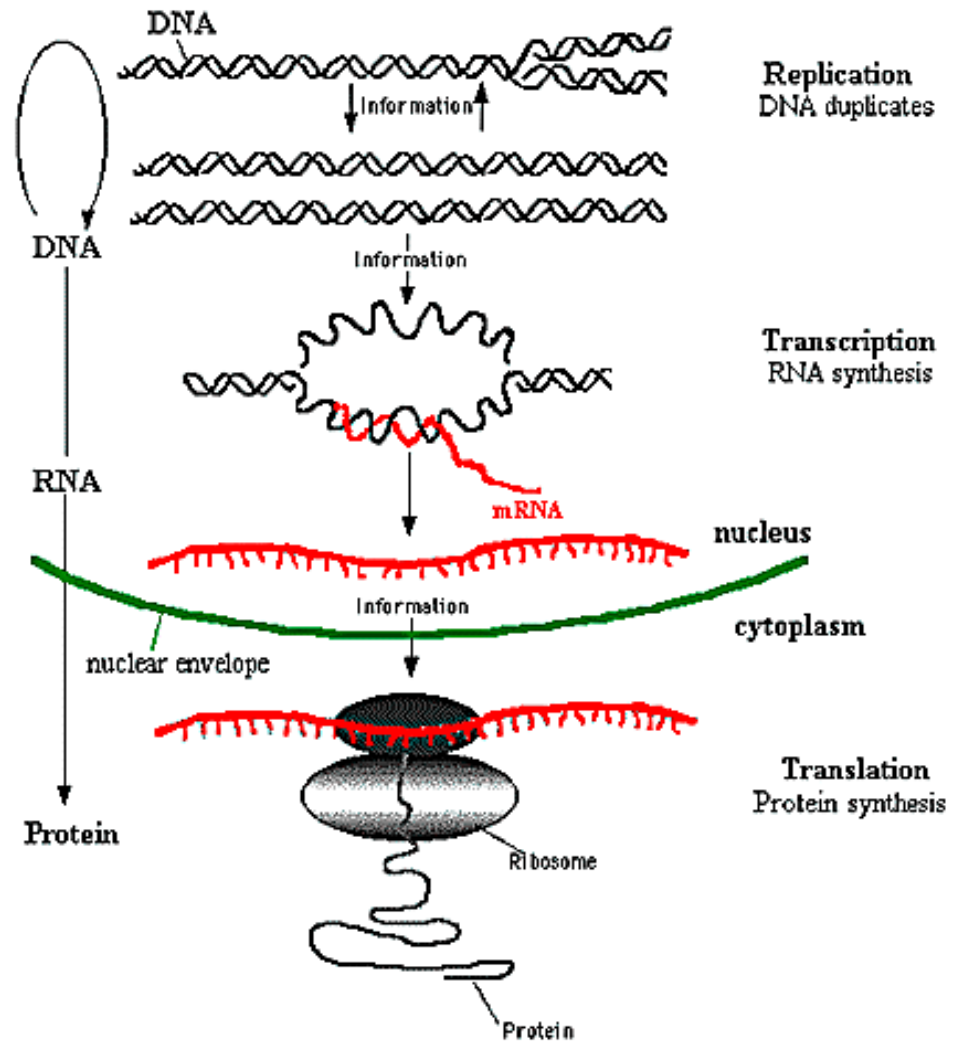
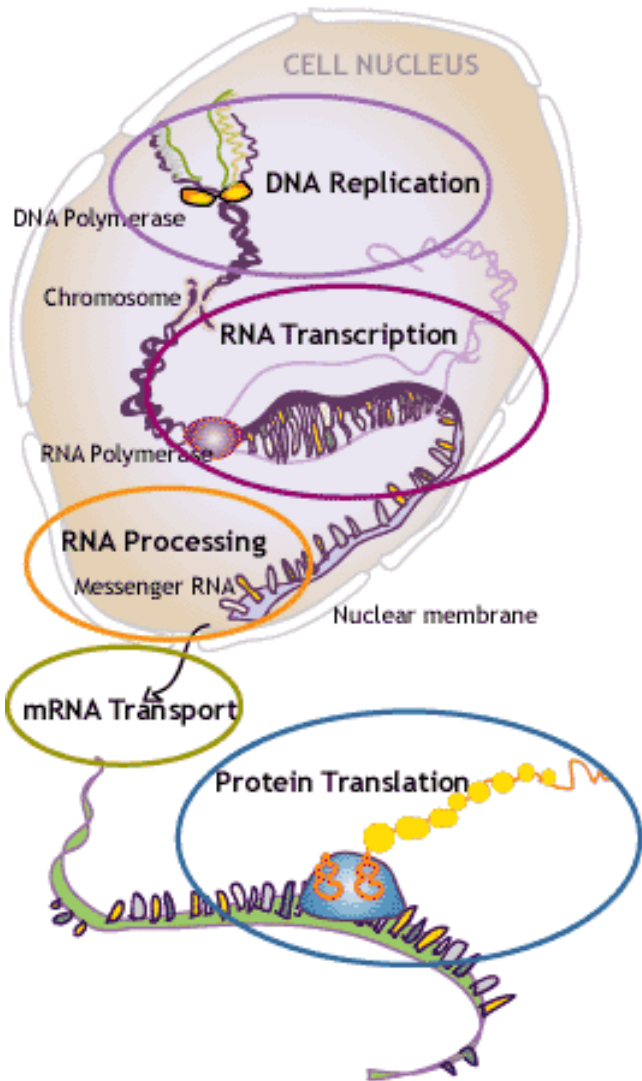


Gene Mapping for Correlated Traits

Brian S. Yandell

University of Wisconsin-Madison

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

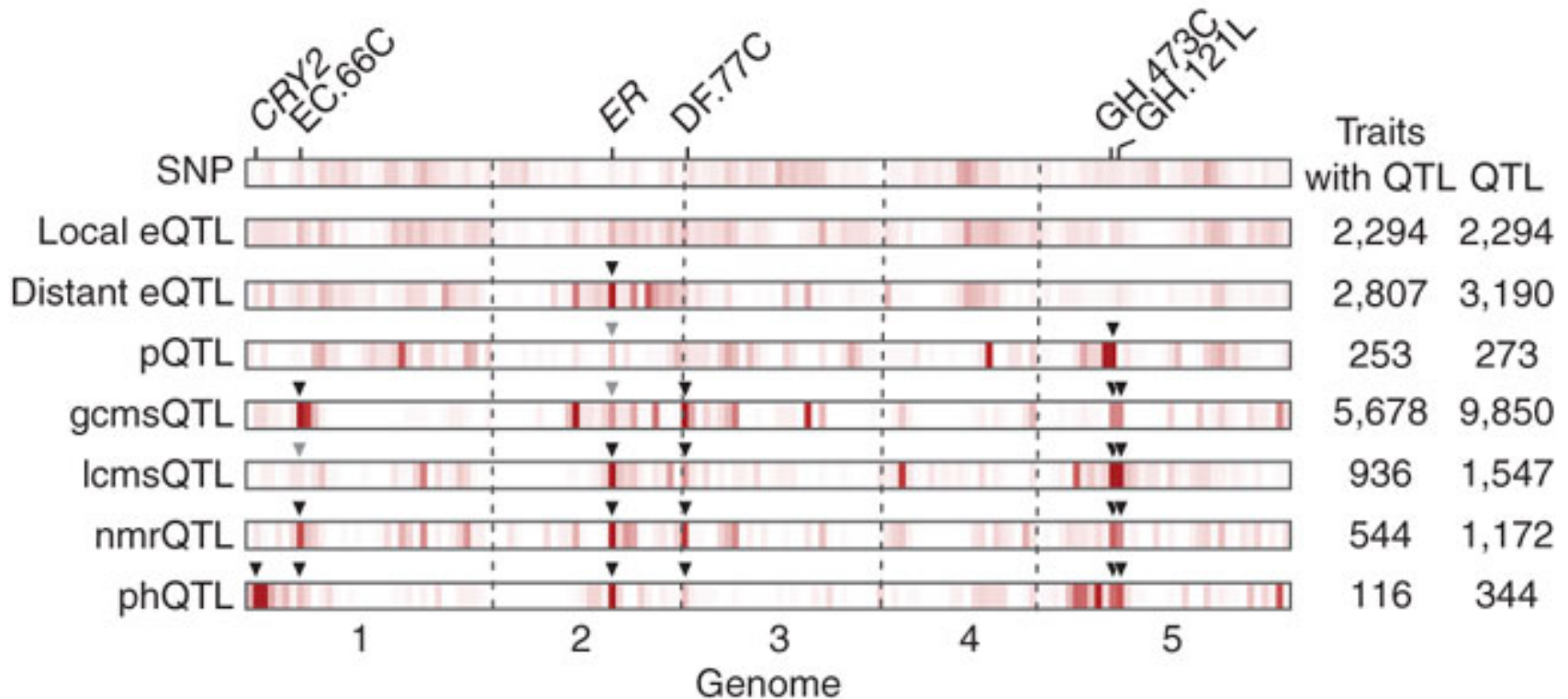


The Central Dogma of Molecular Biology

www.nobelprize.org/educational/medicine/dna

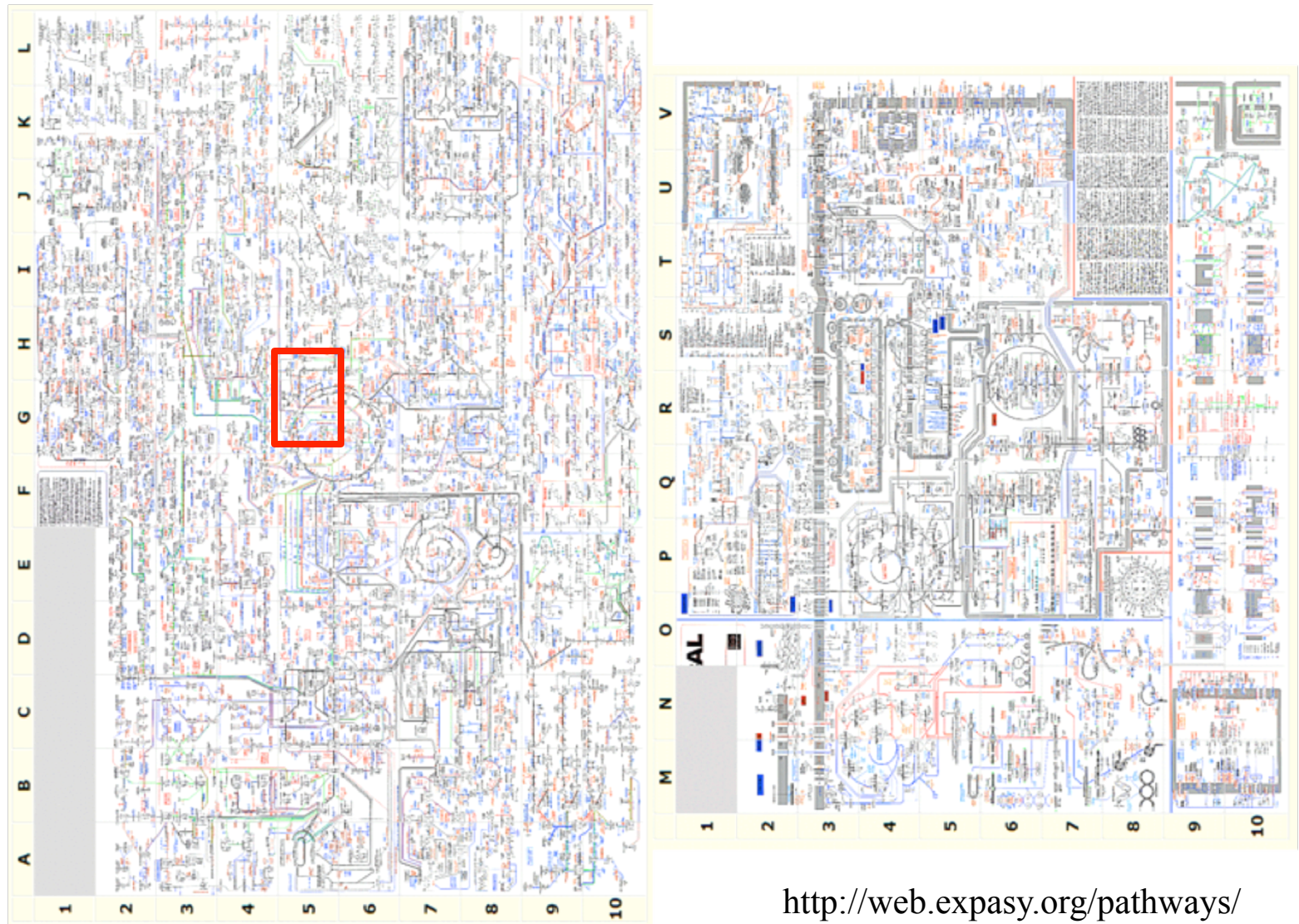
www.accessexcellence.org/RC/VL/GG/central.php

