Inferring Genetic Architecture
of Complex Biological Processes

BioPharmaceutical Technology Center Institute (BTCI)

Brian S. Yandell
University of Wisconsin-Madison

http://www.stat.wisc.edu/~yandell/statgen

Multiple Traits & Microarrays
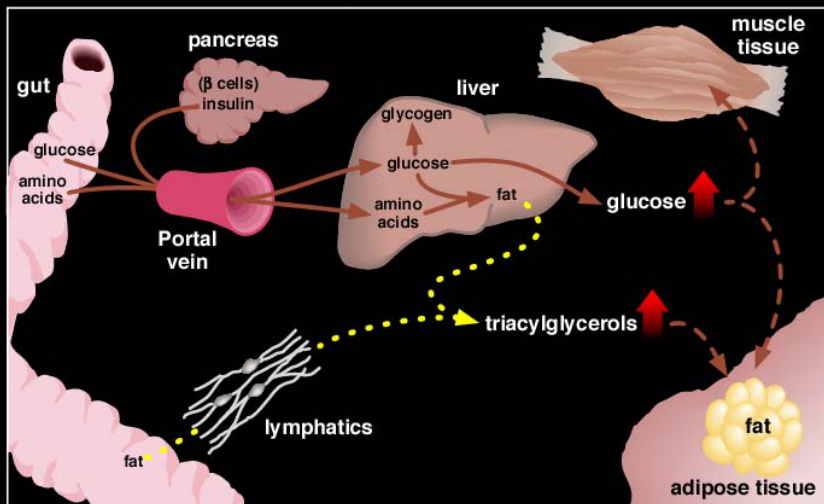
1. why study multiple traits together?

- 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
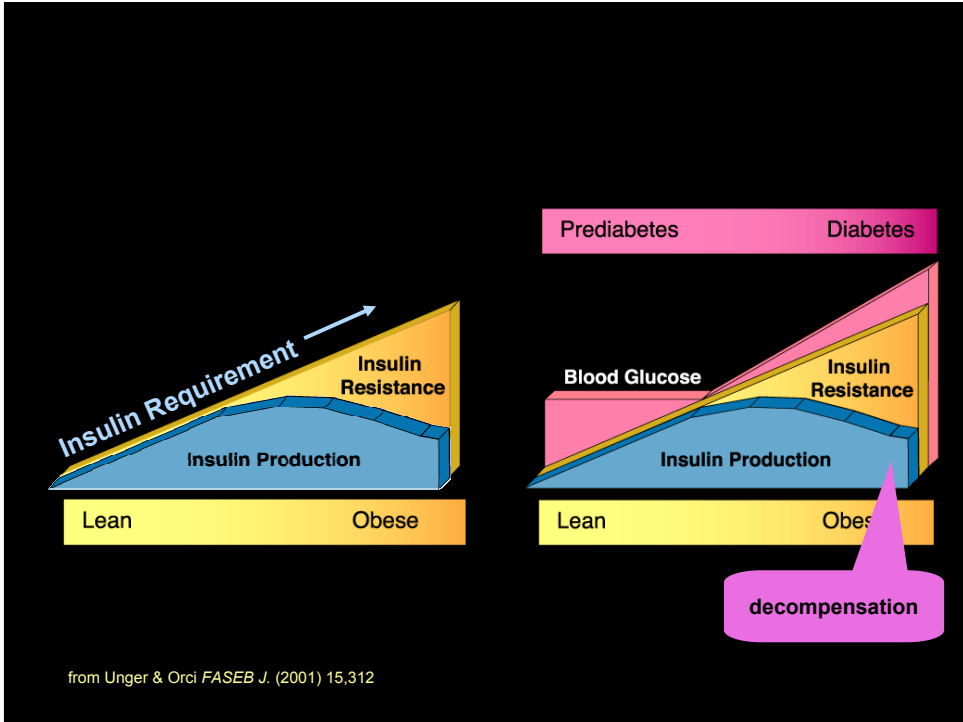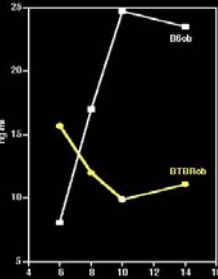- increase power to detect QTL

