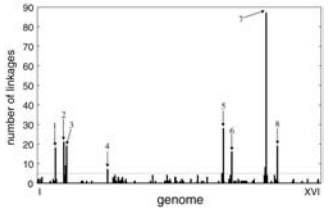
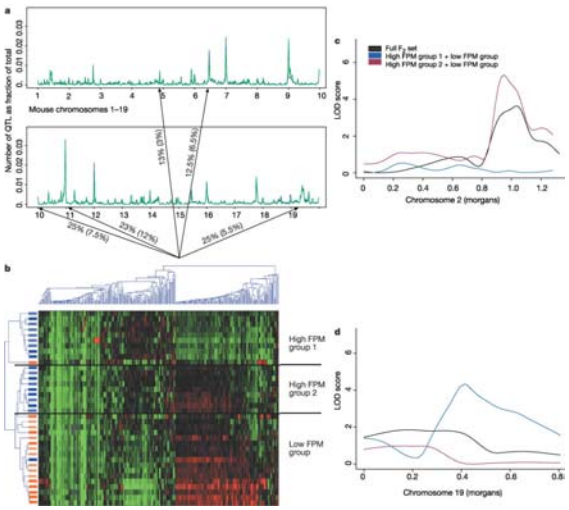
4. modern high throughput biology

- measuring the molecular dogma of biology
  - DNA → RNA → protein → metabolites
  - measured one at a time only a few years ago
- massive array of measurements on whole systems ("omics")
  - thousands measured per individual (experimental unit)
  - all (or most) components of system measured simultaneously
    - whole genome of DNA: genes, promoters, etc.
    - all expressed RNA in a tissue or cell
    - all proteins
    - all metabolites
- systems biology: focus on network interconnections
  - chains of behavior in ecological community
  - underlying biochemical pathways
- genetics as one experimental tool
  - perturb system by creating new experimental cross
  - each individual is a unique mosaic

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



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



finding heritable traits
(from Christina Kendziorski)

- reduce 30,000 traits to 300-3,000 heritable traits
- probability a trait is heritable

  $\text{pr}(H|Y,Q) = \text{pr}(Y|Q,H)\ \text{pr}(H|Q) / \text{pr}(Y|Q)$ Bayes rule

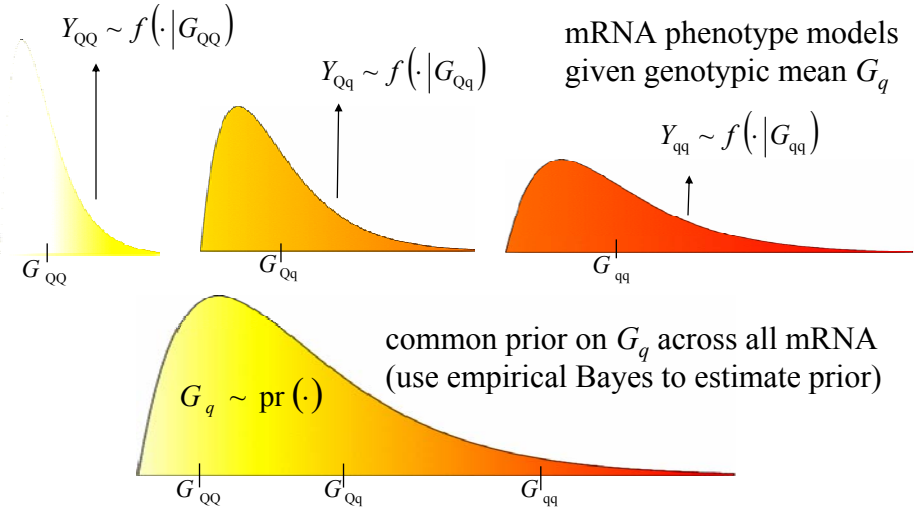
  $\text{pr}(Y|Q) = \text{pr}(Y|Q,H)\ \text{pr}(H|Q) + \text{pr}(Y|Q, \text{not } H)\ \text{pr}(\text{not } H|Q)$
- phenotype averaged over genotypic mean μ