phenotypic buffering of molecular QTL

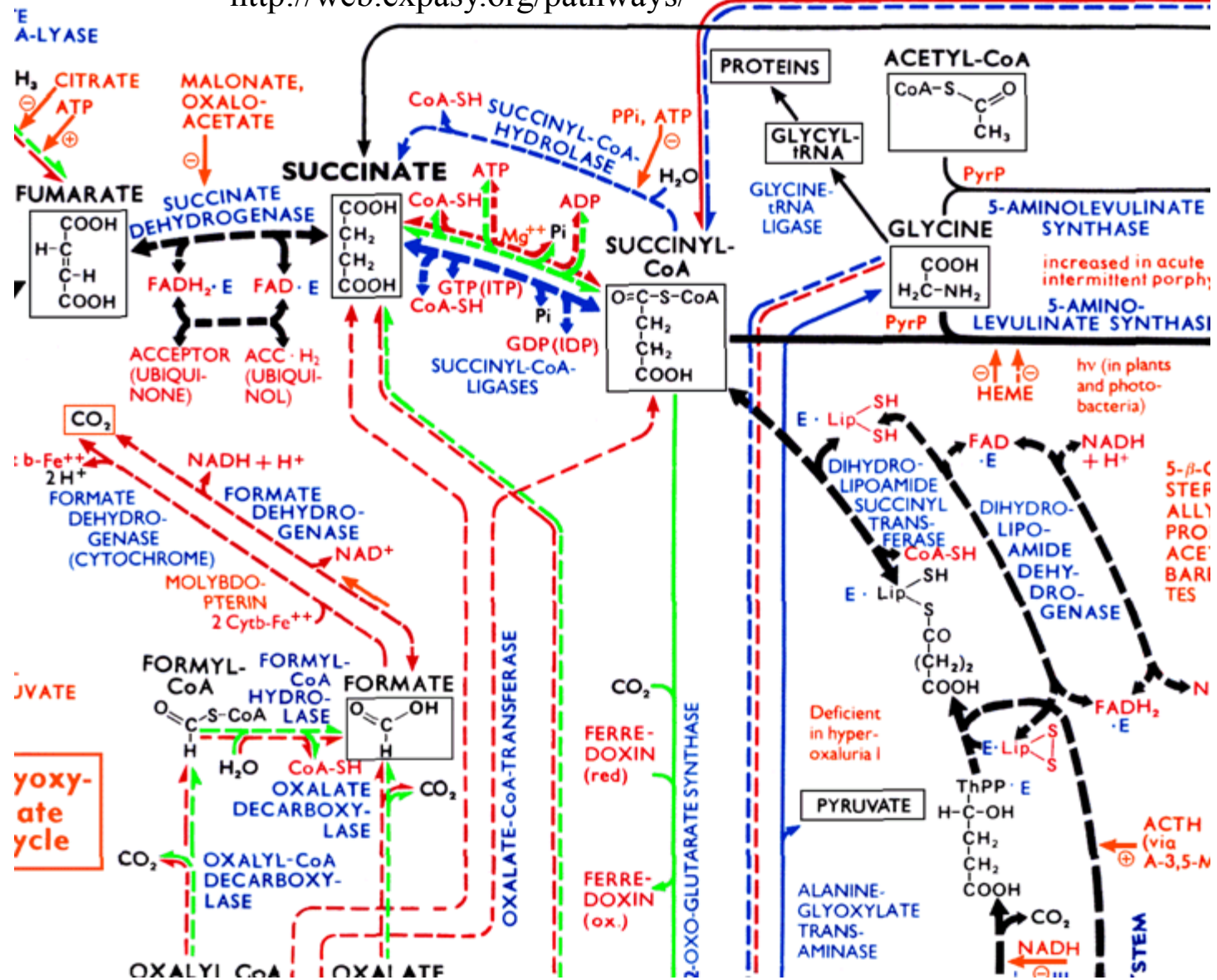


Fu et al. Jansen (2009 *Nature Genetics*)

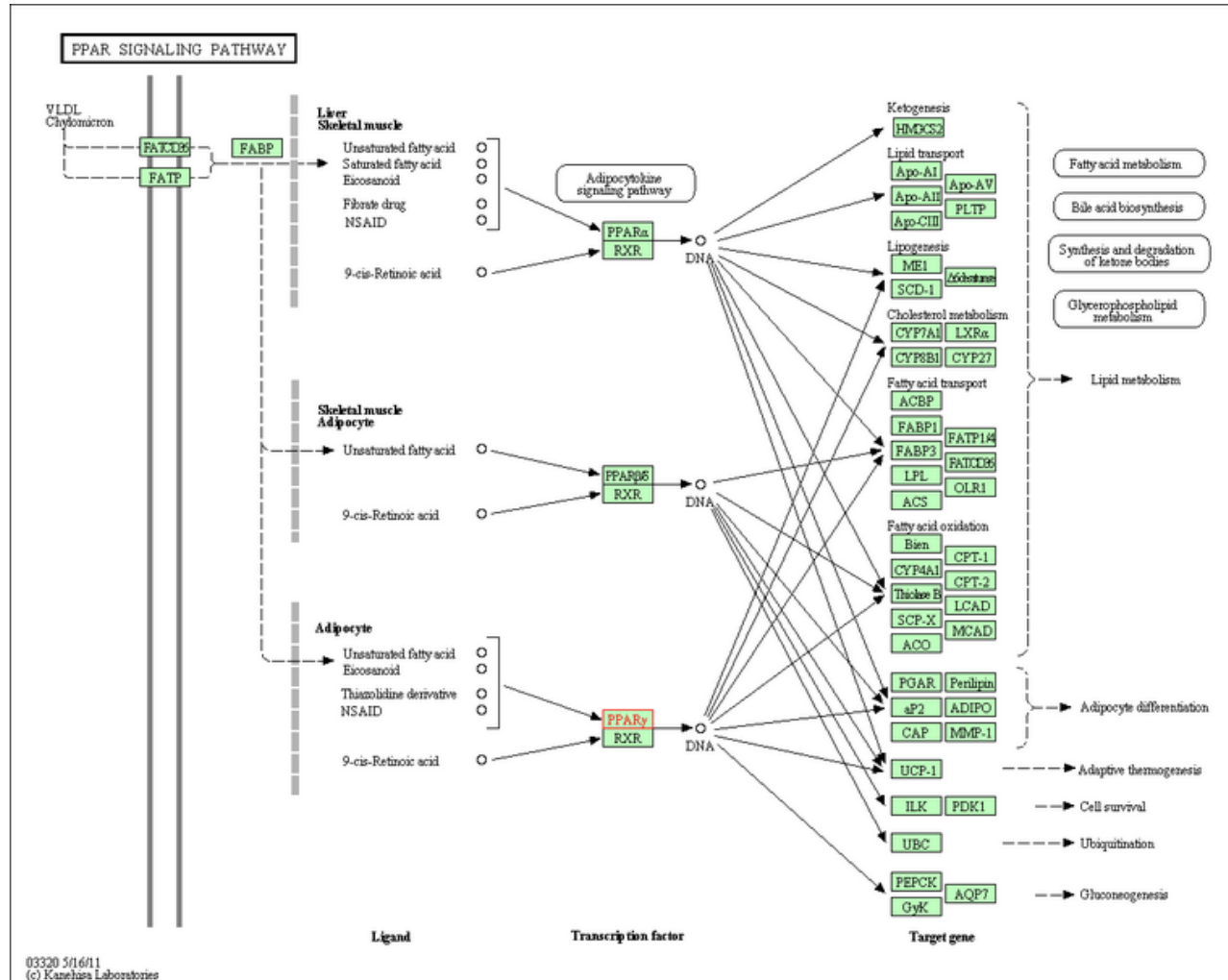
Biochemical Pathways chart, Gerhard Michal, Beohringer Mannheim



<http://web.expasy.org/pathways/>



KEGG pathway: pparg in mouse



systems genetics approach

- study genetic architecture of quantitative traits
 - in model systems, and ultimately humans
- interrogate single resource population for variation
 - DNA sequence, transcript abundance, proteins, metabolites
 - multiple organismal phenotypes
 - multiple environments
- detailed map of genetic variants associated with
 - each organismal phenotype in each environment
- functional context to interpret phenotypes
 - genetic underpinnings of multiple phenotypes
 - genetic basis of genotype by environment interaction

Sieberts, Schadt (2007 *Mamm Genome*); Emilsson et al. (2008 *Nature*)
Chen et al. 2008 *Nature*); Ayroles et al. MacKay (2009 *Nature Genetics*)