Type 2 Diabetes Mellitus

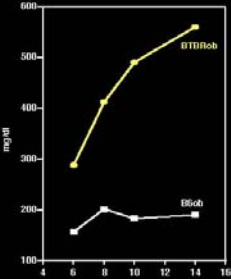from Unger & Orci *FASEB J.* (2001) 15,312

Insulin Resistant Mice

Bill Dove

BTBR strain

glucose   insulin

insulin resistance alleles

+   ???   →   diabetes

obesity

(courtesy AD Attie)
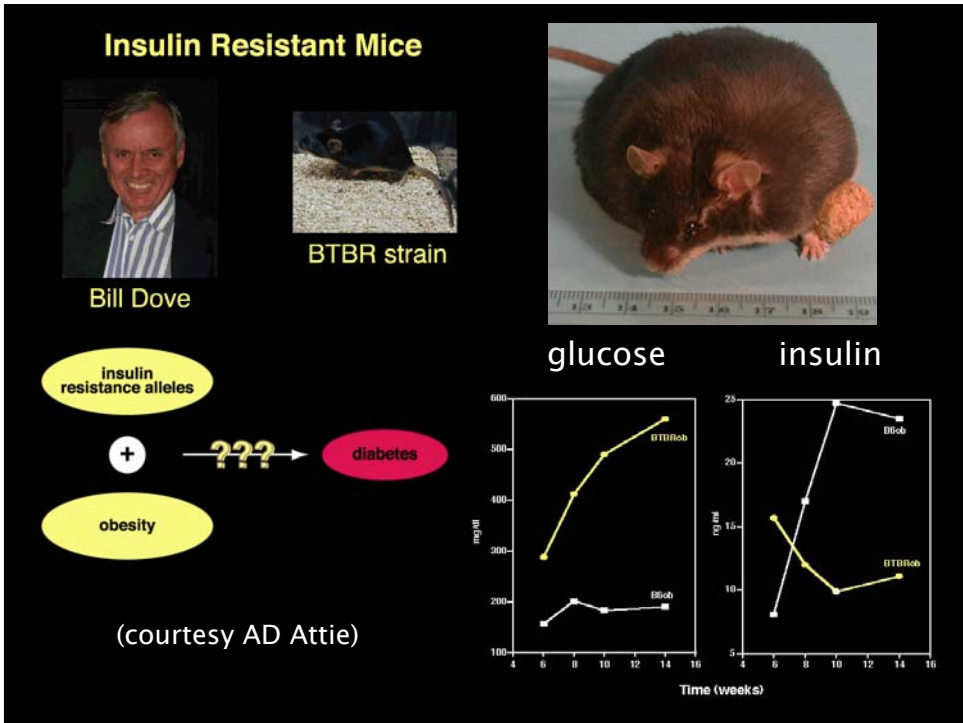
studying diabetes in an F2

- segregating cross of inbred lines
  - B6.ob x BTBR.ob $\rightarrow$ F1 $\rightarrow$ 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.)

