Unraveling complex traits through the genetics of gene expression

ASHG Annual Meeting 2002

Monks SA, Schadt EE (2002) Unraveling complex traits through the genetics of gene expression. Am J Hum Genet 71 (4): 2381

Issues in Identifying Genes for Complex Human Diseases

- Gene mapping has been extremely successful for simple Mendelian diseases; however, finding such genes for diseases, and their associated risk traits, that are of public health interest has proven difficult.

- Reasons for this difficulty include disease heterogeneity (disease sub-types with some or no overlapping genetic causes), misclassification (from using discrete classifications of disease from thresholds and combinations of thresholds), and cumulative environmental influences.

- With the advent of technology to measure changes in gene expression, i.e. changes in mRNA transcript abundance, it should be possible to unravel some of the complexity existing for these common diseases.

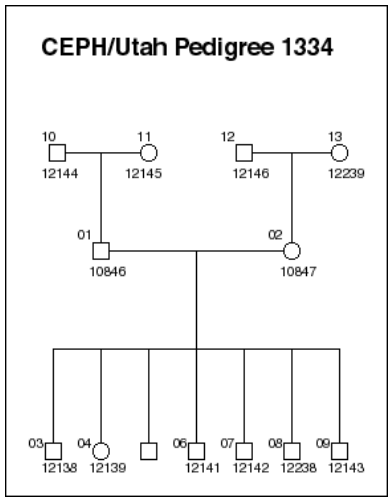
Goal of Present Study

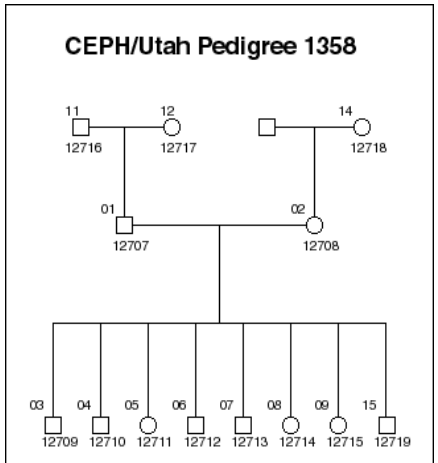
By incorporating genetic variation, patterns of genetic inheritance, measures of relevant environmental influences and gene expression components, it should be possible to dissect complex interacting pathways that lead to susceptibility to complex disease.

As an initial step in establishing the power of such a combined approach, **we study the genetic component of mRNA expression** through the use of a sample of CEPH (Centre d'Etude du Polymorphisme Humain) families.

Data: CEPH families

Representative families





Samples were based on 15 CEPH families: 1334, 1340, 1345, 1346, 1349, 1350, 1358, *1362, 1375, 1377, 1408*, 1418, 1421, 1424 and 1477

Expression profiling of lymphoblastoid cell lines was performed using a 25K human gene oligonucleotide microarray[1]

167 individuals were successfully profiled

Data: Measured Variables

Genotype data across 346 autosomal genetic markers for 210 of the family members

- Genetic markers were selected so that at least 75% of the pedigrees had genotypes available for at least 75% of the family
- Median inter-marker distance of 5.22 cM based on the deCODE map[5]

A total of 23,615 genes with expression data[2]

