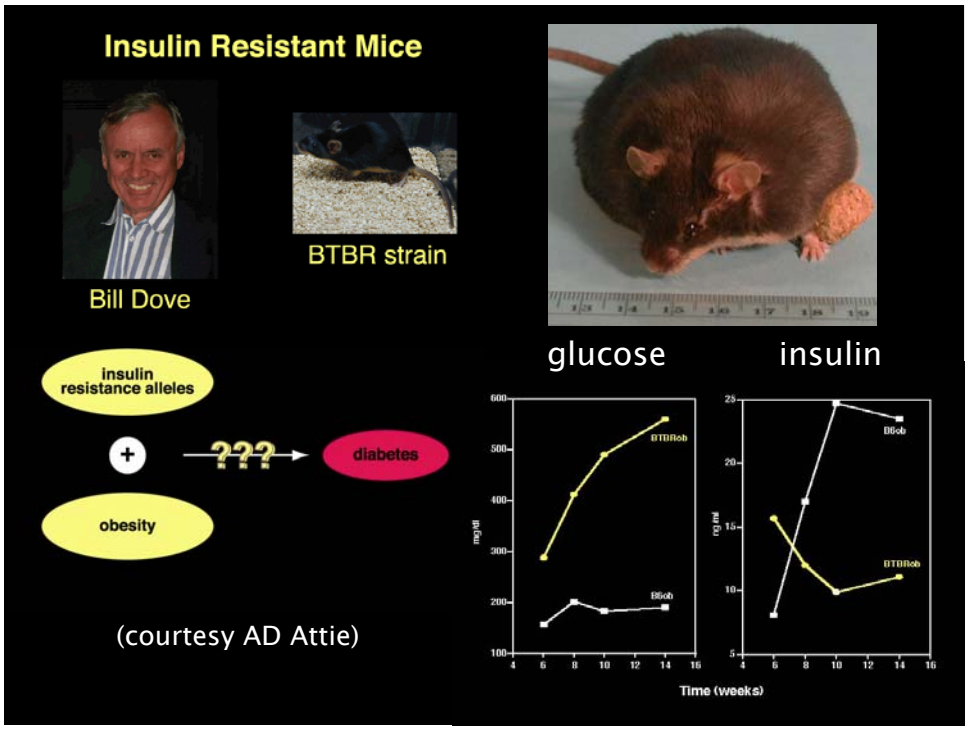
- WinQTL/QTL Cartographer: IM & CIM
  - Jiang Zeng (1995); Vieira et al. (2000)
  - statgen.ncsu.edu/qtlcart
- MultiQTL: 1-2 QTL with PC on residuals
  - Korol et al. (2001)
  - www.multiqtl.com
- QTL Express: Haley-Knott regression
  - Knott Haley (2000)
  - qtl.cap.ed.ac.uk
- SOLAR: outbred pedigrees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

Type 2 Diabetes Mellitus





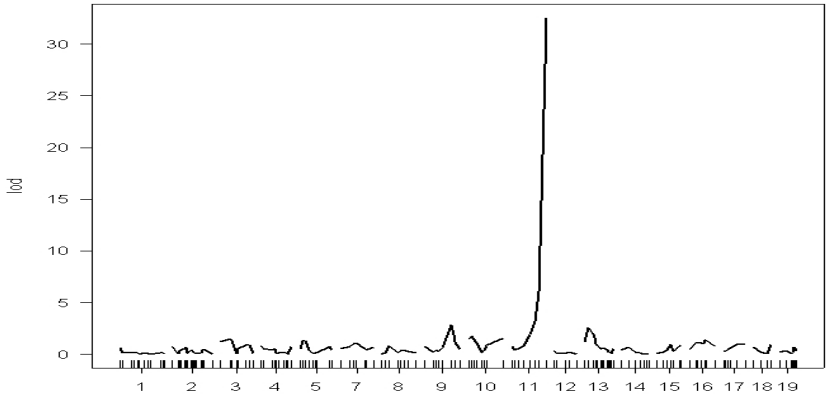
from Unger & Orci *FASEB J.* (2001) 15,312