The intercross (from K Broman)
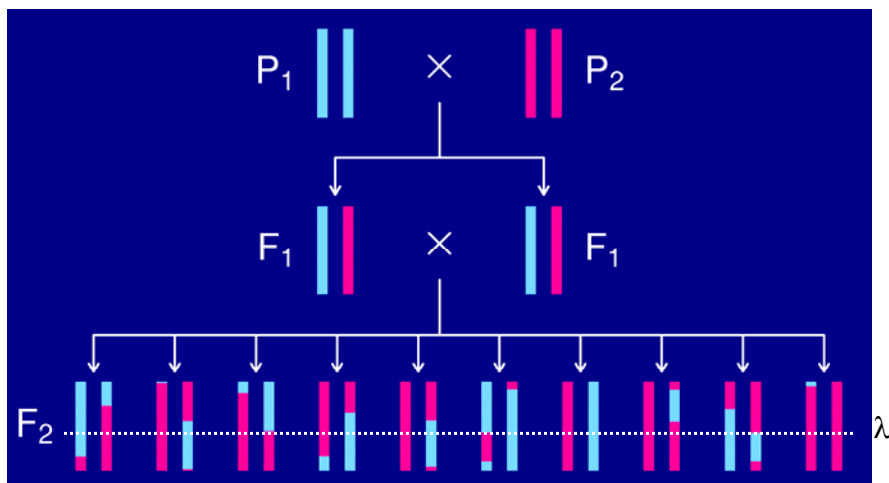
why map gene expression as a quantitative trait?

- *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)
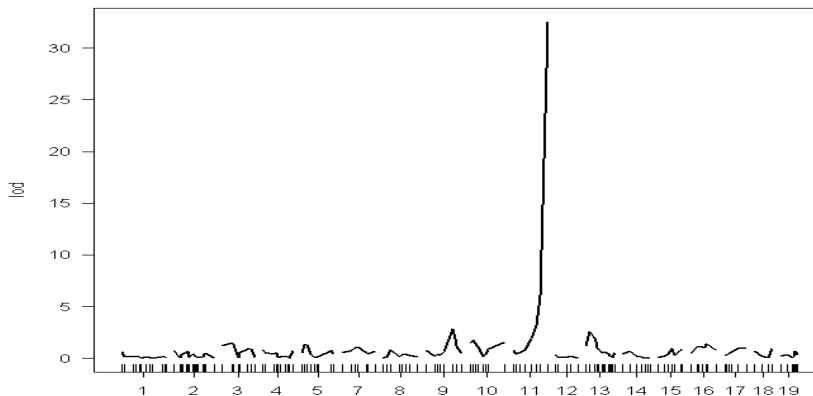 - adjust for multiple testing