brief tutorial on gene mapping for experimental crosses

*** Check out Karl Broman's stuff ***

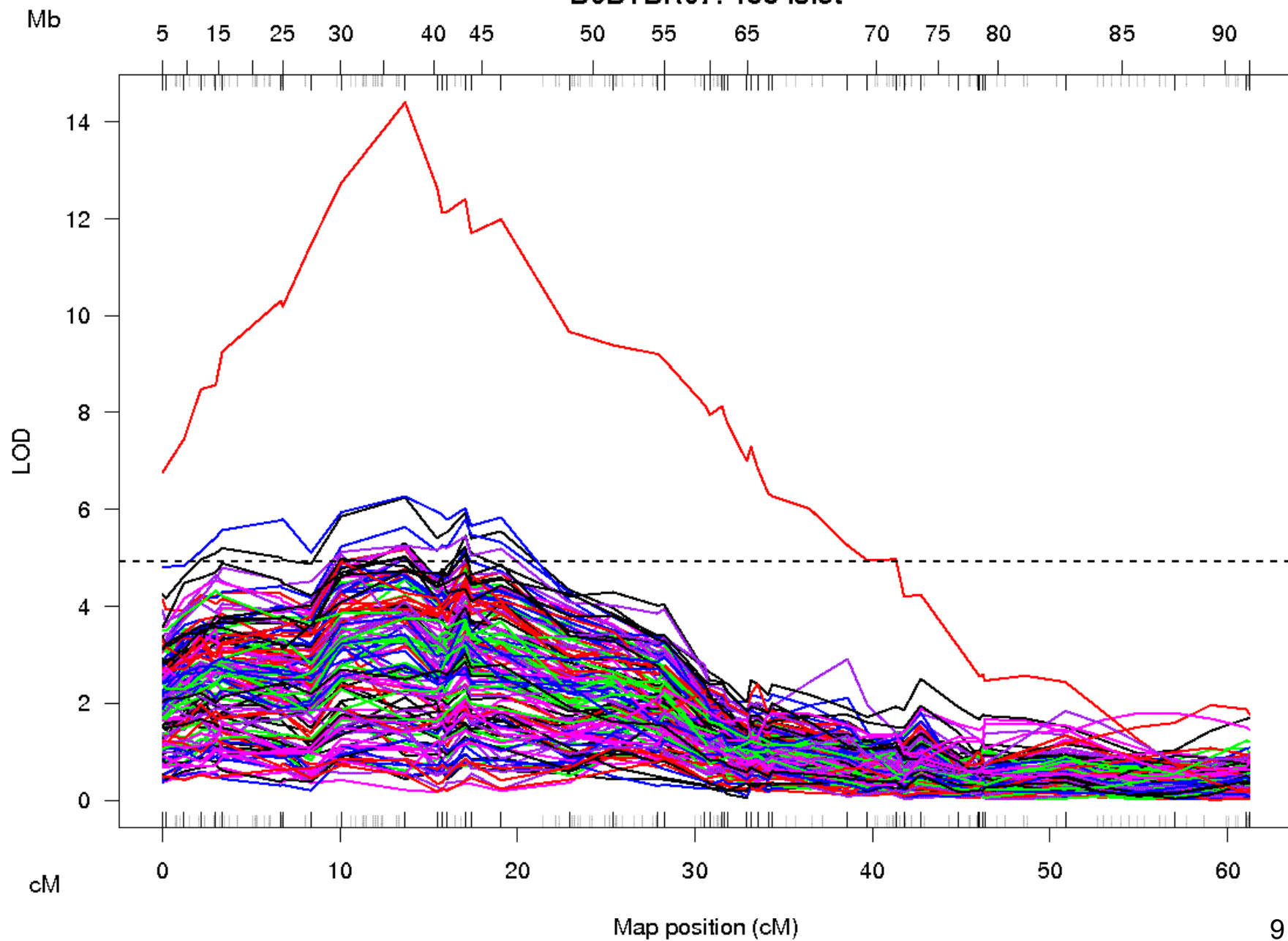
– visit www.rqtl.org

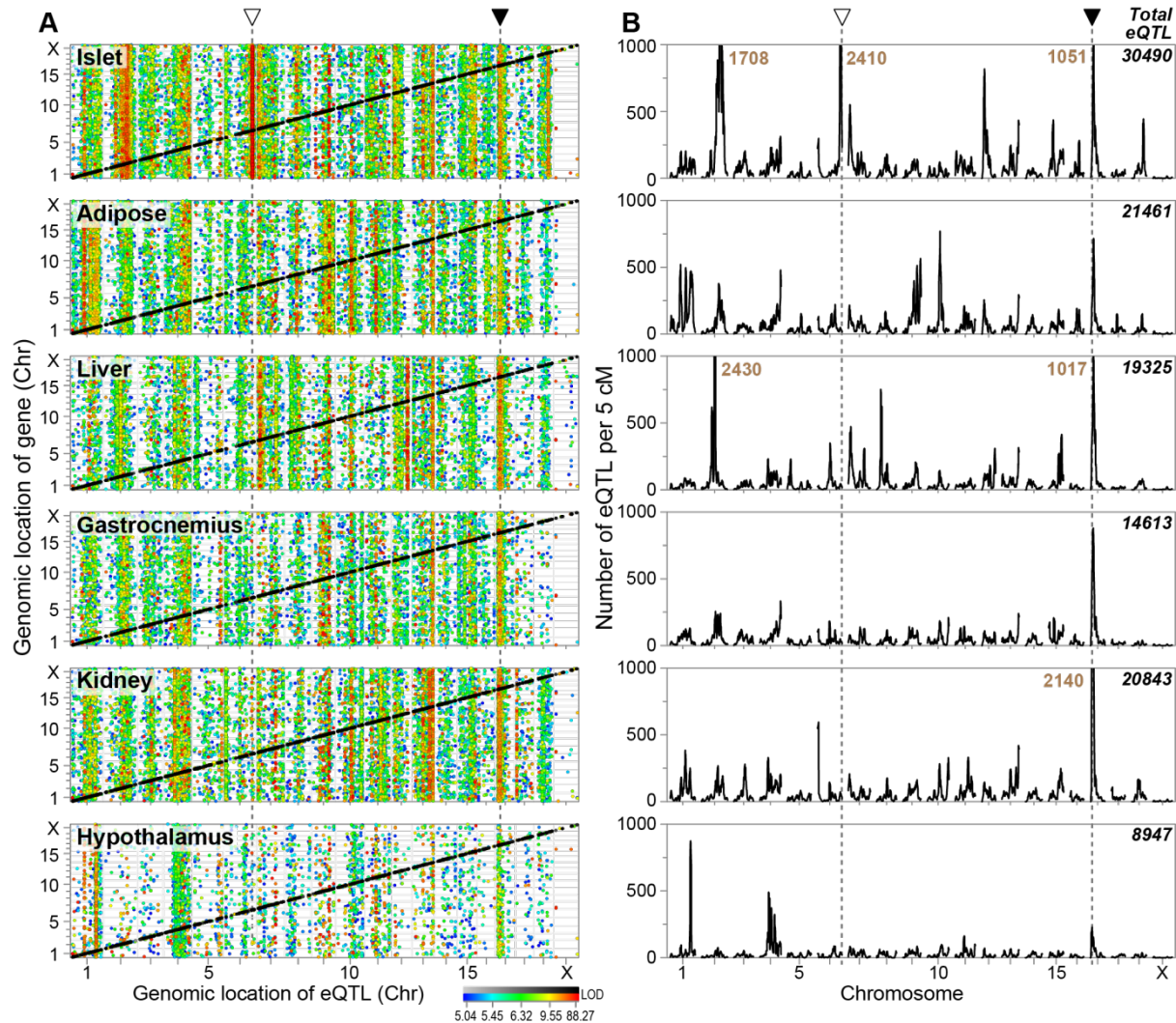
– go through talk on multiple QTLs (for instance)

[http://www.biostat.wisc.edu/~kbroman/
presentations/multiqtl_columbia11.pdf](http://www.biostat.wisc.edu/~kbroman/presentations/multiqtl_columbia11.pdf)

– Work through tutorials at www.rqtl.org

B6BTBR07: 133 islet





Genetic architecture of gene expression in 6 tissues.

A Tissue-specific panels illustrate the relationship between the genomic location of a gene (*y*-axis) to where that gene's mRNA shows an eQTL ($LOD > 5$), as a function of genome position (*x*-axis). Circles represent eQTLs that showed either cis-linkage (black) or trans-linkage (colored) according to LOD score. Genomic hot spots, where many eQTLs map in trans, are apparent as vertical bands that show either tissue selectivity (e.g., Chr 6 in the islet, ∇) or are present in all tissues (e.g., Chr 17, \blacktriangledown). **B** The total number of eQTLs identified in 5 cM genomic windows is plotted for each tissue; total eQTLs for all positions is shown in upper right corner for each panel. The peak number of eQTLs exceeding 1000 per 5 cM is shown for islets (Chrs 2, 6 and 17), liver (Chrs 2 and 17) and kidney (Chr 17).

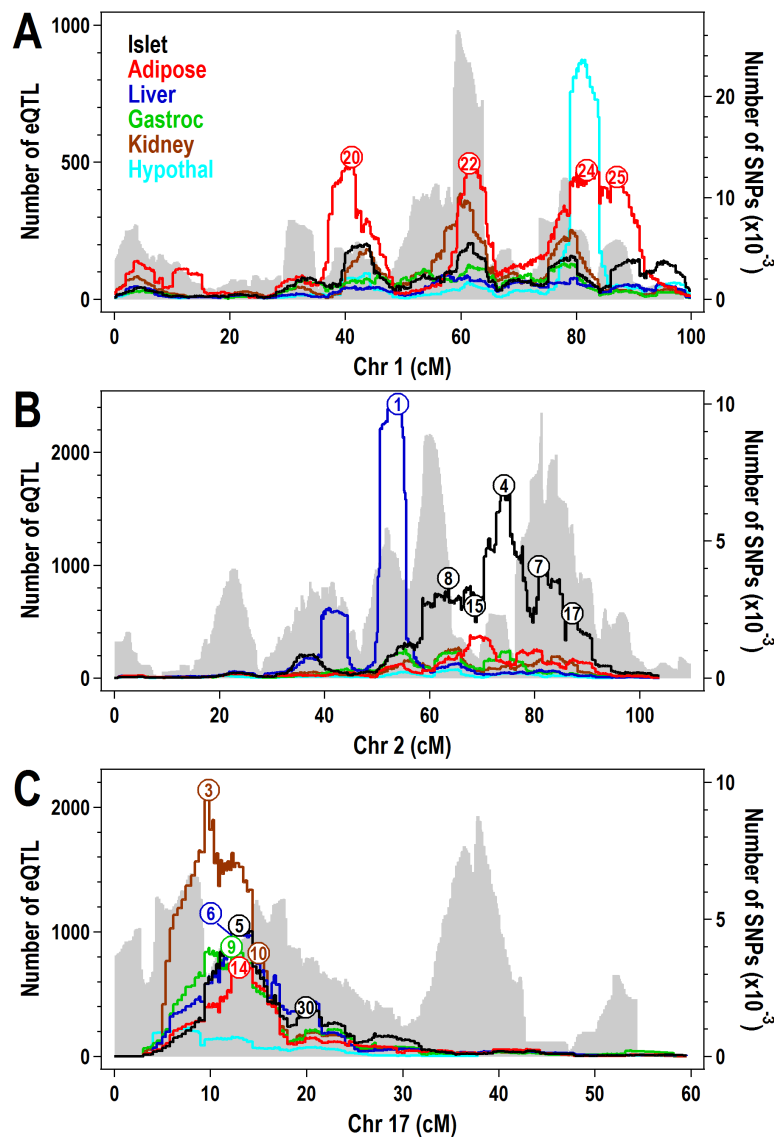
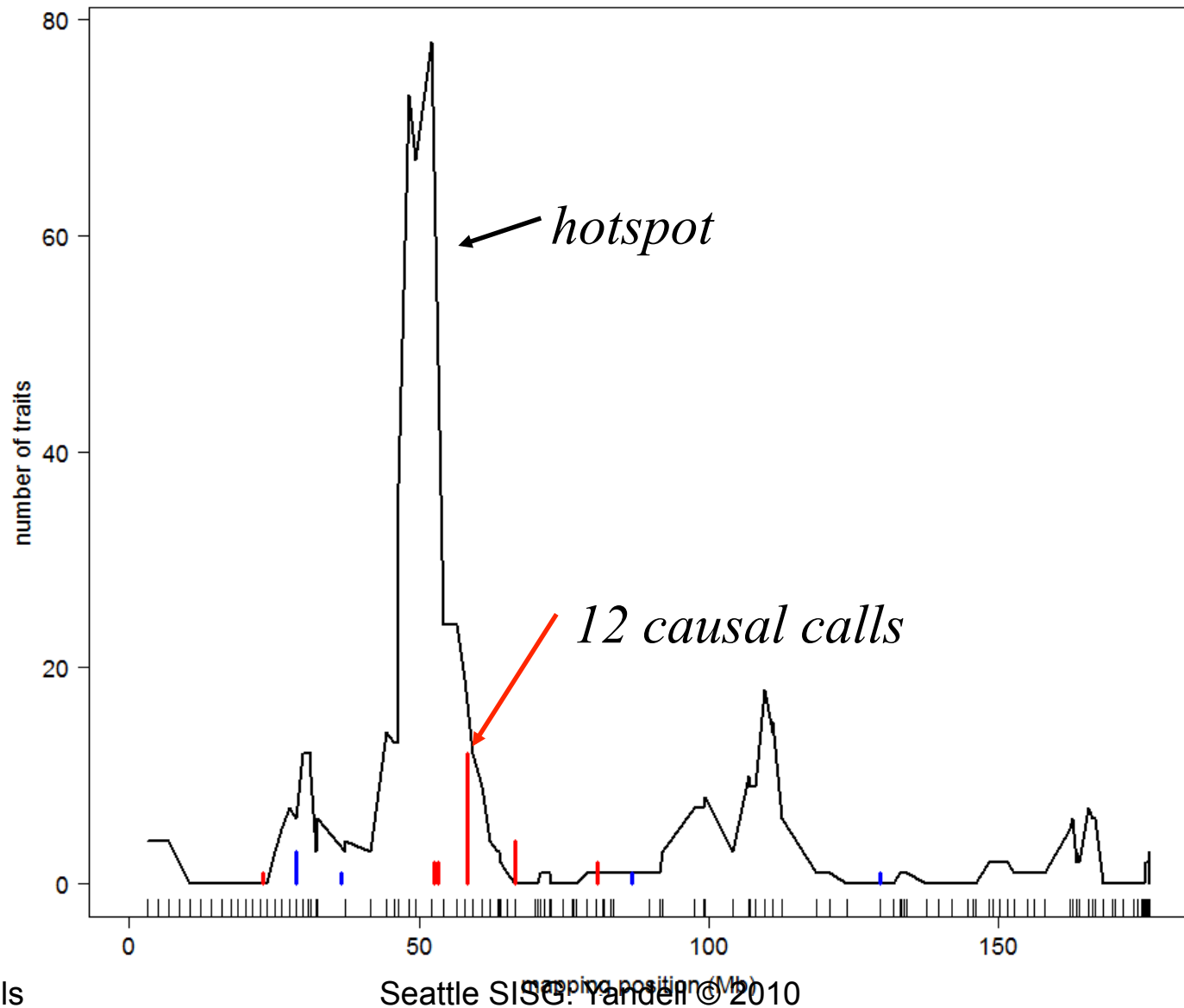