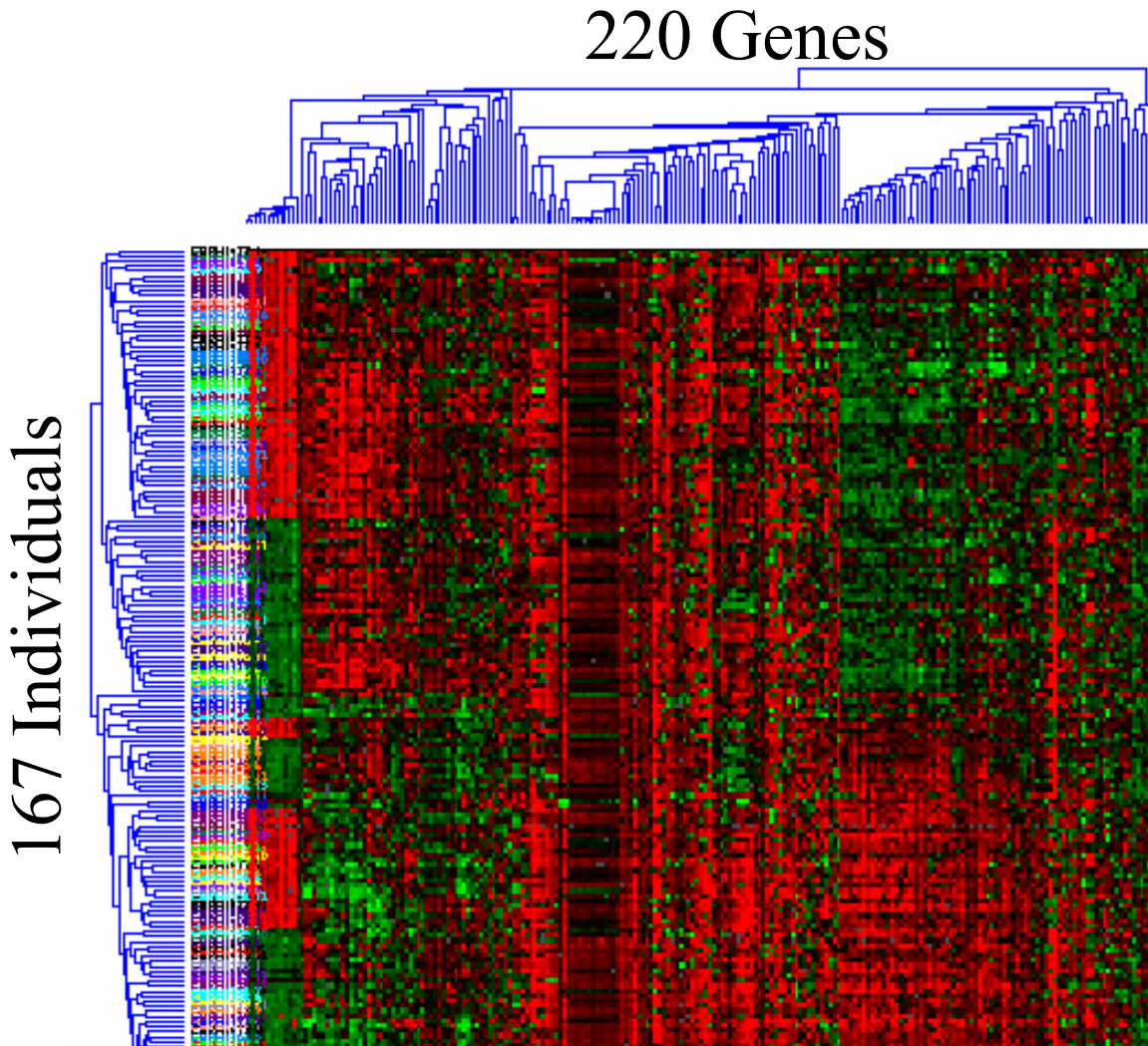
- Individuals were profiled relative to a pool consisting of all founder samples
- Expression was measured by the mean, over fluor-reverse pairs, of $\log_{10}$(expression ratio) where the ratio is between normalized, background-corrected intensity values for the two channels (red and green) for each spot on the array
- 2430 genes were differentially expressed (DE) [3] relative to the pool across more than half of the children at a Type I error rate of 0.05

Clustering of Genes and Individuals

Clustering of 220 genes that were differentially expressed relative to the pool in at least half of the children (p-value < 0.05)