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
    - key tissues: adipose, liver, muscle, β-cells
    - novel discoveries of differential expression (Nadler et al. 2000 PNAS; Lan et al. 2002 in review; Ntambi et al. 2002 PNAS)
  - RT-PCR on 108 F2 mice liver tissues
    - 15 genes, selected as important in diabetes pathways
    - SCD1, PEPCK, ACO, FAS, GPAT, PPARgamma, PPARalpha, G6Pase, PDI,…

LOD map for PDI: *cis*-regulation



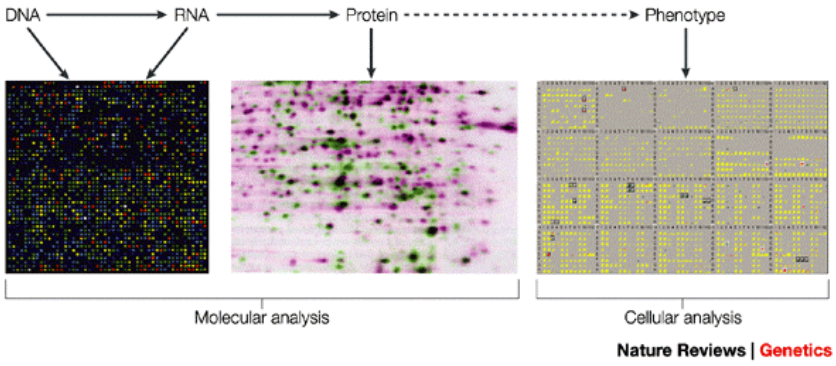
why map gene expression as a quantitative trait?

- *cis*- or *trans*-action?
  - does gene control its own expression?
  - evidence for both modes (Brem et al. 2002 Science)
- mechanics of gene expression mapping
  - measure gene expression in intercross (F2) population
  - map expression as quantitative trait (QTL technology)
  - adjust for multiple testing via false discovery rate
- research groups working on expression QTLs
  - review by Cheung and Spielman (2002 *Nat Gen Suppl*)
  - Kruglyak (Brem et al. 2002 *Science*)
  - Doerge et al. (Purdue); Jansen et al. (Waginingen)
  - Williams et al. (U KY); Lusis et al. (UCLA)
  - Dumas et al. (2000 *J Hypertension*)