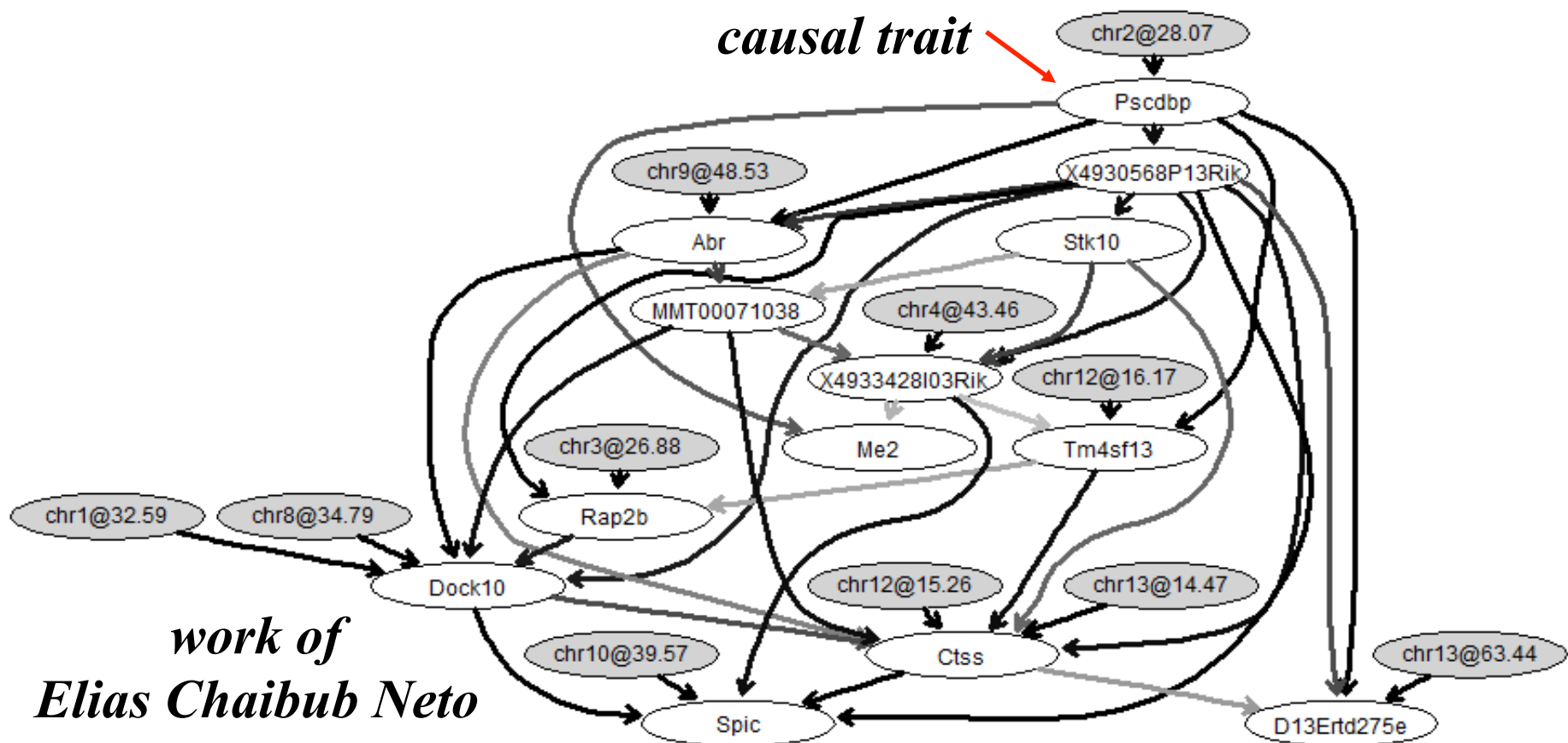
Figure 4 Tissue-specific hotspots with eQTL and SNP architecture for Chrs 1, 2 and 17.

The number of eQTLs for each tissue (left axis) and the number of SNPs between B6 and BTBR (right axis) that were identified within a 5 cM genomic window is shown for Chr 1 (A), Chr 2 (B) Chr 17 (C). The location of tissue-specific hotspots are identified by their number corresponding to that in Table 1. eQTL and SNP architecture is shown for all chromosomes in supplementary material.

BxH ApoE-/- chr 2: causal architecture



BxH ApoE-/- causal network for transcription factor Pscdbp



*work of
Elias Chaibub Neto*

Multiple Correlated Traits

- Pleiotropy vs. close linkage
- Analysis of covariance
 - Regress one trait on another before QTL search
- Classic GxE analysis
- Formal joint mapping (MTM)
- Seemingly unrelated regression (SUR)
- Reducing many traits to one
 - Principle components for *similar* traits

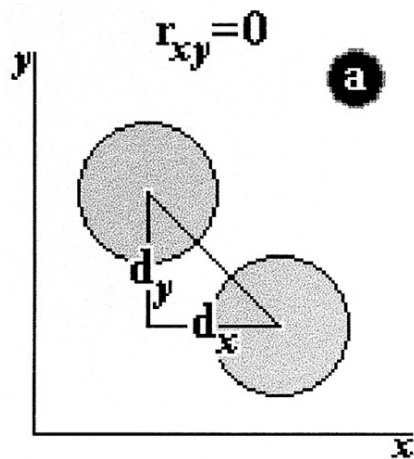
co-mapping multiple traits

- 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

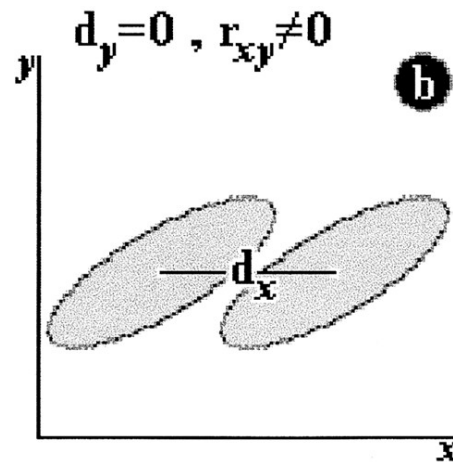
Two types of data

- Design I: multiple traits on same individual
 - Related measurements, say of shape or size
 - Same measurement taken over time
 - Correlation within an individual
- Design II: multiple traits on different individuals
 - Same measurement in two crosses
 - Male vs. female differences
 - Different individuals in different locations
 - No correlation between individuals

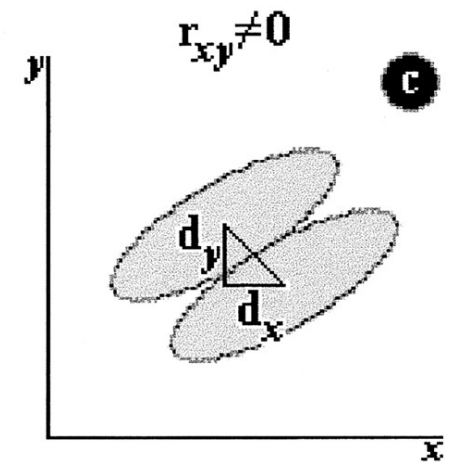
interplay of pleiotropy & correlation



pleiotropy only



correlation only



both

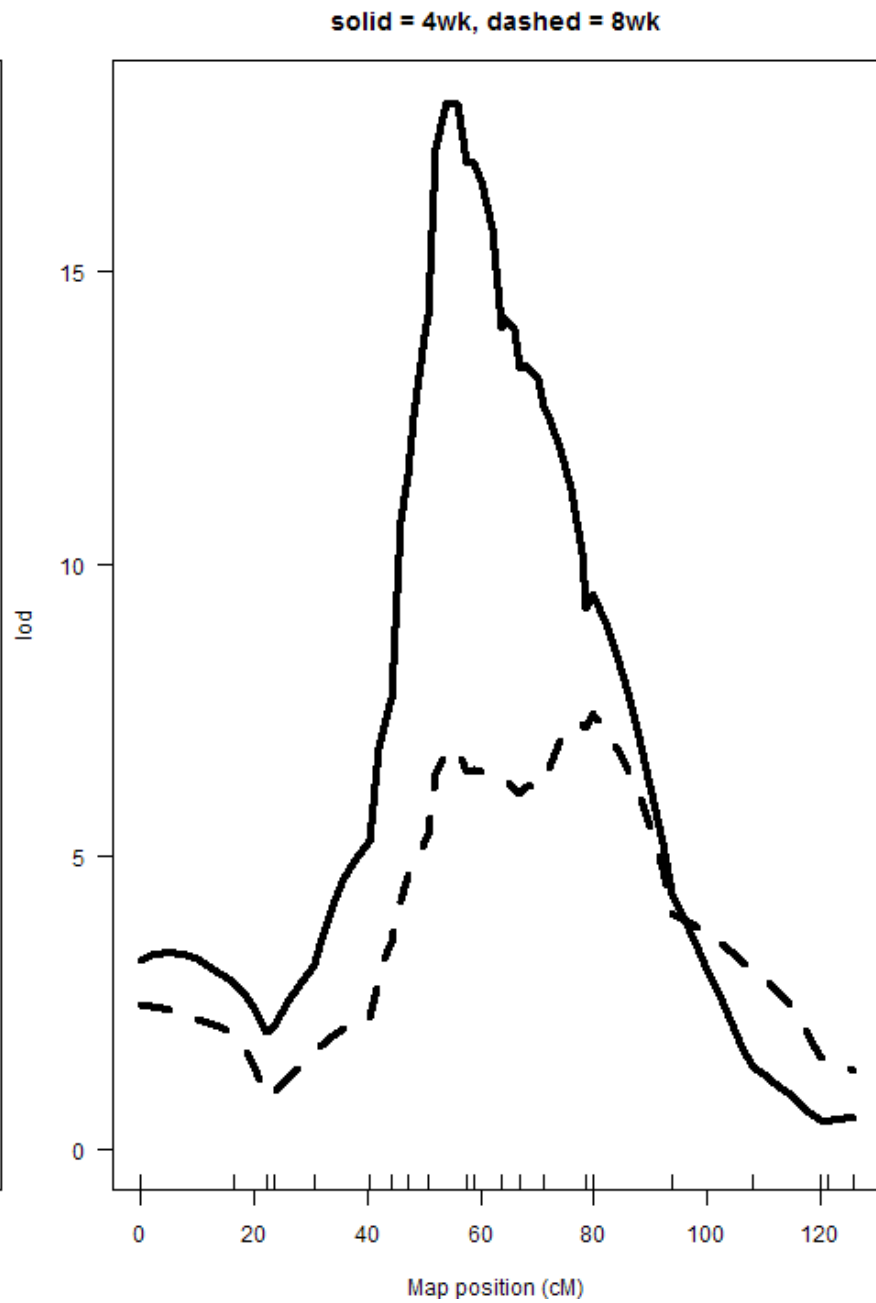
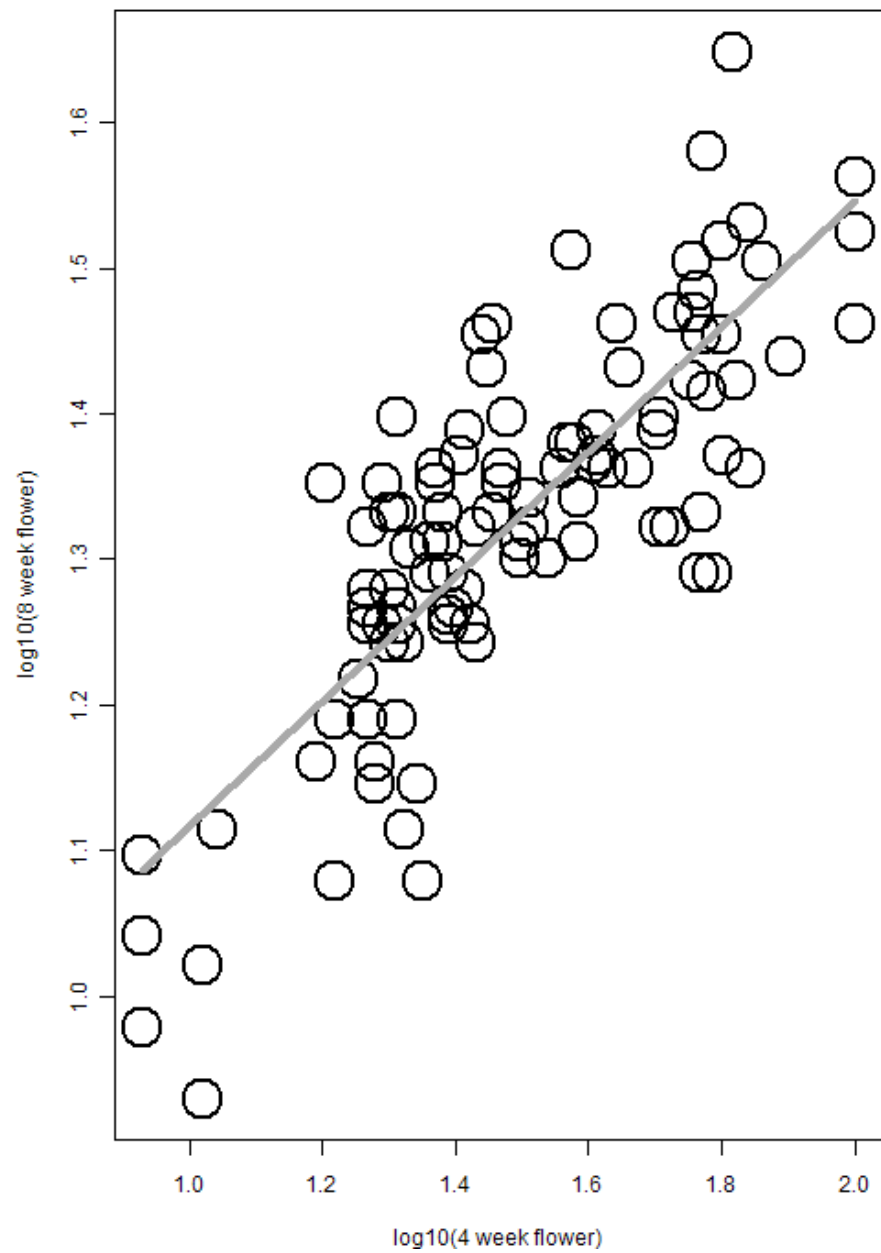
Korol et al. (2001)