Clustering of 167 individuals, color coding by family

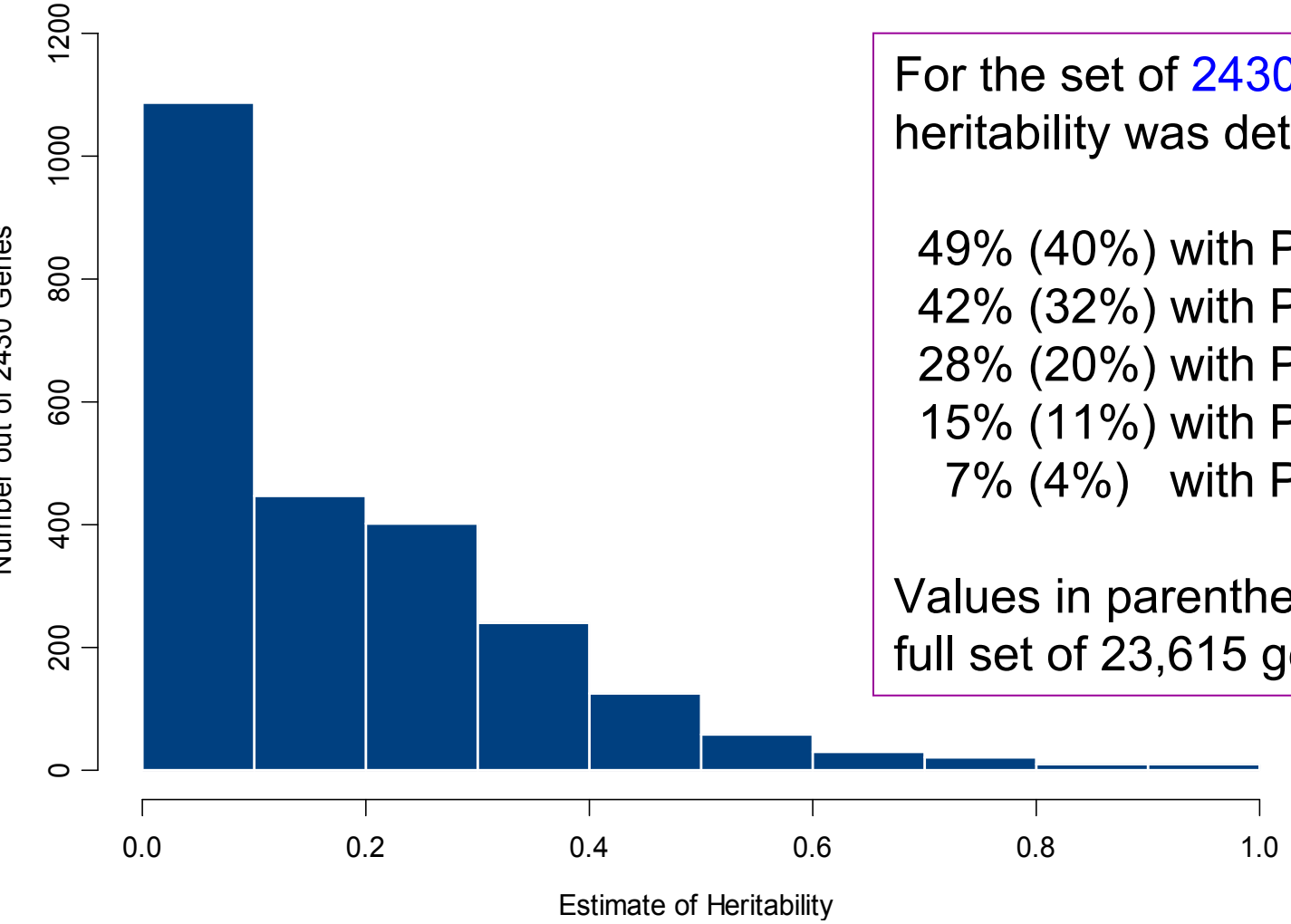
Based on highly expressed genes, clustering of individuals is not tightly familial



Statistical Methods: Overview

- For each gene, expression was treated as a quantitative trait.
- Variance-component linkage methods, as implemented in SOLAR, were utilized[4] for:
  - segregation analyses with additive and dominance variance components
  - multipoint linkage analysis (based on deCODE map estimates[5])
  - and bivariate segregation analysis[6]
  - All analyses were adjusted for age, gender and age*gender interaction.

Heritability* of Gene Expression



For the set of 2430 DE genes, heritability was detected for:

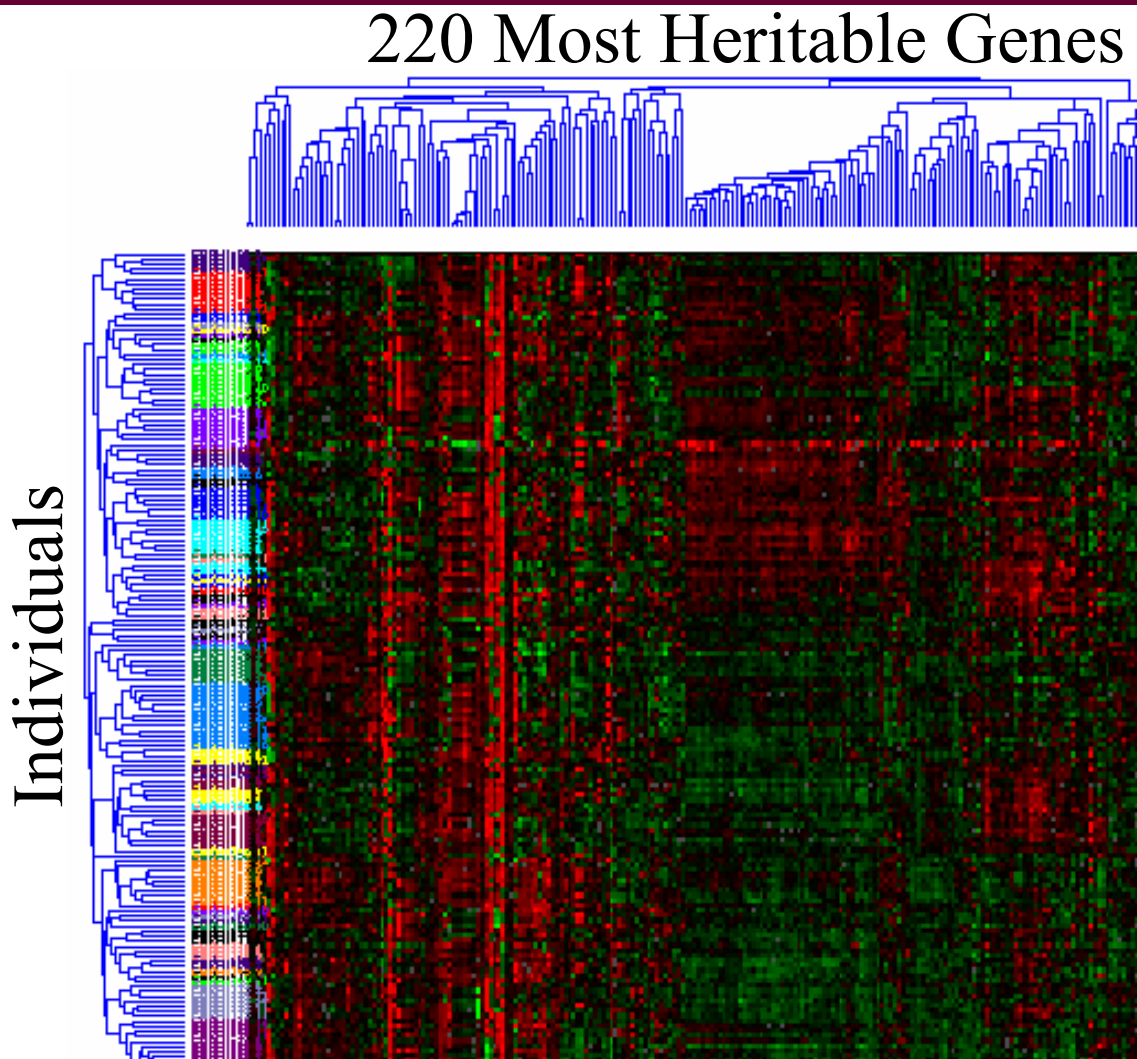
49% (40%) with P-value < 0.1
42% (32%) with P-value < 0.05
28% (20%) with P-value < 0.01
15% (11%) with P-value < 0.001
7% (4%) with P-value < 0.05/2430

Values in parentheses correspond to the full set of 23,615 genes.