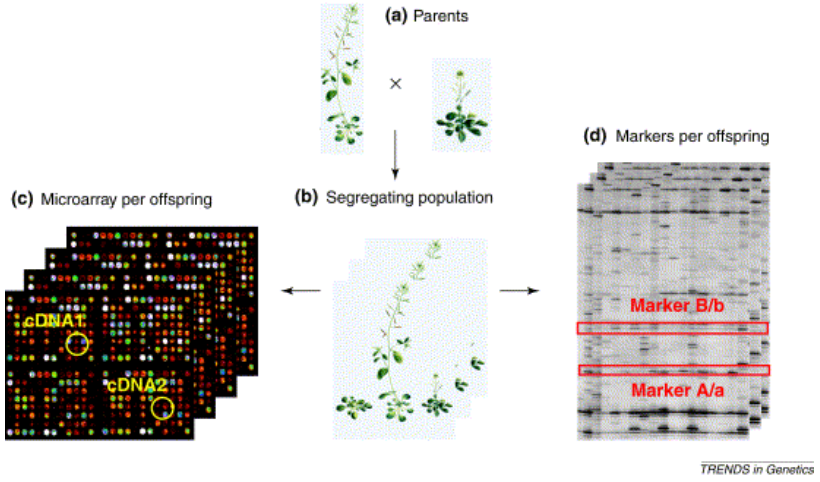
mapping microarray data

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

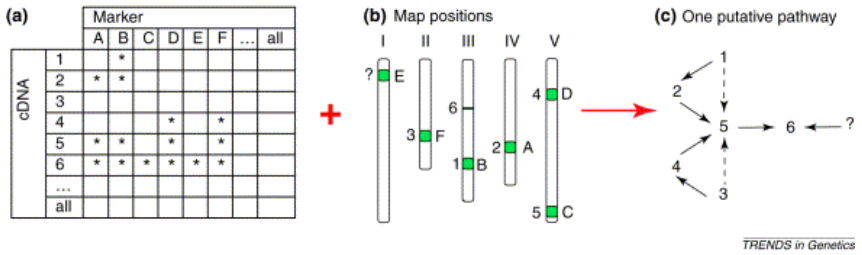
central dogma via microarrays (Bochner 2003)



idea of mapping microarrays (Jansen Nap 2001)



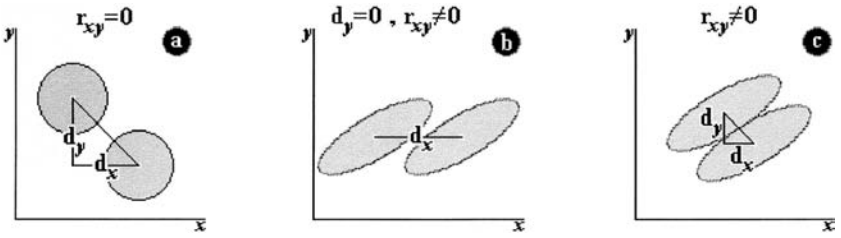
goal: unravel biochemical pathways (Jansen Nap 2001)



2 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



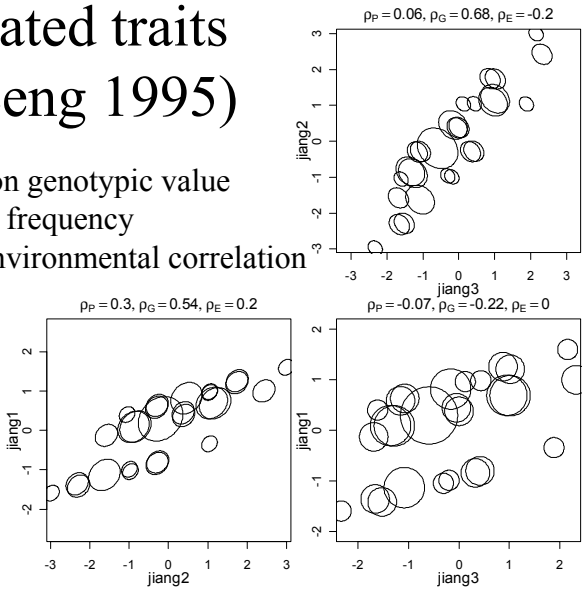
pleiotropy only | correlation only | both

Korol et al. (2001)

3 correlated traits (Jiang Zeng 1995)

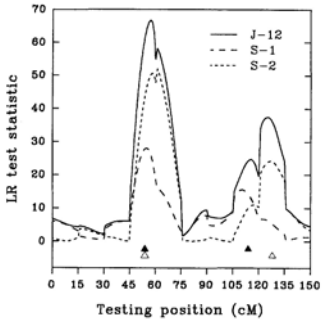
ellipses centered on genotypic value
width for nominal frequency
main axis angle environmental correlation
3 QTL, F2
27 genotypes

note signs of genetic and environmental correlation



pleiotropy or close linkage?