Brassica napus: 2 correlated traits

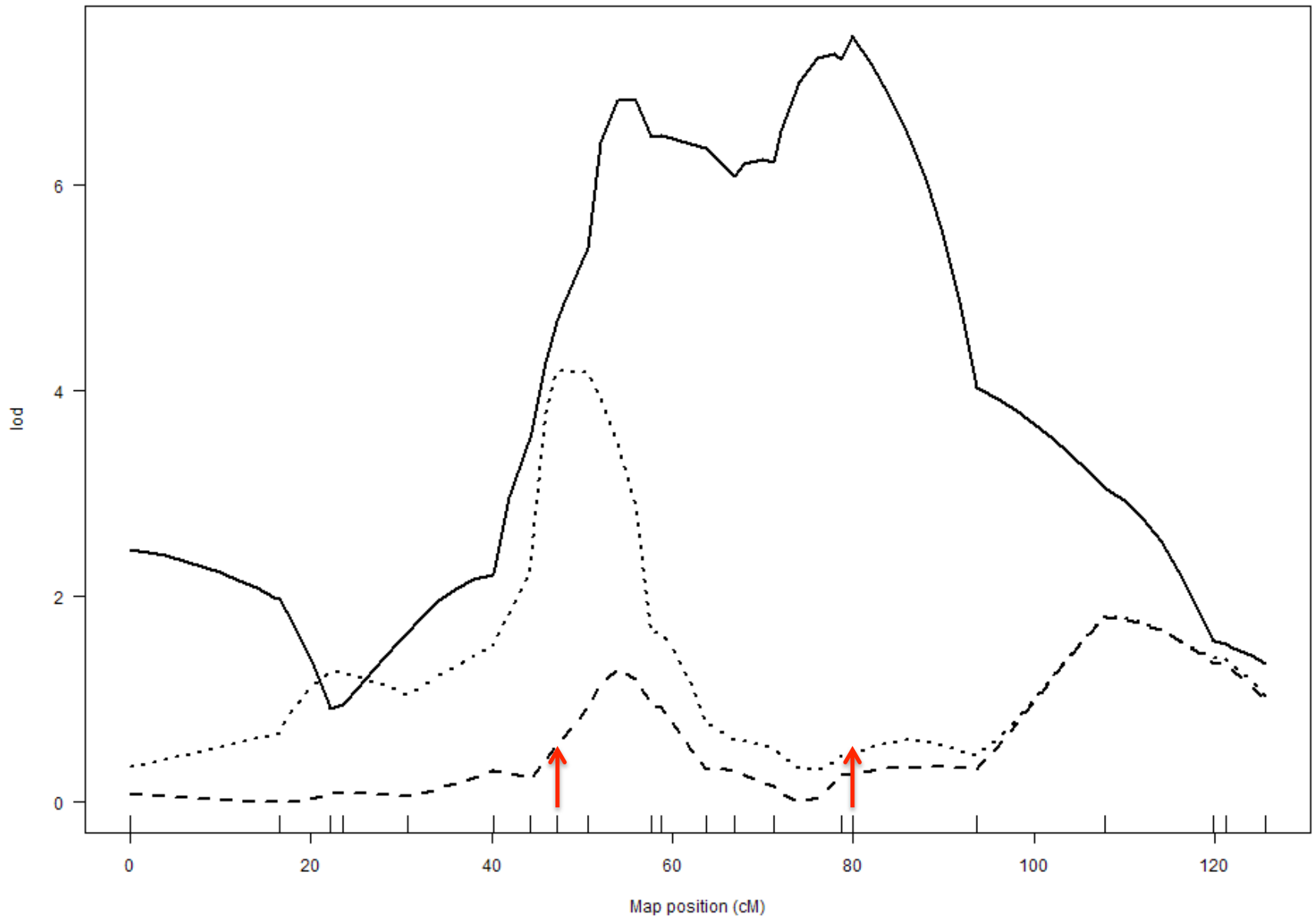
- 4-week & 8-week vernalization effect
 - log(days to flower)
- genetic cross of
 - Stellar (annual canola)
 - Major (biennial rapeseed)
- 105 F1-derived double haploid (DH) lines
 - homozygous at every locus (*QQ* or *qq*)
- 10 molecular markers (RFLPs) on LG9
 - two QTLs inferred on LG9 (now chromosome N2)
 - corroborated by Butruille (1998)
 - exploiting synteny with *Arabidopsis thaliana*

QTL with GxE or Covariates

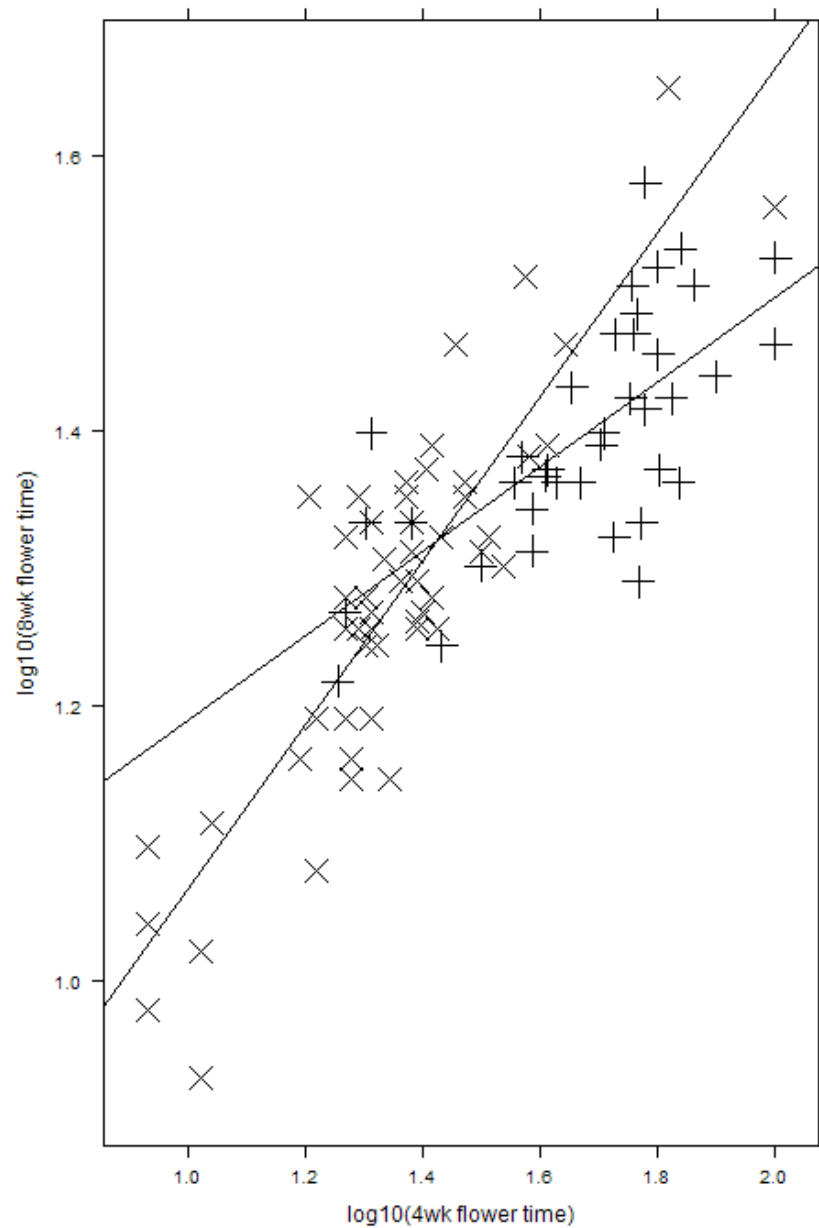
- adjust phenotype by covariate
 - covariate(s) = environment(s) or other trait(s)
- additive covariate
 - covariate adjustment same across genotypes
 - “usual” analysis of covariance (ANCOVA)
- interacting covariate
 - address GxE
 - capture genotype-specific relationship among traits
- another way to think of multiple trait analysis
 - examine single phenotype adjusted for others