  $\text{pr}(Y|Q, \text{not } H) = f_0(Y) = \int f(Y|G)\ \text{pr}(G)\ dG$ if not H

  $\text{pr}(Y|Q, H) = f_1(Y|Q) = \prod_q f_0(Y_q)$ if heritable

  $Y_q = \{Y_i \mid Q_i = q\}$ = trait values with genotype $Q=q$

hierarchical model for expression phenotypes
(EB arrays: Christina Kendziorski)

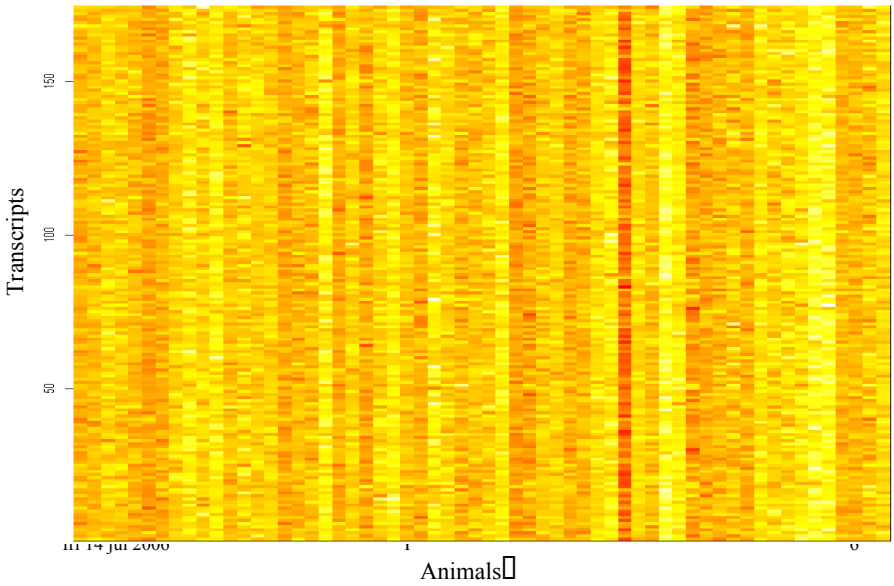


Improving the power for eQTL mapping (Chen et al. 2005)

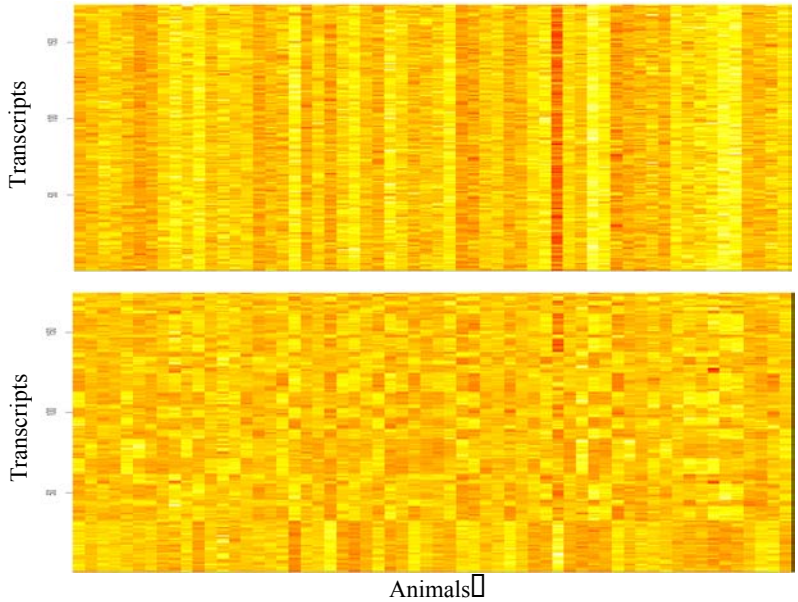
- Consider 6000 transcripts that show heritability (as seeds)
- For each seed, construct list of correlated transcripts ($\rho > 0.7$)
- 1300 lists are enriched for at least one GO term