LOD map for PDI:
cis-regulation (Lan et al. 2003)
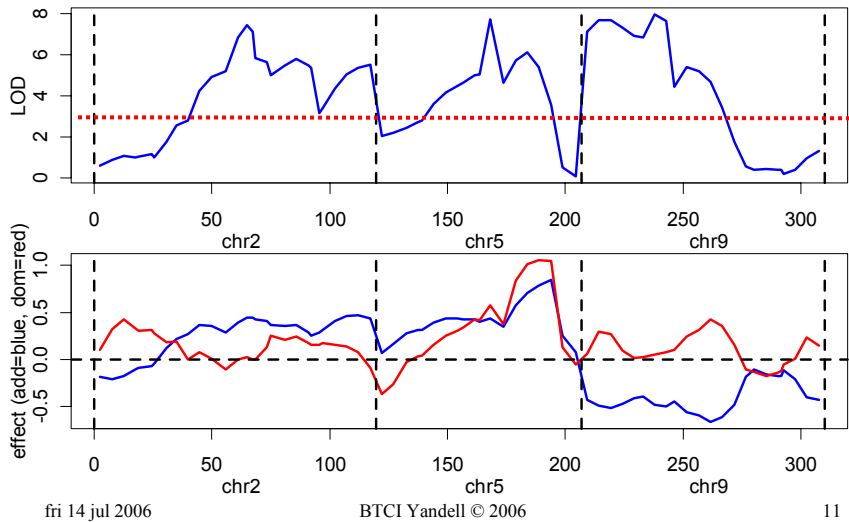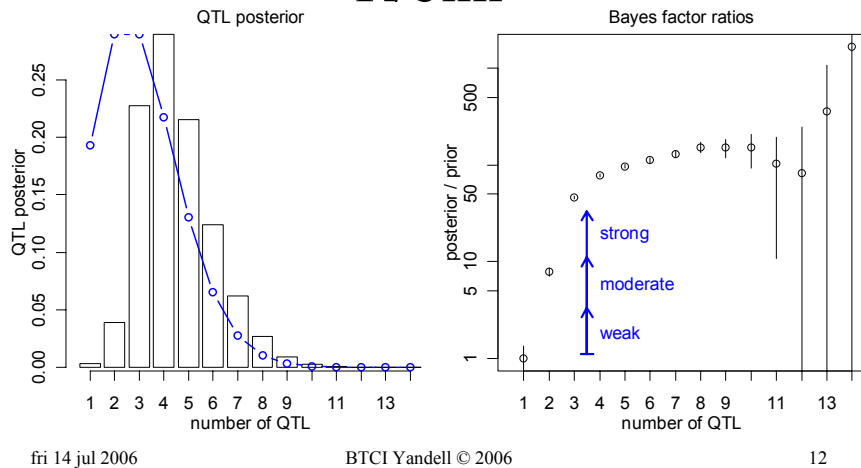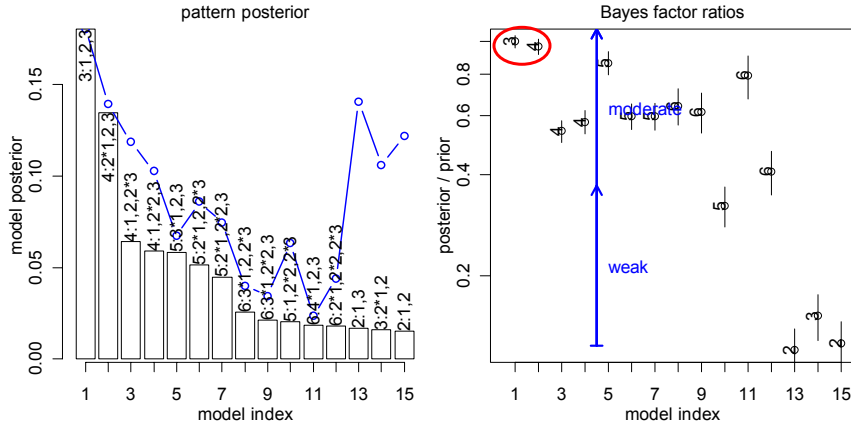
mRNA expression as phenotype:
interval mapping for SCD1 is complicated

Bayesian model assessment:
number of QTL for SCD1 with R/bim

Bayesian model assessment
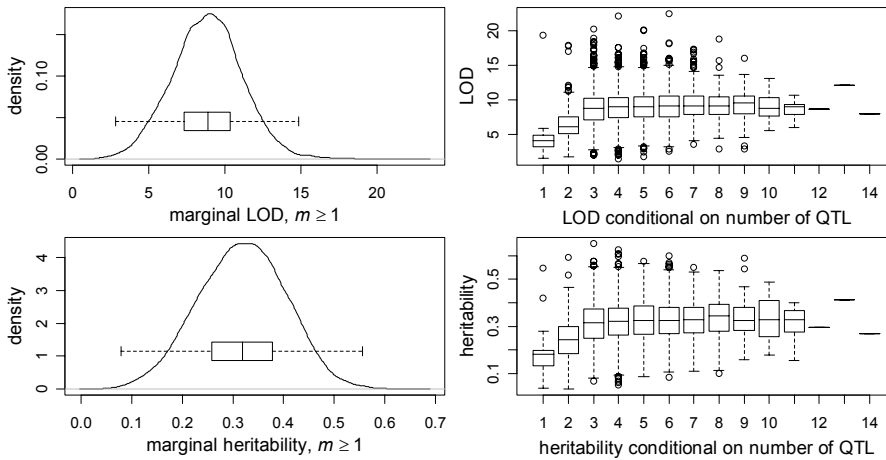## genetic architecture: chromosome pattern

pattern posterior

Bayes factor ratios

Bayesian LOD and h^2 for SCD1
(summaries from R/bim)

trans-acting QTL for SCD1
Bayesian model averaging with R/bim

hong7scd.bim summaries with pattern ≥ ch2,ch5,ch9

additive QTL?