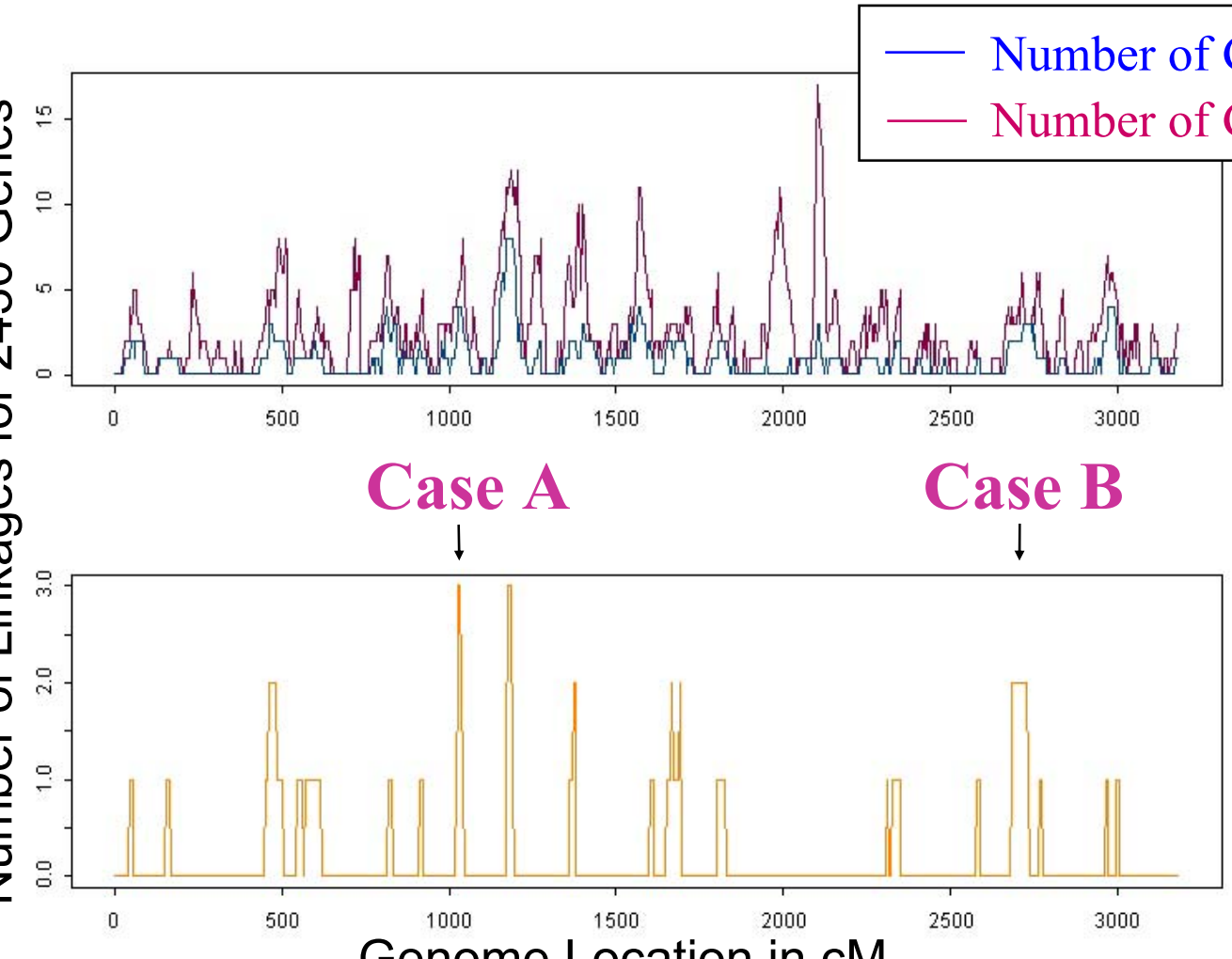
*comparison of additive genetic model versus sporadic model after adjustment for age, gender and age*gender;

Refined Clustering Based on Heritable Genes

- As expected, familial clustering is refined
- Note relationships among the expression of the top heritable genes
- Linkage analyses may reveal causes of gene expression correlation
- Linkage analyses may identify common genetic effects even for genes that are not highly correlated based on raw expression data