	G1	G2	G3	…	G860
L1	x		x		
L2		x	x		
L3		x			x
⋮					
L1300	x		x	x	x

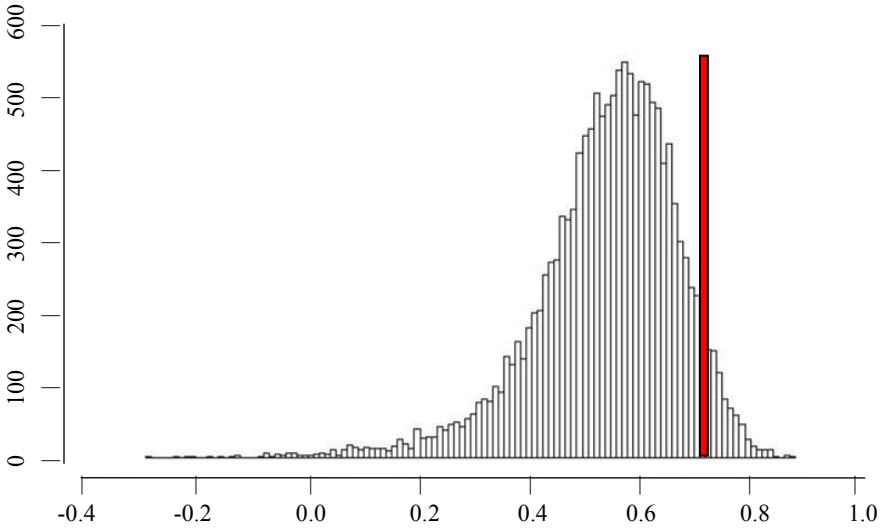
184 transcripts from 31 lists in Lipid Metabolism GO category