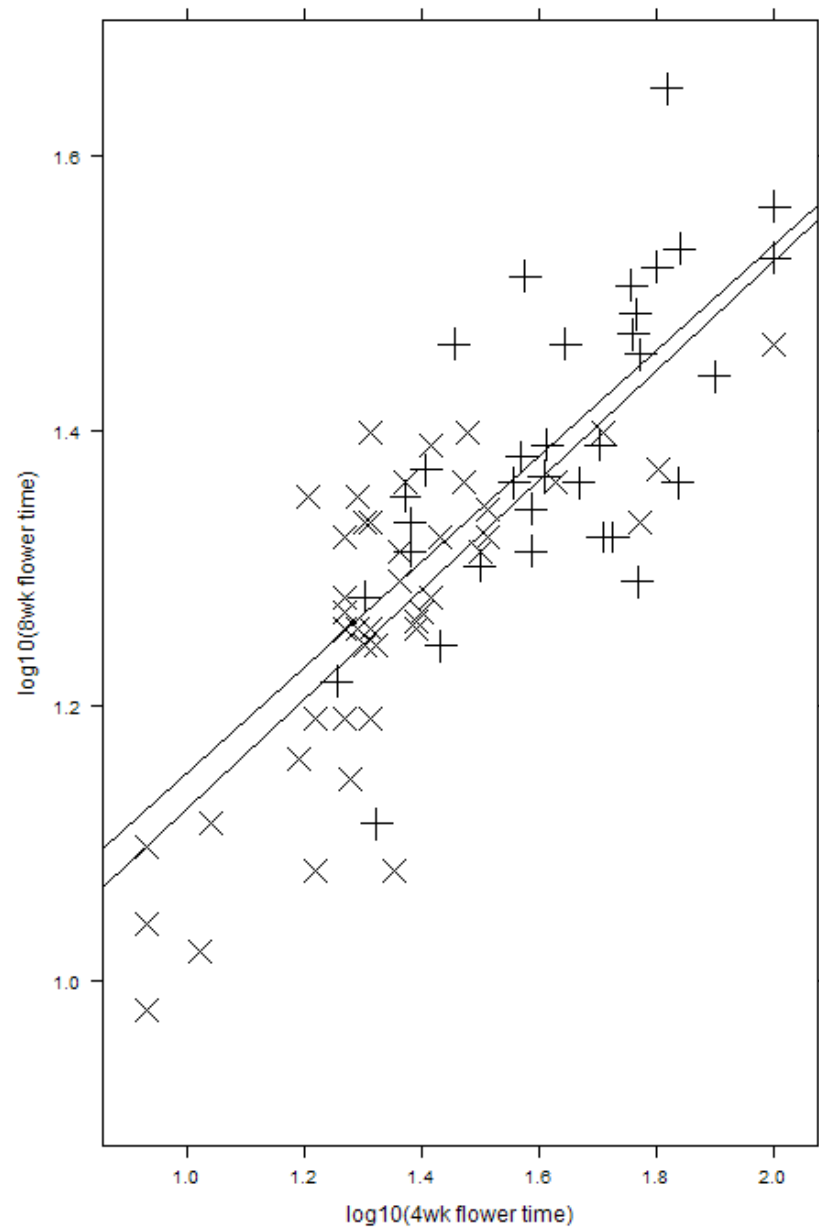
solid = 8wk, dashed = addcov, dotted = intcov



marker at 47cM



marker at 80cM



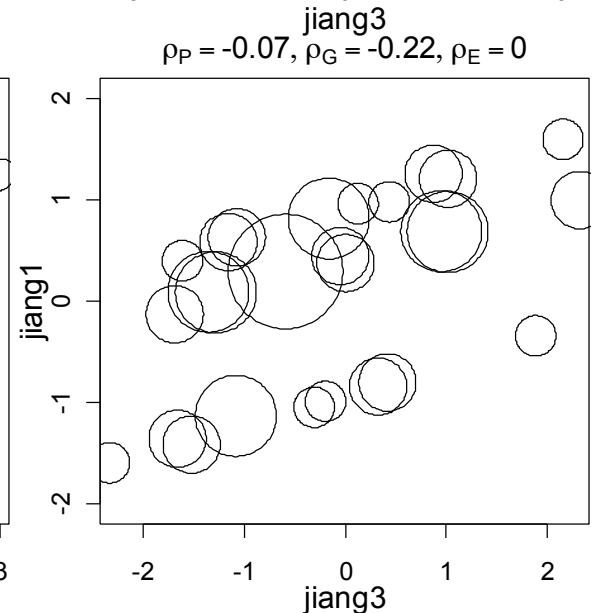
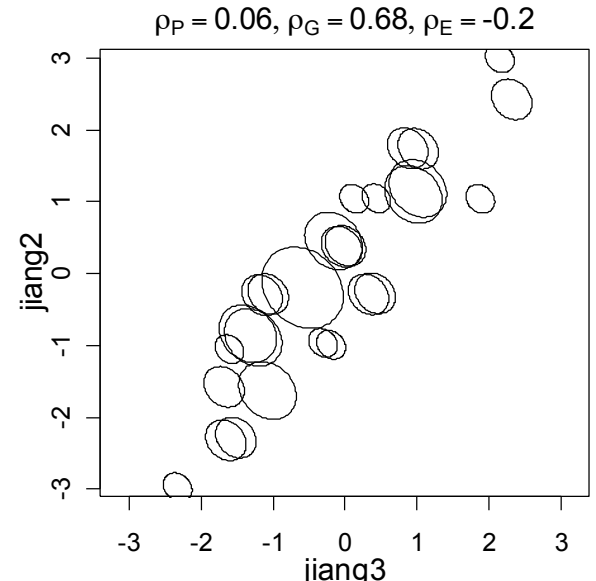
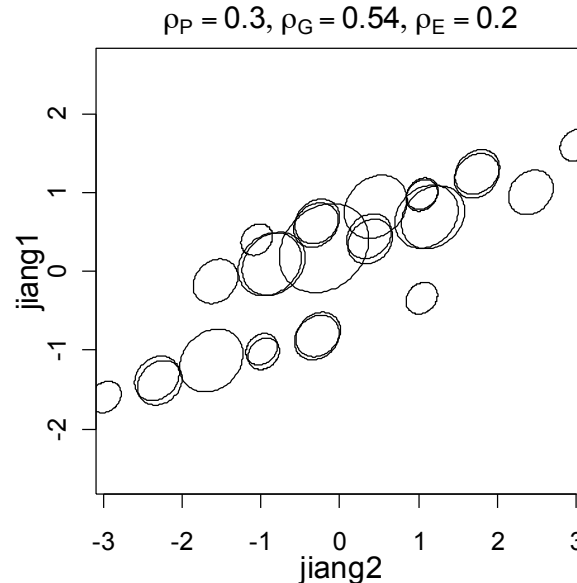
Multiple trait mapping

- Joint mapping of QTL
 - testing and estimating QTL affecting multiple traits
- Testing pleiotropy vs. close linkage
 - One QTL or two closely linked QTLs
- Testing QTL x environment interaction
- Comprehensive model of multiple traits
 - Separate genetic & environmental correlation

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)

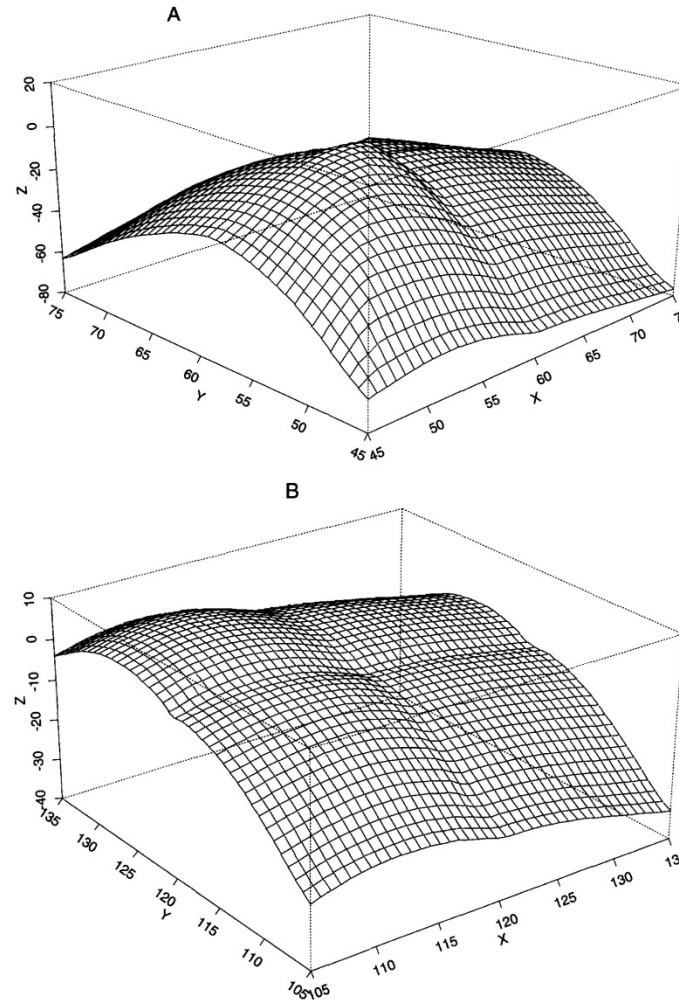
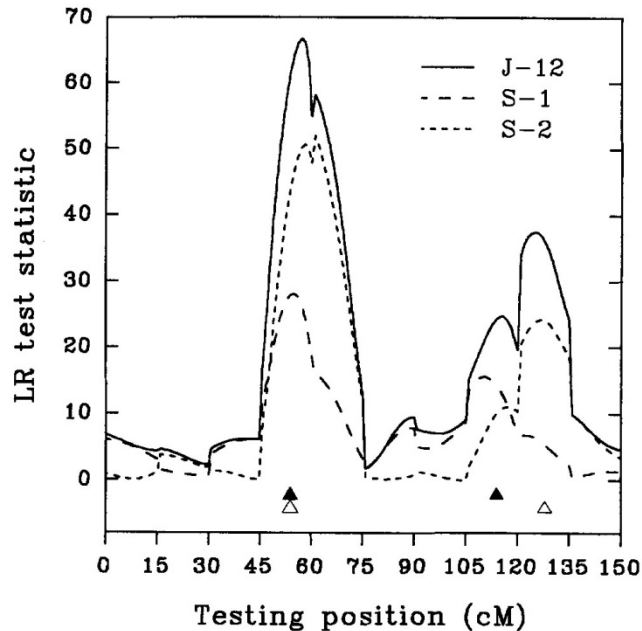


FIGURE 2.—Two-dimensional log-likelihood surfaces (expressed as deviations from the maximum of the log-likelihoods on the diagonal) for the test of pleiotropy *vs.* close linkage are presented for two regions: the region between 45 and 75 cM of Figure 1 (A) and the region between 105 and 135 cM (B).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

Formal Tests: 2 traits

$y_1 \sim N(\mu_{q1}, \sigma^2)$ for group 1 with QTL at location λ_1

$y_2 \sim N(\mu_{q2}, \sigma^2)$ for group 2 with QTL at location λ_2

- Pleiotropy vs. close linkage
 - test QTL at same location: $\lambda_1 = \lambda_2$
 - likelihood ratio test (LOD): null forces same location
- if pleiotropic ($\lambda_1 = \lambda_2$)
 - test for same mean: $\mu_{q1} = \mu_{q2}$
 - Likelihood ratio test (LOD)
 - null forces same mean, location
 - alternative forces same location
 - only make sense if traits are on same scale
 - test sex or location effect