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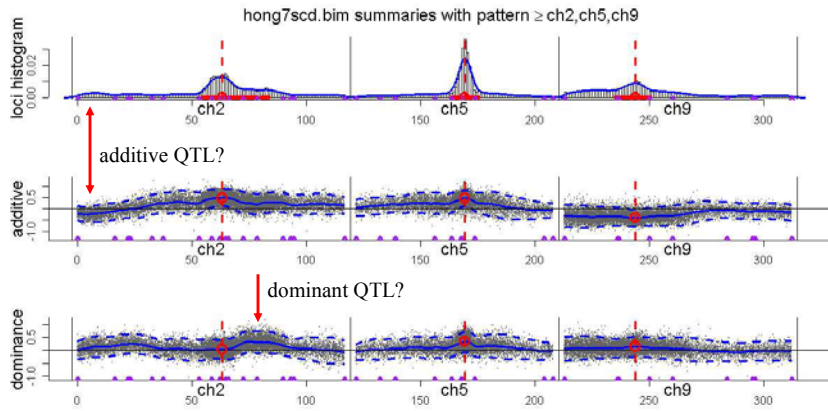
1-D and 2-D marginals
pr(QTL at λ | *Y,X, m*)

unlinked loci   linked loci
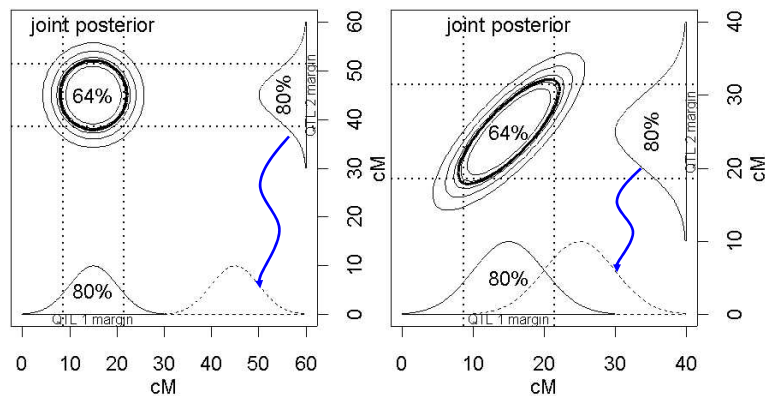
2-D scan: assumes only 2 QTL!

SCD on chr 2,5,9

epistasis
LOD
peaks

joint
LOD
peaks
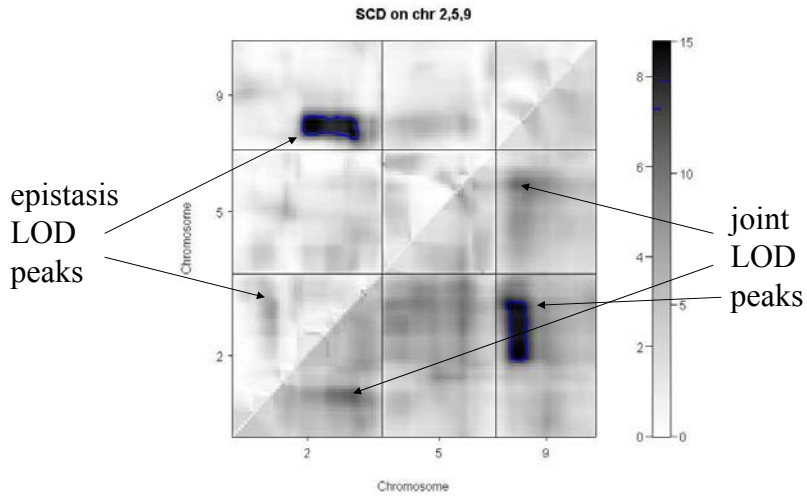
sub-peaks can be easily overlooked!