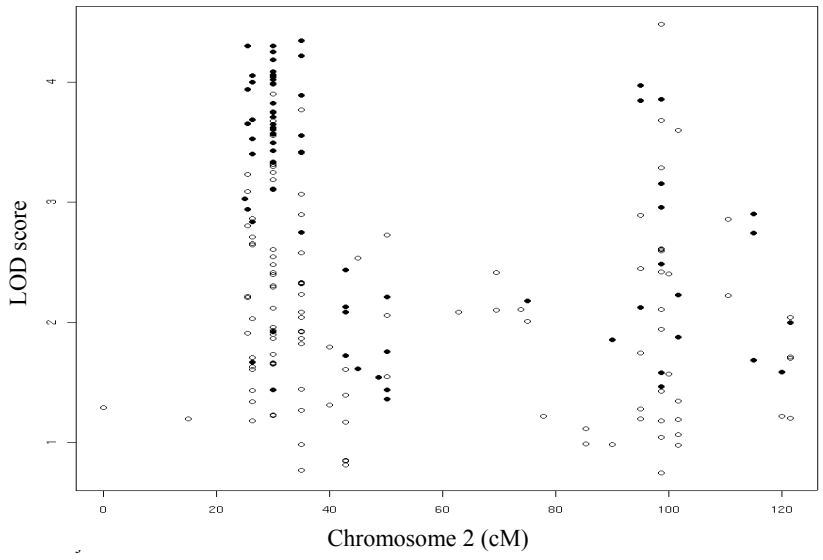
Lipid Metabolism vs. Control



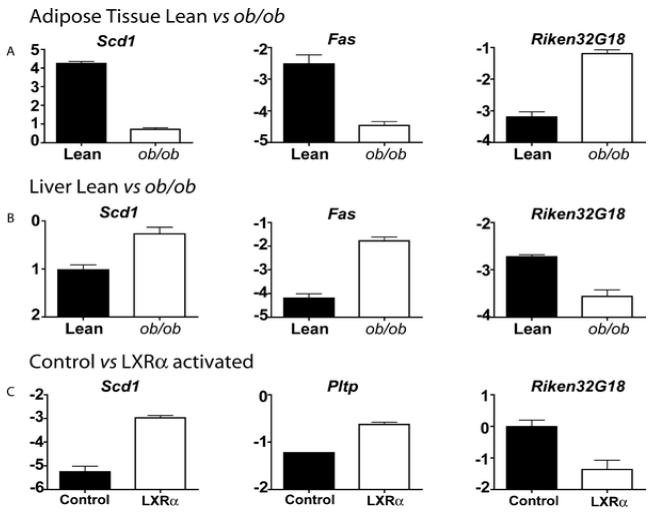
Pairwise correlation of traits from different lipid metabolism lists



LOD scores for lipid metabolism transcripts



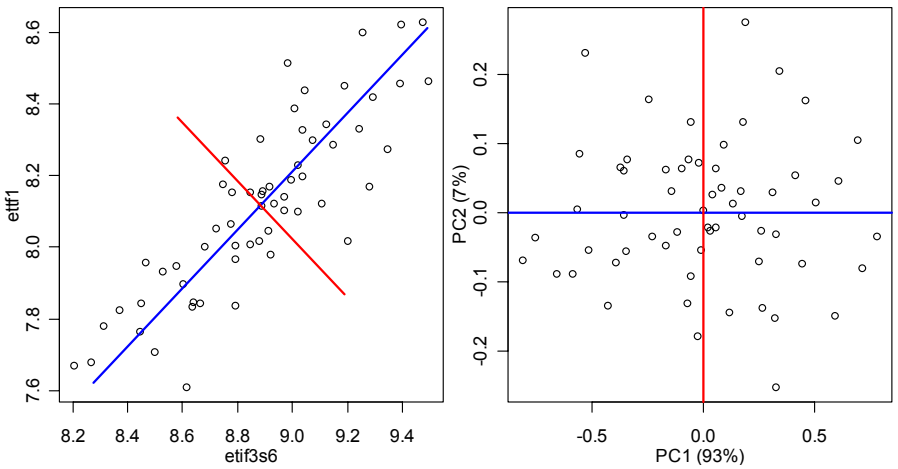
Validation of Riken32G18



expression meta-traits: pleiotropy

- reduce 3,000 heritable traits to 3 meta-traits(!)
- what are expression meta-traits?
  - pleiotropy: a few genes can affect many traits
    - transcription factors, regulators
  - weighted averages: $Z = YW$
    - principle components, discriminant analysis
- infer genetic architecture of meta-traits
  - model selection issues are subtle
    - missing data, non-linear search
    - what is the best criterion for model selection?
  - time consuming process
    - heavy computation load for many traits
    - subjective judgement on what is best

PC for two correlated mRNA



PC across microarray functional groups

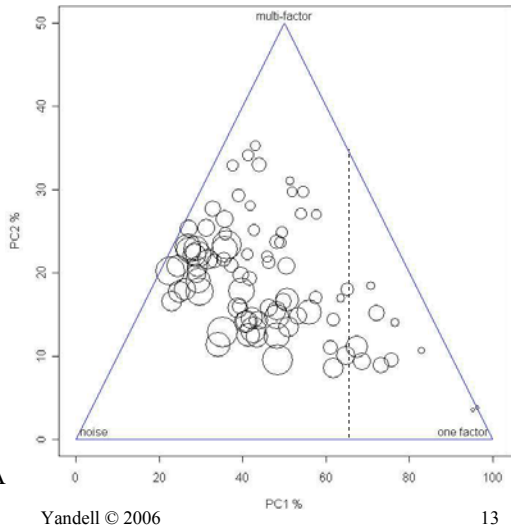
Affy chips on 60 mice
~40,000 mRNA

2500+ mRNA show DE
(via EB arrays with
marker regression)