More detail for 2 traits

$y_1 \sim N(\mu_{q1}, \sigma^2)$ for group 1

$y_2 \sim N(\mu_{q2}, \sigma^2)$ for group 2

- two possible QTLs at locations λ_1 and λ_2
- effect β_{kj} in group k for QTL at location λ_j

$$\mu_{q1} = \mu_1 + \beta_{11}(q_1) + \beta_{12}(q_2)$$

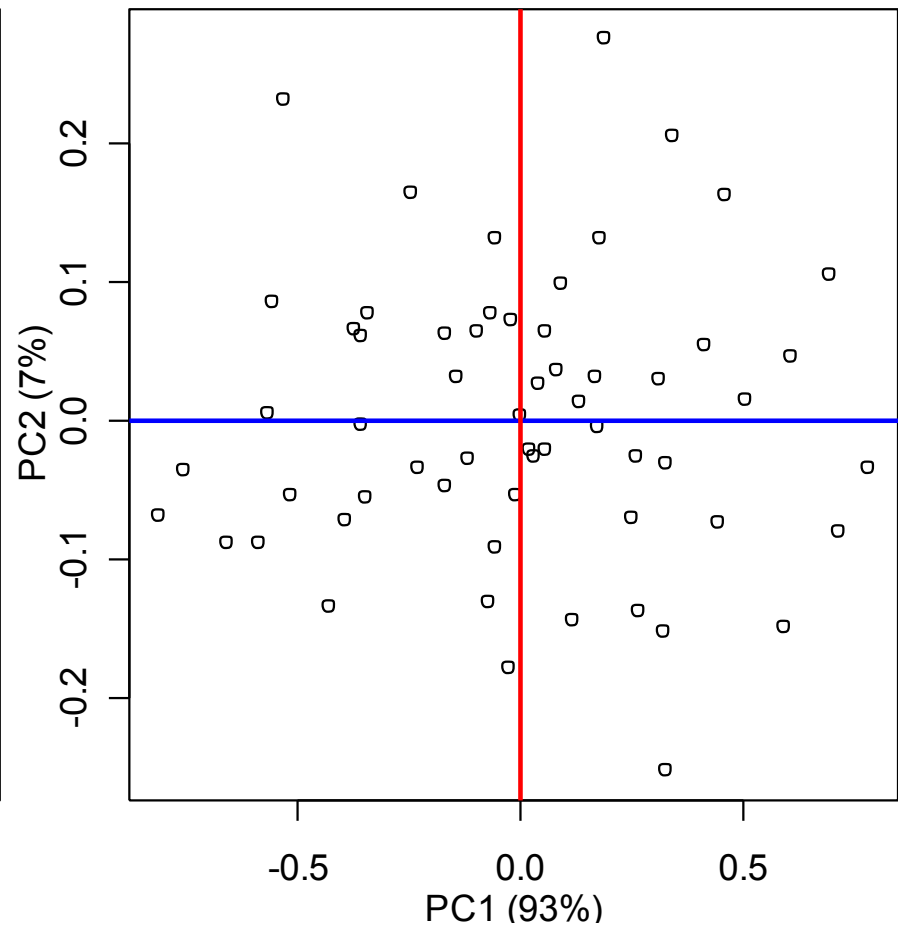
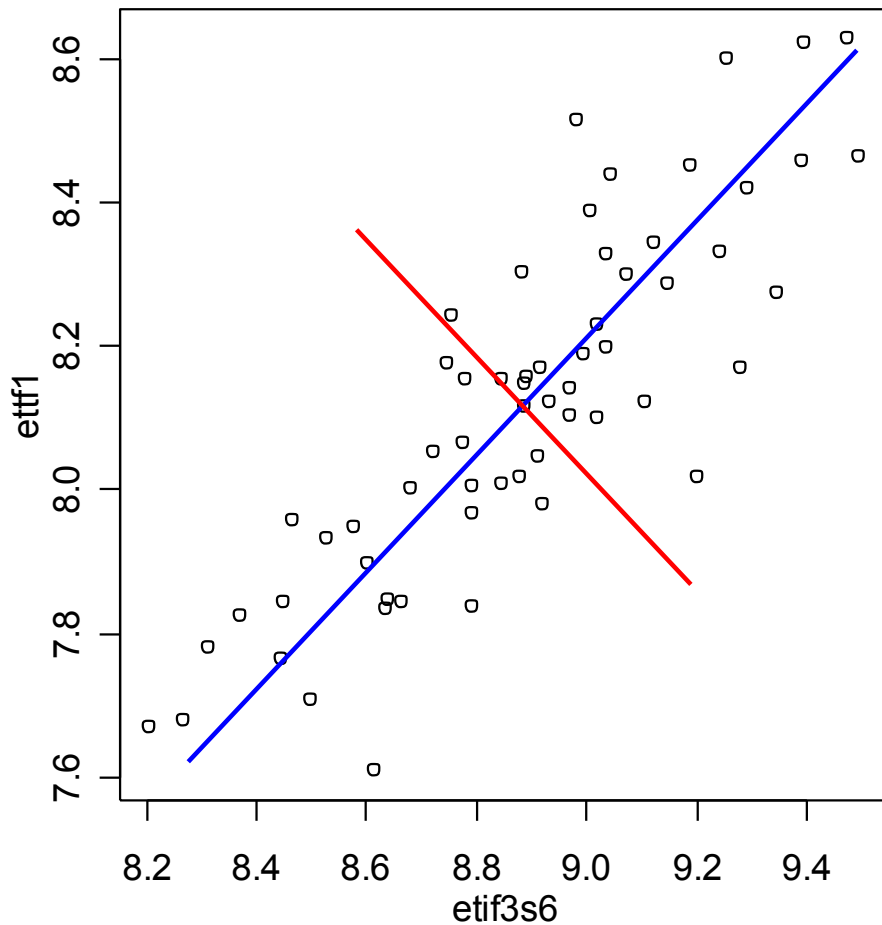
$$\mu_{q2} = \mu_2 + \beta_{21}(q_1) + \beta_{22}(q_2)$$

- classical: test $\beta_{kj} = 0$ for various combinations

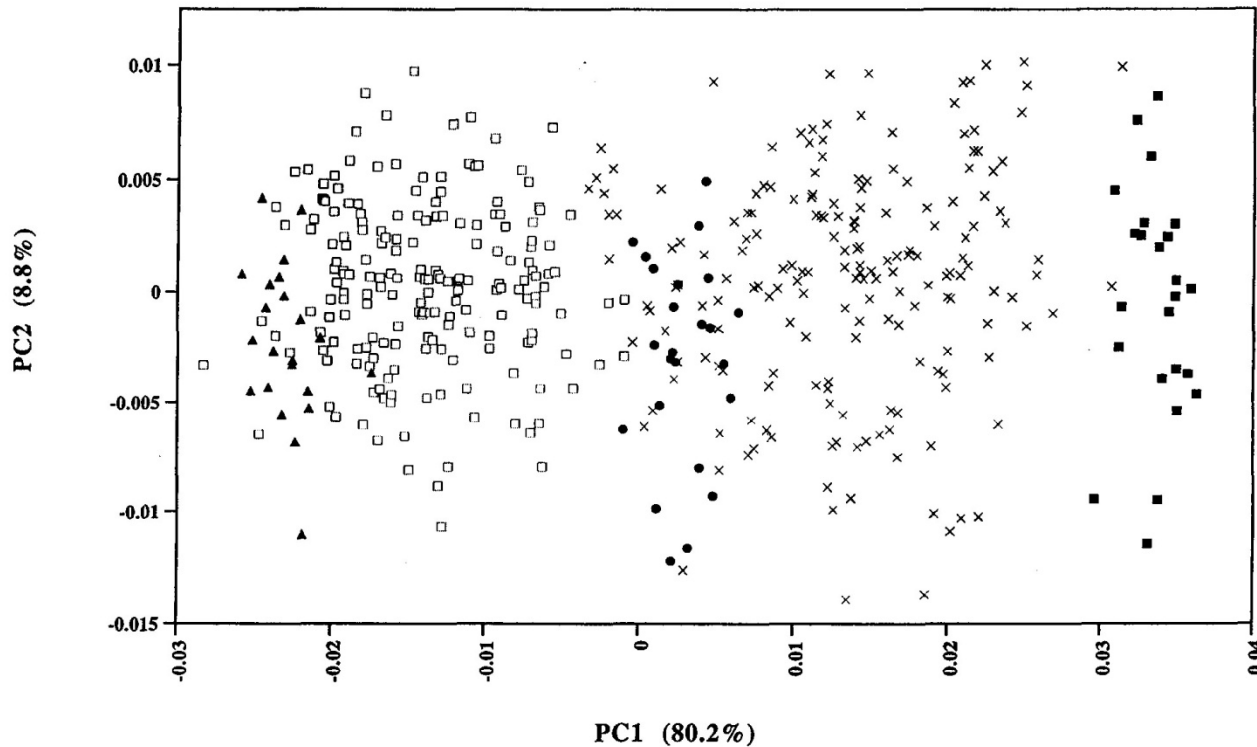
reducing many phenotypes to 1

- *Drosophila mauritiana* x *D. simulans*
 - reciprocal backcrosses, ~500 per bc
- response is “shape” of reproductive piece
 - trace edge, convert to Fourier series
 - reduce dimension: first principal component
- many linked loci
 - brief comparison of CIM, MIM, BIM

PC for two correlated phenotypes



shape phenotype via PC



■ *simulans* ▲ *mauritiana* • F1
 □ *mauritiana* backcross × *simulans* backcross

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*

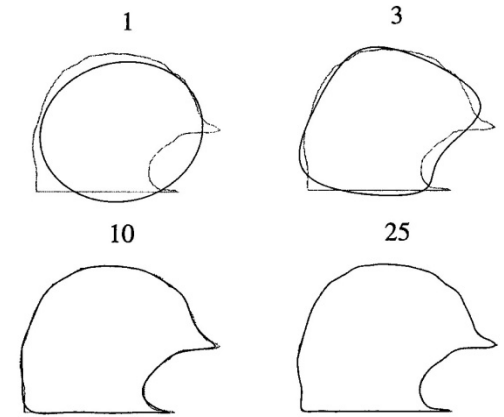


FIGURE 2.—The effect of harmonic number on the accuracy of reconstruction of a posterior lobe outline by elliptical Fourier analysis.

shape phenotype in BC study indexed by PC1

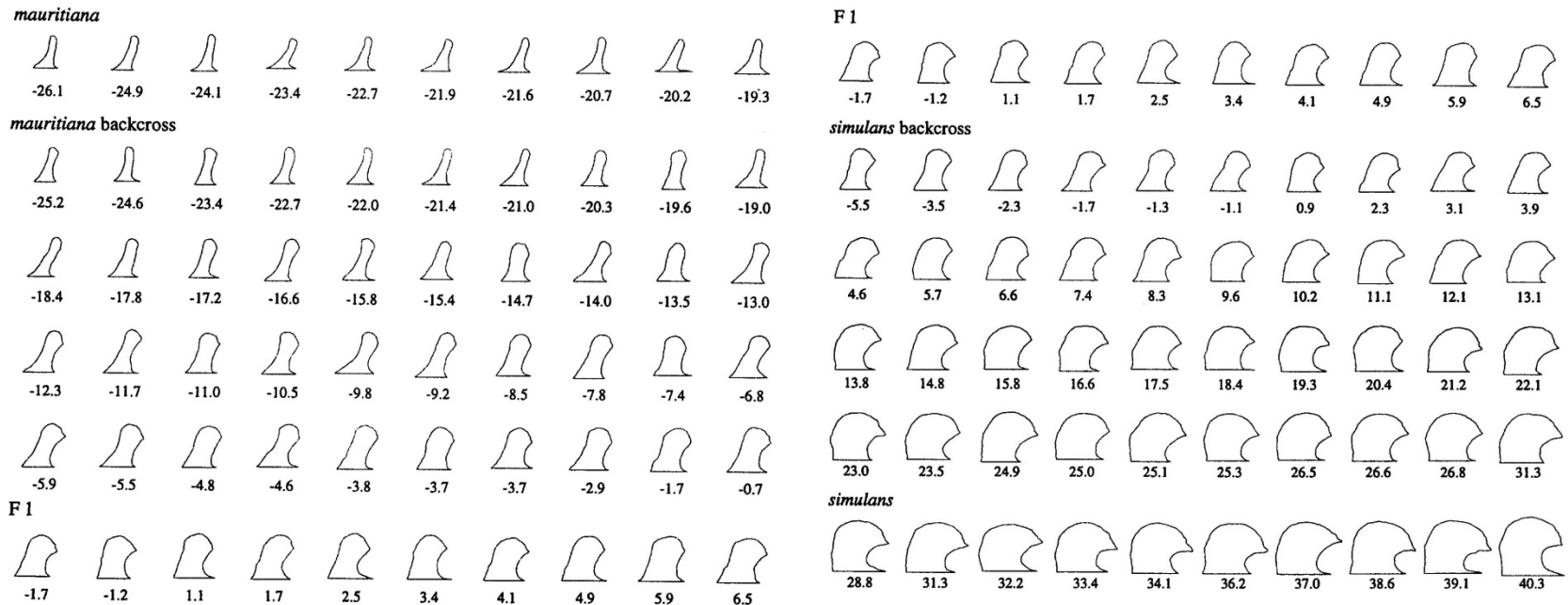


FIGURE 6.—Outlines of the posterior lobe from a sample of individuals from each of the five groups: pure *mauritiana*, *mauritiana* backcross, F₁, *simulans* backcross, and pure *simulans*. Within each group, the outlines are presented in order of their PC1 score (sampled at even intervals from the range of variation). The number below each specimen is its PC1 score. The outlines are drawn to scale with the origin at the centroid of each outline and with all baselines parallel.