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



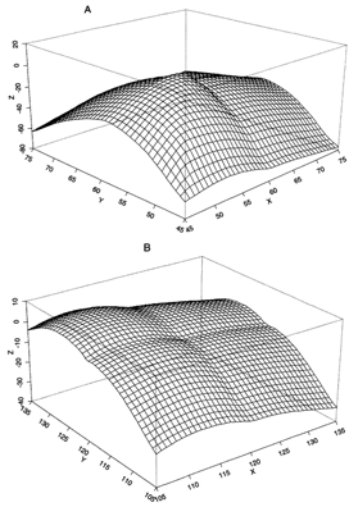
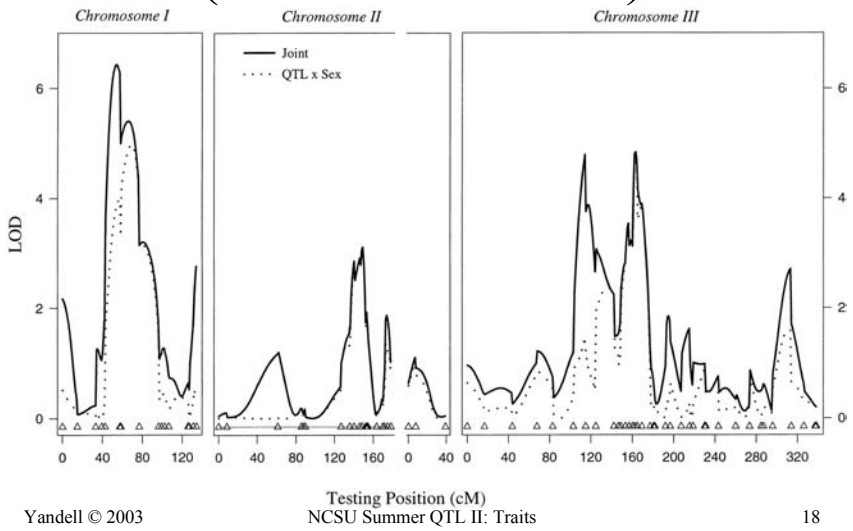


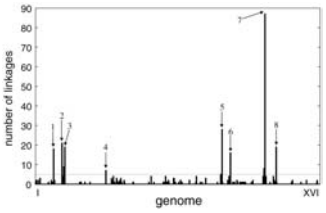
FIGURE 2.—Two-dimensional log-likelihood surfaces (expressed as deviations from the maximum of the log-likelihoods on the diagonal) for the test of pleiotropy vs. close linkage are presented for two regions: the region between 45 and 75 cM of Figure 1 (A) and the region between 105 and 135 cM (B). *X* 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.

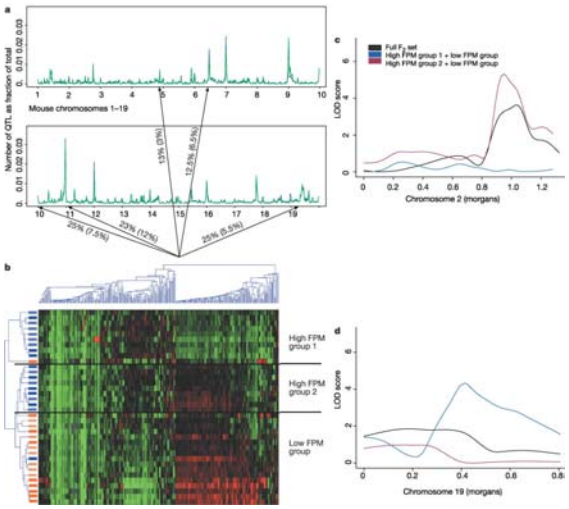
QTL x sex interaction (Vieira et al. 2000)



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

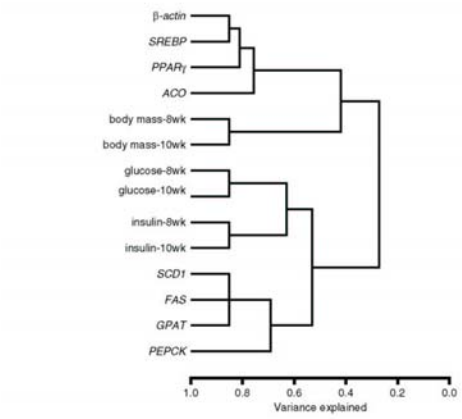
expression pleiotropy in yeast genome (Brem et al. 2002)