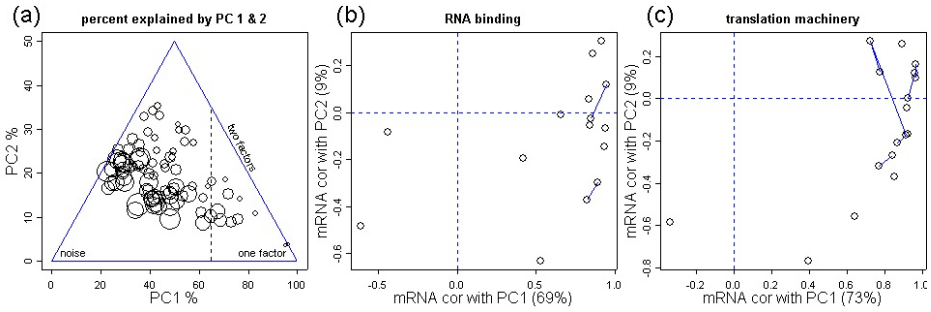
1500+ organized in
85 functional groups
2-35 mRNA / group

which are interesting?
examine PC1, PC2

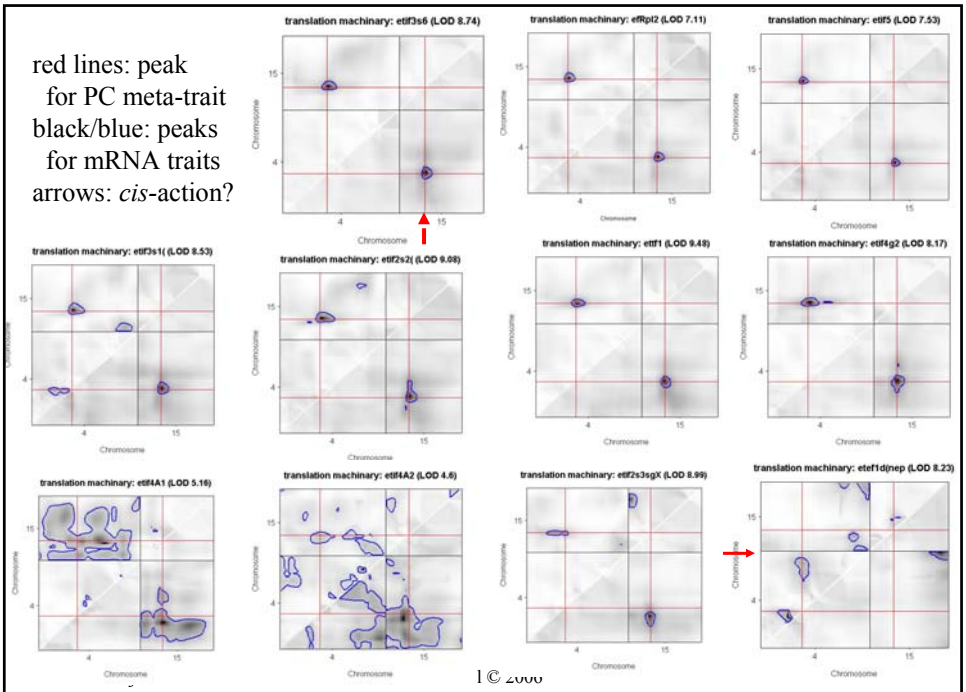
circle size = # unique mRNA



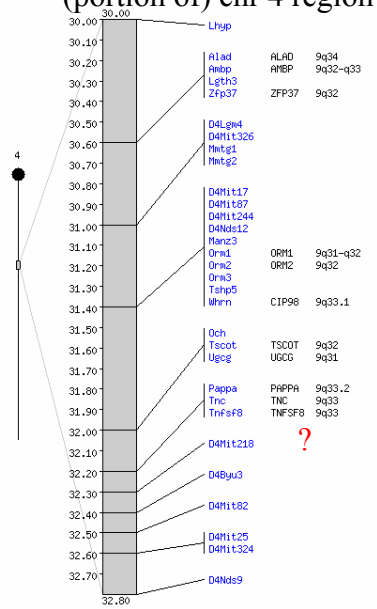
84 PC meta-traits by functional group
focus on 2 interesting groups