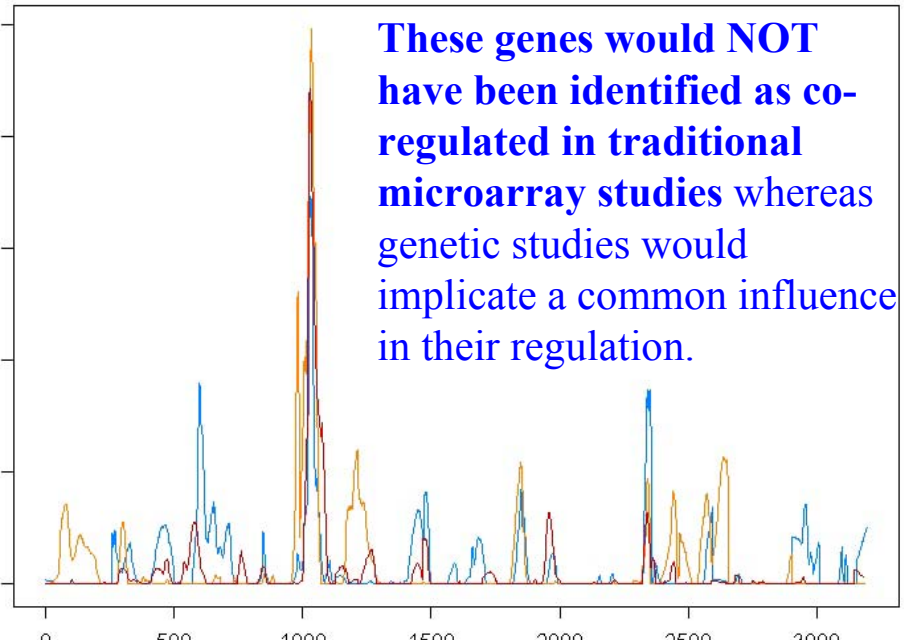
Summary of Linkages for DE Genes



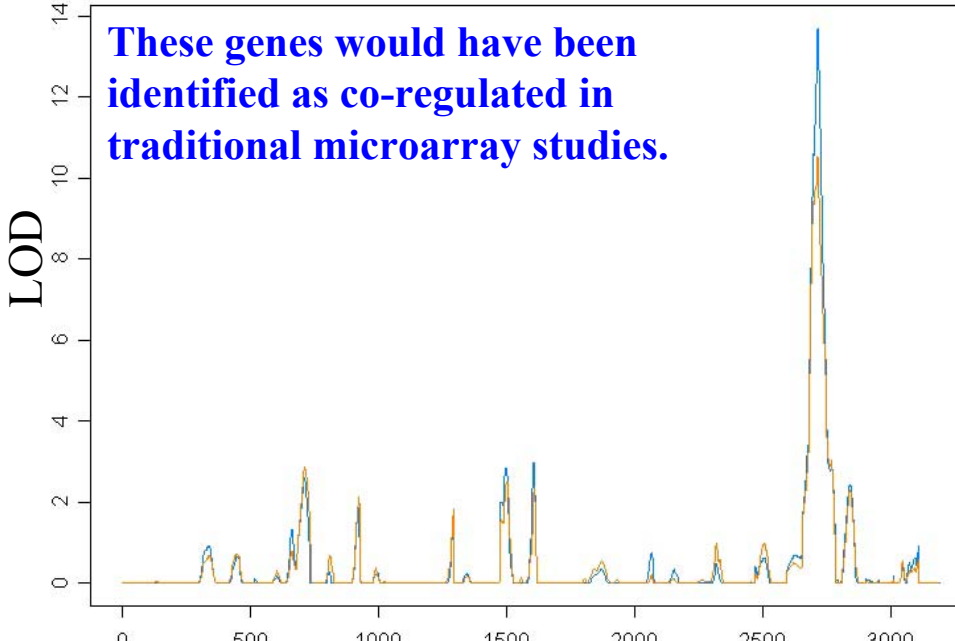
Number of genes with significant linkage using genome-wide significance per gene = 0.05/2430

Additional Power through the Genetics of Gene Expression

Case A: Overlaid LOD score plots for 3 genes that are not highly correlated (strongest correlation=0.09). Linkage results identify underlying common genetic influences. Bivariate segregation analyses failed to identify high genetic correlation with strongest genetic correlation equal to -0.14.



Case B: Overlaid LOD score plots for 2 highly correlated genes (correlation=0.98). Hence, linkage results are also highly correlated.



Conclusions for Human Study

Heritability of gene expression in human samples is detectable for a large number of genes. Similar results have been seen in yeast.[7] We have also detected this in mouse and maize.[8]

For genes identified through gene expression studies, it is possible to propose functional reasons for their co-regulation by using common regions of linkage.

Using the genetics of gene expression, genes that are co-regulated through a genetic mechanism can be identified even for genes with non-correlated expression. These genes would have been missed in traditional gene expression studies.

Overlaying this type of analysis with the genetics of complex traits will allow for a finer understanding of the underlying etiology of complex traits. We have shown this for obesity-related traits in mouse[8] and plan to extend such studies to humans in the near

Literature Cited

1. van 't Veer, L. J. et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-6. (2002).
2. Hughes, T. R. et al. Functional discovery via a compendium of expression profiles. *Cell* 102, 109-26. (2000).
3. Roberts, C. J. et al. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* 287, 873-80. (2000).
4. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62, 1198-211. (1998).
5. Kong, A. et al. A high-resolution recombination map of the human genome. Nat Genet 31, 241-7. (2002).
6. Williams, J. T. et al. Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulations and simulations results. Am J. Hum Genet 65:1134-1147 (1999).
7. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. Science 296, 752-5. (2002).
8. Schadt E. E., Monks S. A. et al. The genetics of gene expression: a survey of maize, mouse and man. Nature (submitted)