SCD: peak LOD = 11.02   SCD: peak LOD = 10.74
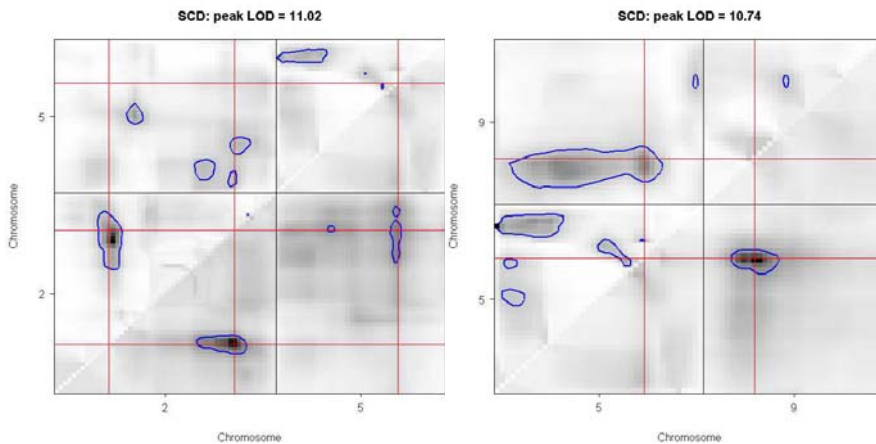
epistatic model fit
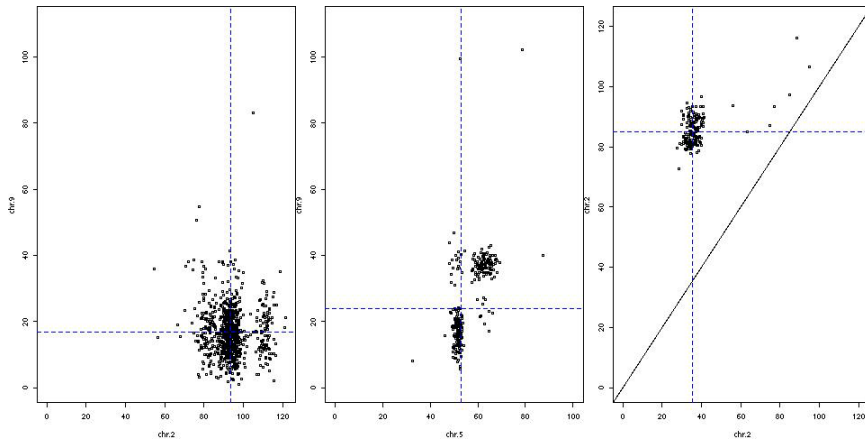
Cockerham epistatic effects
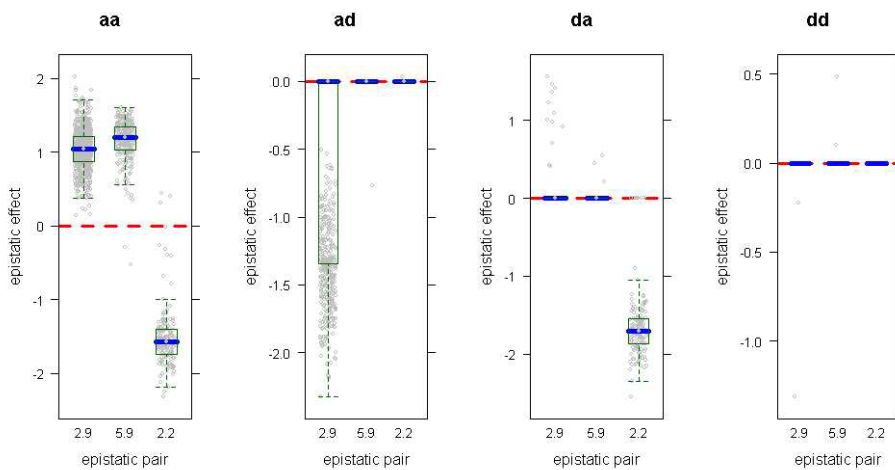
mapping microarray data