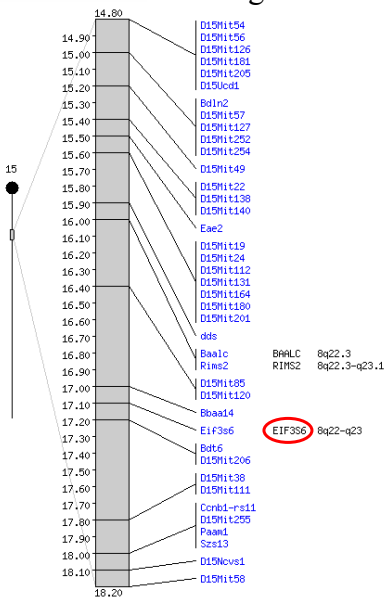
red lines: peak
for PC meta-trait
black/blue: peaks
for mRNA traits
arrows: *cis*-action?



(portion of) chr 4 region



chr 15 region



interaction plots for DA meta-traits

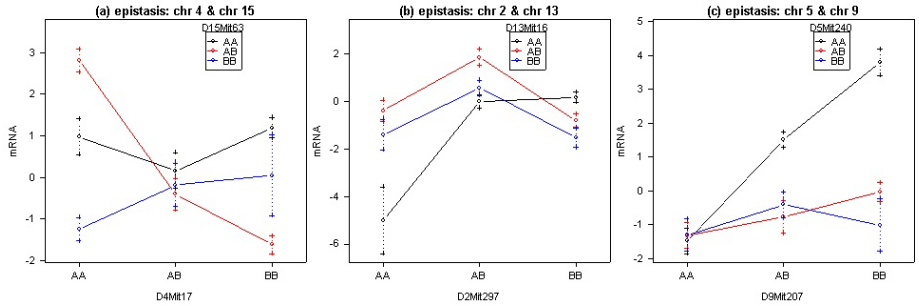
DA for all pairs of markers:

separate 9 genotypes based on markers

(a) same locus pair found with PC meta-traits

(b) Chr 2 region interesting from biochemistry (Jessica Byers)

(c) Chr 5 & Chr 9 identified as important for insulin, SCD

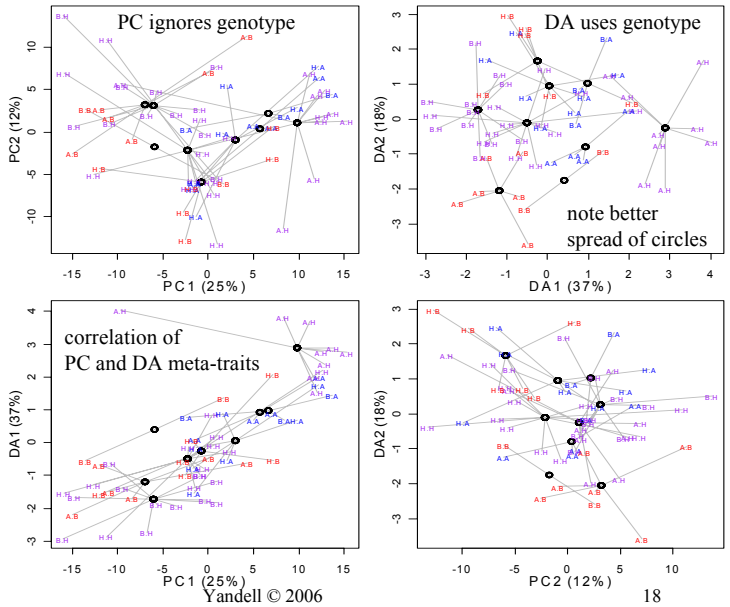


comparison of PC and DA meta-traits on 1500+ mRNA traits

genotypes from Chr 4/Chr 15 locus pair (circle=centroid)