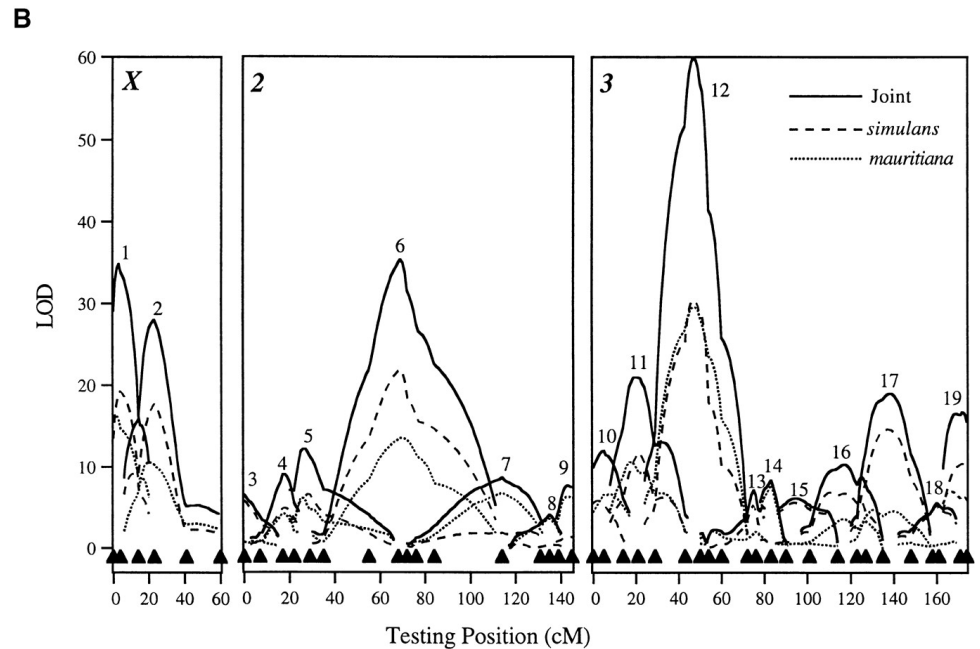
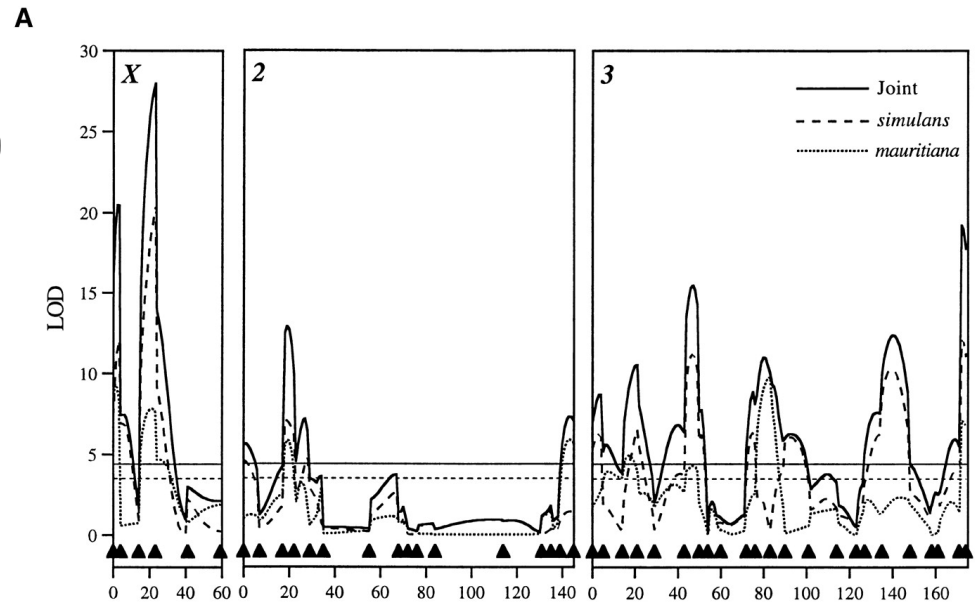
Liu et al. (1996) *Genetics*

Zeng et al. (2000) CIM vs. MIM

composite interval mapping
(Liu et al. 1996)
narrow peaks
miss some QTL

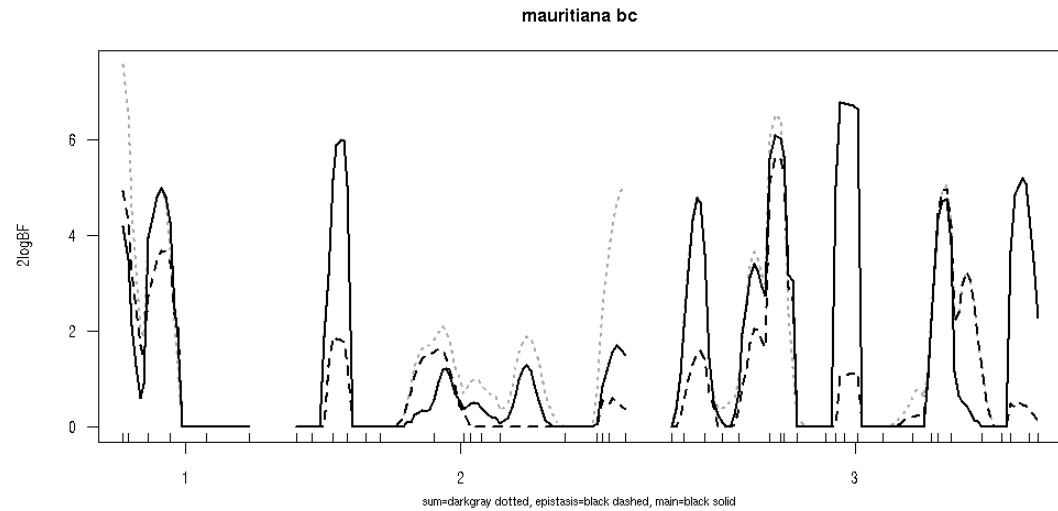
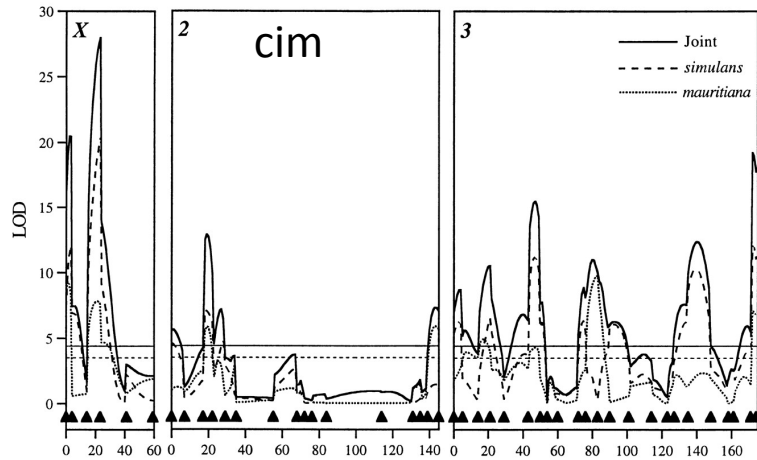
multiple interval mapping
(Zeng et al. 2000)
triangular peaks

both conditional 1-D scans
fixing all other "QTL"

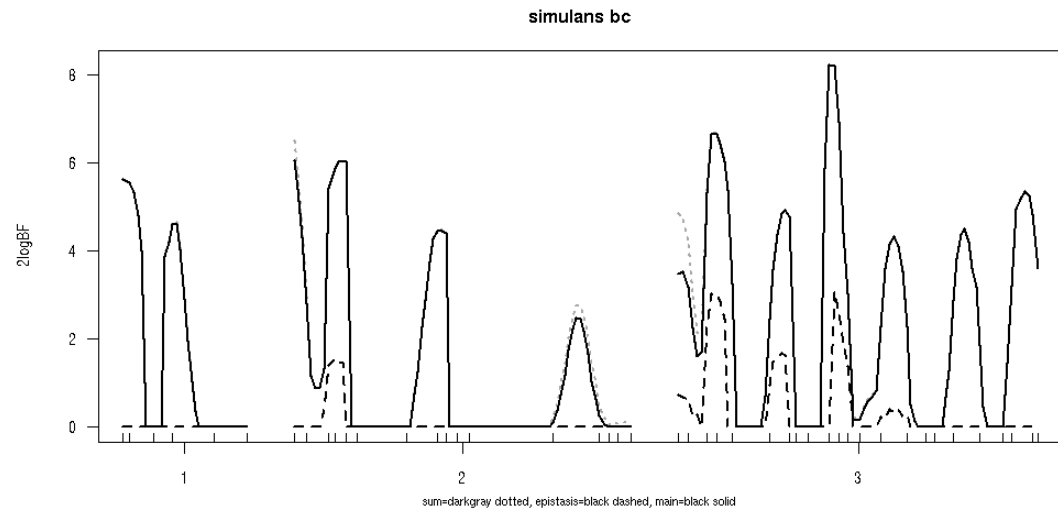
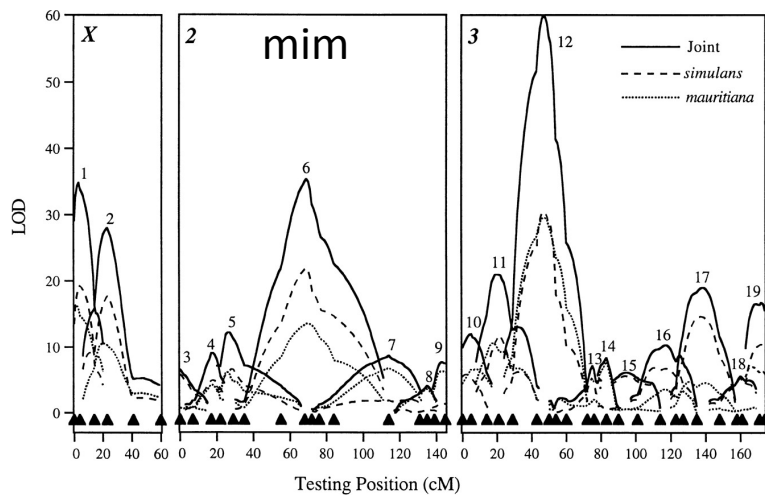


multiple QTL: CIM, MIM and BIM

A



B



Multi-trait strategy

- Use some data-reduction method
 - Principal components or clustering
 - WGCNA modules
 - Gene-mapping Hotspots
 - Functional/pathway information
- Deeper look at high priority groups of traits
 - Gene set enrichment
 - Correlation with key clinical traits

Multi-trait strategy (cont.)

- Use genetics to fine-tune search
 - Test Causal pairs (Elias talk, Mark on Tuesday)
 - Screen out genes based on IBD, SNP effects
- Identify small subset (5-15) for networks
 - Infer causal network (Elias talk)
 - Use biological pathway information (PPI, TF, ...)
 - Avoid hairballs (but hubs may be useful)
- Validation tests of key drivers: new study ...

how to use functional information?

- functional grouping from prior studies
 - may or may not indicate direction
 - gene ontology (GO), KEGG
 - knockout (KO) panels
 - protein-protein interaction (PPI) database
 - transcription factor (TF) database
- methods using only this information
- priors for QTL-driven causal networks
 - more weight to local (*cis*) QTLs?

Thanks!

- Grant support
 - NIH/NIDDK 58037, 66369
 - NIH/NIGMS 74244, 69430
 - NCI/ICBP U54-CA149237
 - NIH/R01MH090948
- Collaborators on papers and ideas
 - Alan Attie & Mark Keller, Biochemistry
 - Karl Broman, Aimee Broman, Christina Kendziorski