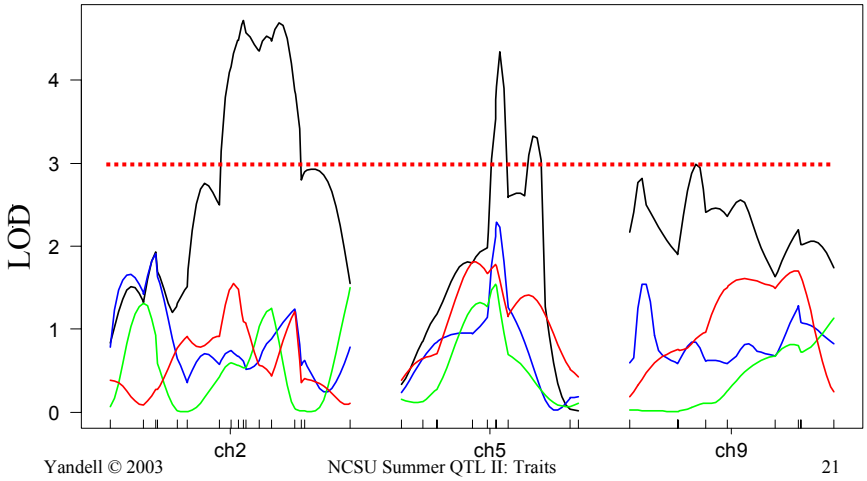
high throughput: which genes are the key players?

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?

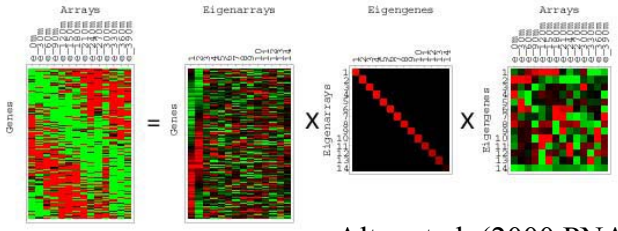


high throughput dilemma

- 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
- mapping principle component as quantitative trait
  - multiple interval mapping with epistatic interactions
  - Liu et al. (1996 *Genetics*); Zeng et al. (2000 *Genetics*) Mahler et al. (2002 *Genomics*)
  - elicit biochemical pathways (Henderson et al. Hoeschele 2001; Ong Page 2002)

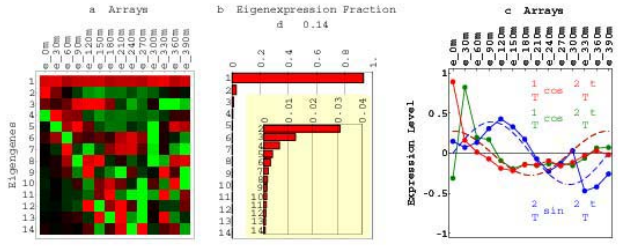
QTL via Principal Components

- PC or SVD decomposition of multiple traits
  - $Y = n \times t$ matrix: $Y = UDW^{\mathrm{T}}$
  - U, W = orthonormal transforms (eigen-vectors)
  - D = diagonal matrix with singular values
- transform problem to principal components
  - $Y^* = D^{-1/2}U^{\mathrm{T}}Y$ has uncorrelated PC traits
  - $Y^* = \mu^* + G^*_Q + e^*$
- interval map each PC separately
  - $Y^*_{1i} = \mu^*_{1i} + G^*_1(Q_i) + e^*_{1i}$
  - may only need to map a few PCs



$Y = UDW^{\mathrm{T}}$

Alter et al. (2000 PNAS)
yeast cell cycle



QTL via Principal Components

- example: *Drosophila* reproduction
  - Liu et al. (1996); Zeng et al. (2000); ch. 7
- other refs
  - 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