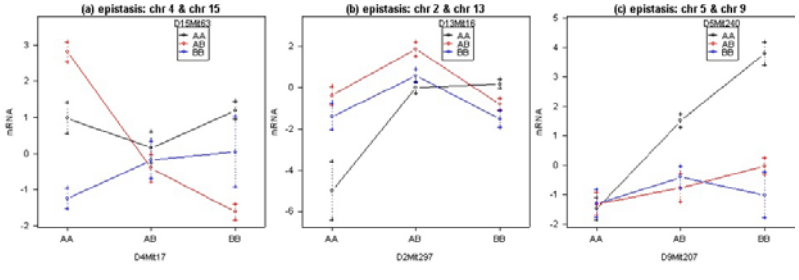
PC captures spread without genotype

DA creates best separation by genotype

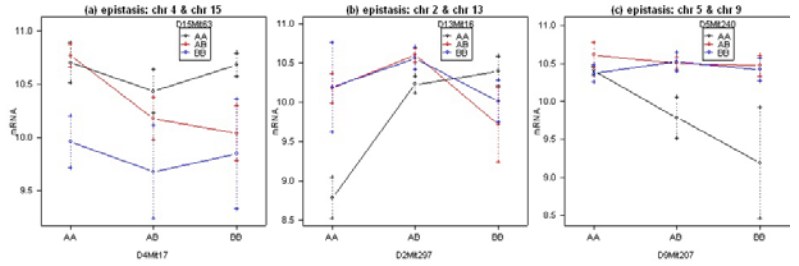


relating meta-traits to mRNA traits

DA meta-trait standard units

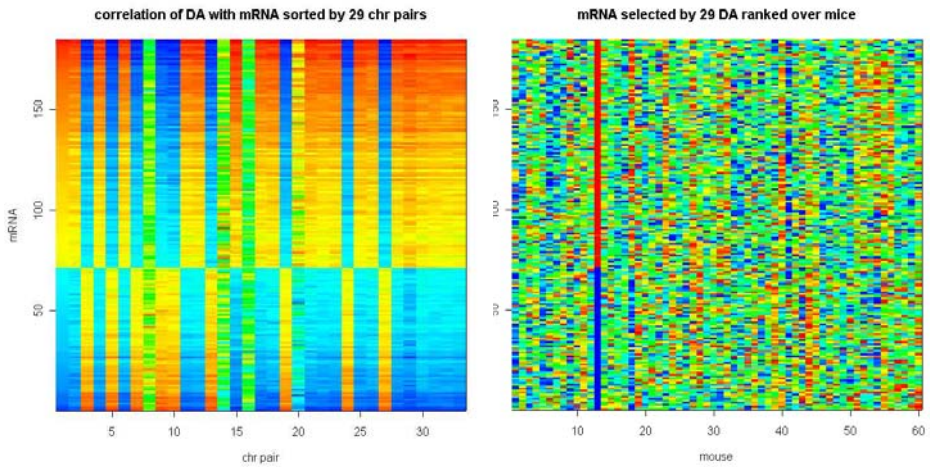


SCD trait log2 expression



DA: a cautionary tale

(184 mRNA with |cor| > 0.5; mouse 13 drives heritability)



building graphical models

- infer genetic architecture of meta-trait
  - $E(Z \mid Q, M) = \mu_q = \beta_0 + \sum_{\{q \text{ in } M\}} \beta_{qk}$
- find mRNA traits correlated with meta-trait
  - $Z \approx \underline{YW}$ for modest number of traits $\underline{Y}$
- extend meta-trait genetic architecture
  - $\underline{M}$ = genetic architecture for $\underline{Y}$
  - expect subset of QTL to affect each mRNA
  - may be additional QTL for some mRNA

posterior for graphical models

•posterior for graph given multivariate trait & architecture

$\text{pr}(G \mid \underline{Y}, Q, \underline{M}) = \text{pr}(\underline{Y} \mid Q, G)\ \text{pr}(G \mid \underline{M}) / \text{pr}(\underline{Y} \mid Q)$