- 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*)

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

selective phenotyping

- emphasize additive effects in F2
  - F2 design: 1QQ:2Qq:1qq
  - best design for additive only: 1QQ:1Qq
  - drop heterozygotes (Qq)
  - reduce sample size by half with no power loss
- emphasize general effects in F2
  - best design: 1QQ:1Qq:1qq
  - drop half of heterozygotes (25% reduction)
- multiple loci
  - same idea but care is needed
 - drop 7/16 of sample for two unlinked loci

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

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
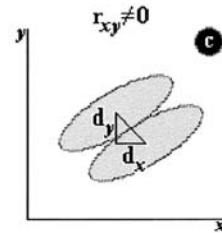   - genotypes Q are collinear

interplay of pleiotropy & correlation
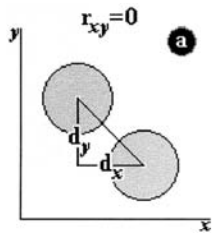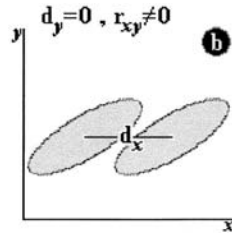
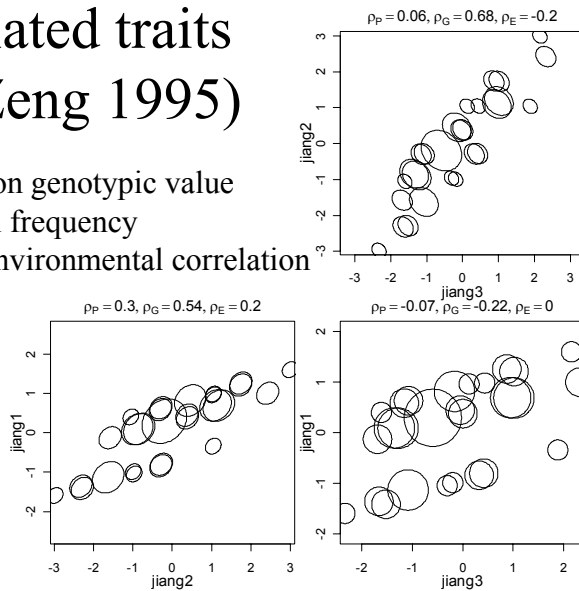pleiotropy only   correlation only   both

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
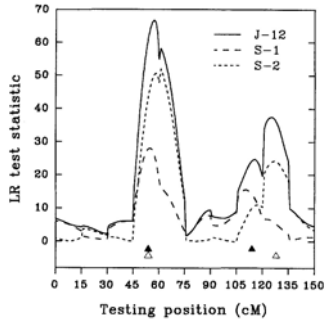genetic and
environmental
correlation

pleiotropy or close linkage?

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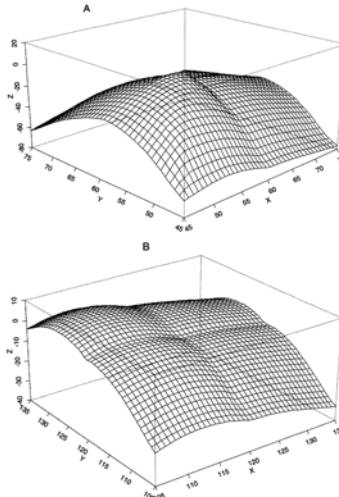linkage @ 114,128cM
Jiang Zeng (1995)

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.