PC summary of shape phenotype

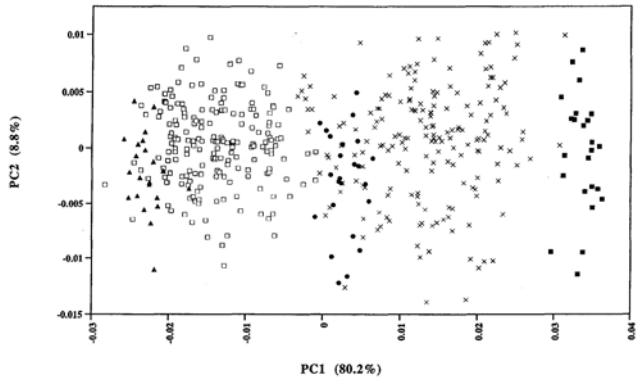


FIGURE 5.—A plot of the first two principal components of the Fourier coefficients from posterior lobe outlines. Many individuals from each of five genotypic classes are represented. Each point represents an average of scores from the left and right sides of an individual (with a few exceptions for which the score is from one side only). The percentage of variation in the Fourier coefficients accounted for by each principal component is given in parentheses. Liu et al. (1996) *Genetics*

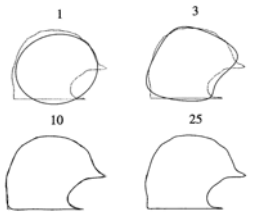
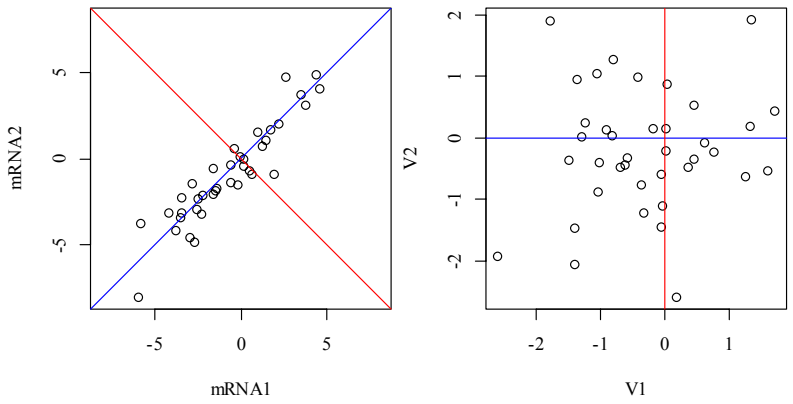
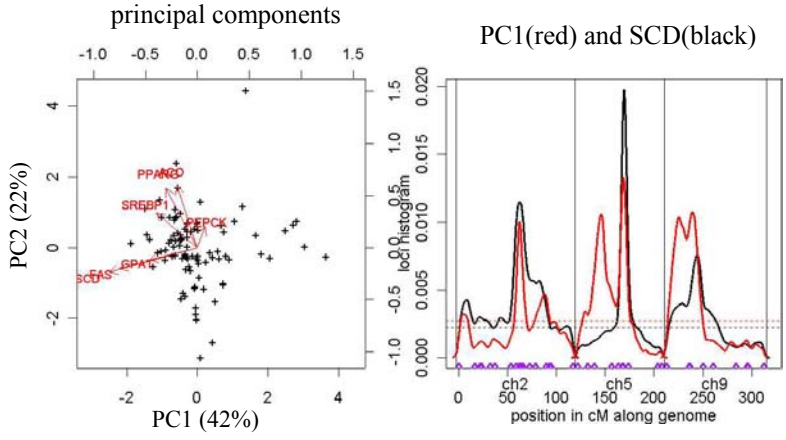


FIGURE 2.—The effect of harmonic number on the accuracy of reconstruction of a posterior lobe outline by elliptical Fourier analysis.

PC simply rotates & rescales to find major axes of variation



multivariate screen for gene expressing mapping



Relation of Composite Phenotypes to Individual mRNA Expressions
(after West et al. 2000)