–$\text{pr}(G \mid \underline{M})$ = prior on valid graphs given architecture

•multivariate phenotype averaged over genotypic mean μ

$\text{pr}(\underline{Y} \mid Q, G) = f_1(\underline{Y} \mid Q, G) = \prod q\ f_0(\underline{Y}q \mid G)$

$f_0(\underline{Y}_q \mid G) = \int f(\underline{Y}_q \mid \underline{\mu}, G)\ \text{pr}(\underline{\mu})\ \text{d}\underline{\mu}$

•graphical model G implies correlation structure on $\underline{Y}$

•genotype mean prior assumed independent across traits

$\text{pr}(\underline{\mu}) = \prod_t \text{pr}(\mu_t)$

from graphical models to pathways

- build graphical models

  QTL $\rightarrow$ RNA1 $\rightarrow$ RNA2

  – class of possible models

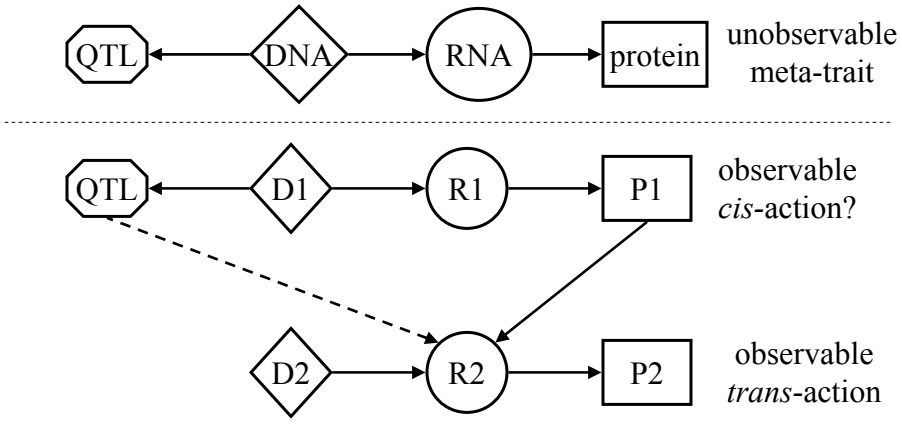
  – best model = putative biochemical pathway
- parallel biochemical investigation

  – candidate genes in QTL regions

  – laboratory experiments on pathway components

graphical models (with Elias Chaibub)

$$f_1(\underline{Y} \mid Q, G=g) = f_1(Y_1 \mid Q)\ f_1(Y_2 \mid Q, Y_1)$$



summary

- expression QTL are complicated
  - need to consider multiple interacting QTL
- coherent approach for high-throughput traits
  - identify heritable traits
  - dimension reduction to meta-traits
  - mapping genetic architecture
  - extension via graphical models to networks
- many open questions
  - model selection
  - computation efficiency
  - inference on graphical models

references: eQTL

- Brem et al. Kruglyak (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science 296*: 752-755.
- Lan et al. Yandell, Attie (2003) Dimension reduction for mapping mRNA abundance as quantitative traits. *Genetics 164*: 1607-1614.
- Lan, Chen, Flowers, Yandell et al. Kendziorski, Attie (2006) Combined expression trait correlations and expression quantitative trait locus mapping. *PLoS Genet 2*: e6.
- Schadt, Monks et al. Friend (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature 422*: 297-302.
- Schadt et al. Lusis (2005) An integrative genomics

references: multiple traits

- Doerge (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet 3*: 43-52.
- Jiang, Zeng (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics 140*: 1111-1127.
- Korol et al. Nevo (2001) Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. *Genetics 157*: 1789-1803.
- Vieira, Pasyukova, Zeng et al. Mackay (2000) Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. *Genetics 154*: 213-227.

thanks

- USDA CSREES
- NIH/NIDDK, 5803701 and 66369-01
- ADA Innovation Award, 7-03-IG-01
- HHMI 133-ES29 (CMK)
- National Council for Scientific and Technological Development (CNPq), Brazil (ENC)