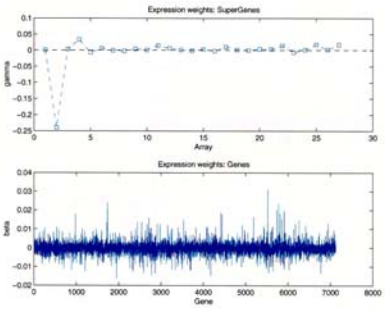


Figure 6: Summary of binary regression fit: Regression parameters

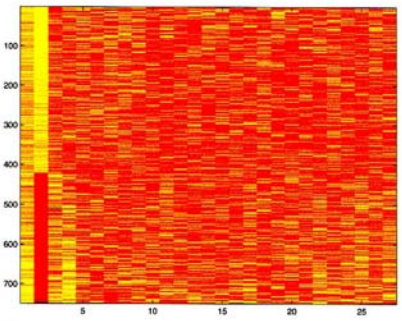
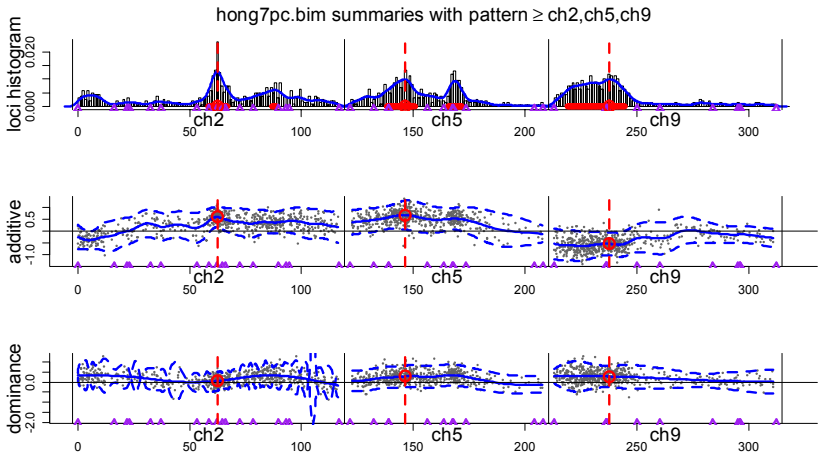
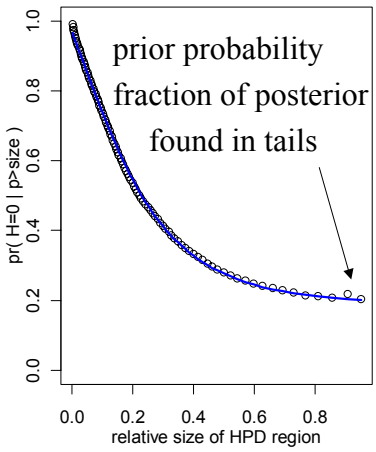


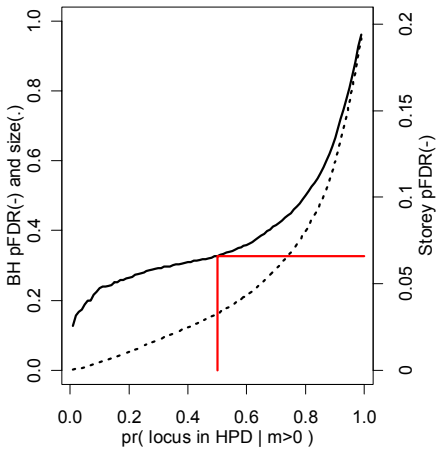
Figure 9: Factor loadings **A** for top 750 genes

mapping first diabetes PC as a trait



pFDR for PC1 analysis





interval-dependent PC

- want to reduce dimensionality while focusing on QTL differences
- interval-dependent PC on residuals
  - reduces dimensionality by identifying patterns in residuals not explained by QTL
  - not quite discriminant analysis: does not aim to best discriminate among QTL genotypes
- pleiotropy highlighted by PCs
  - find strongest correlation
  - interval map using transformed data
- Allison et al. (1998); Korol et al. (2001)

interval-dependent PC details

- flanking marker based PC
  - remove effect of flanking markers (Haley-Knott)
  - PC decomposition of residuals
  - transform original data to new PC axes
  - map transformed data on interval
  - problem: bias, variance for Haley-Knott
- improvement through iteration
  - PC decomposition of residuals of transformed data
  - repeat until estimates stabilize