Inferring Genetic Architecture of Complex Biological Processes

Brian S. Yandell[*†], Christina Kendziorski[*‡], Hong Lan[§],

Elias N. Chaibub[*], Alan D. Attie[§]

[*]Department of Statistics

[†]Department of Horticulture

[‡]Department of Biostatistics & Medical Informatics

[§]Department of Biochemistry

University of Wisconsin-Madison

Madison, Wisconsin 53706 USA

Running Title: Inferring Genetic Architecture

Corresponding Author: Brian S. Yandell

Departments of Statistics and Horticulture

1300 University Avenue, MSC-1239

University of Wisconsin-Madison

Madison, Wisconsin 53706 USA

Phone: 608-262-1157

Fax: 608-262-0032

Email: byandell@wisc.edu

ABSTRACT

Gene mapping infers the relationship between genotype and phenotype in a segregating population. We map thousands of mRNA expression phenotypes, or expression QTL, using dimension reduction methods to uncover correlated genetic architecture, including number and location of genomic regions as well as gene action and epistasis. We show a novel blending of principal components and discriminant analysis with functional information to detect multiple expression QTL that together may affect the expression of many correlated mRNA. These common patterns of gene action are largely overlooked by simple interval mapping when conducted separately for each mRNA. In our current study with 60 F_2 mice from a B6-BTBR *ob/ob* model of diabetes and over 40,000 mRNA transcripts measured with Affymetrix chips, we find three pairs of genomic regions of particular interest associated with signal transduction, apoptosis, and lipid metabolism. We propose to join genetic architecture with graphical models of biochemical activity. Our approach is directly applicable to gene mapping for other "omic" measurements on the horizon.

INTRODUCTION

Gene mapping infers the relationship between genotype and phenotype in a segregating population (LANDER and BOTSTEIN 1989). The mRNA abundance data obtained from microarrays, "expression phenotypes", can be genetically mapped to identify loci that regulate gene expression. We map thousands of mRNA expression phenotypes to find expression trait loci (ETL), using dimension reduction methods to uncover their complex genetic architecture. Genetic architecture involves the number and location of genomic regions that affect phenotypic traits, as well as their gene action and possible epistatic interactions (KAO *et al.* 1999). We show how the novel use of principal components and discriminant analysis, in conjunction with functional information, can detect multiple expression QTL. These approaches identify loci that act together to coordinately regulate the levels of many mRNAs. Such patterns of gene action are largely overlooked by simple interval mapping conducted separately for each mRNA.

Our ultimate goal is to model causal biochemical networks based on biological process measurements on a panel of individuals from an experimental cross. Causal biochemical networks, or directed graphical models connecting genes, consist of nodes (genes and/or their mRNA) and connecting arrows indicating causal direction in a putative biochemical pathway. Several authors have proposed models to discover associations in genomic expression data using correlation (ZHOU *et al.* 2002), partial correlation (DE LA FUENTE *et al.* 2004) and Bayesian networks (FRIEDMAN 2004; SEGAL *et al.* 2002; SEGAL *et al.* 2003). Causal direction from gene to mRNA follows from the experimental cross, a perturbation to create unique genetic mosaics of the parent strains. Our approach extends existing causal biochemical networks using expression arrays in experimental crosses (SCHADT *et al.* 2003b; ZHU *et al.* 2004) by incorporating greater

complexity of genetic architecture and using meta-traits to identify sets of correlated expression traits for network construction.

Our model system consists of two mouse strains carrying a mutation that causes morbid obesity, the *leptin*^{ob} mutation, "*ob*". One strain, B6, is insulin resistant, but not diabetic. The other strain, BTBR, is insulin resistant and severely diabetic (LAN *et al.* 2003a). This allows us to search for genetic differences between the strains that are responsible for these dramatically different interactions with obesity. Our current study involves 60 mice selected for phenotyping (JIN *et al.* 2004) from an intercross of B6 and BTBR *ob/ob* mice as a model of diabetes, with over 40,000 mRNA traits measured with Affymetrix chips. Here, we report three pairs of genomic regions of particular interest that may be associated with signal transduction, apoptosis, and lipid metabolism.

Central to the identification of such causal networks is finding expression (or other 'omic') phenotypes that are heritable in the broad sense and inferring patterns of genetic architecture that influence that heritability. The strategy is as follows: (1) find all mRNA showing evidence for heritability at one or more genomic loci; (2) organize these heritable mRNA into functional groups and create meta-traits (defined below); (3) map meta-traits to infer the genetic architecture of the biological process; (4) infer graphical models of biochemical pathways using the genetic architecture. In short, we build on a Bayesian identification scheme (KENDZIORSKI *et al.* 2004) by formally incorporating genetic architecture, allowing for multiple QTL gene mapping. We map meta-traits, weighted averages of heritable expression phenotypes, to infer the regions of the genome that may contain transcription factors or other regulators that broadly influence heritability.

Three important issues have not been addressed in previous efforts at gene mapping of expression traits (BREM *et al.* 2002; LAN *et al.* 2003b; SCHADT *et al.* 2003a; YVERT *et al.* 2003). First, it is important to establish the false discovery rate across inference for all mRNA (STOREY and TIBSHIRANI 2003). Second, many genes probably affect each phenotype, and several allelic combinations may yield the same observable phenotype through heterogeneity. Thus multiple expression QTL affect mRNA expression, and multiple QTL models are crucial to uncover genetic architecture. Even apparently *cis*-acting mRNA could be affected by *trans*-acting QTL, missed in a single QTL scan. Finally, one gene can affect the expression of multiple mRNAs through pleiotropy. Meta-trait gene mapping provides a coherent framework for uncovering the pleiotropic action of a few genetic loci on multiple co-regulated mRNA. Post-hoc analysis without such a formal framework is potentially quite misleading.

METHODS

Meta-trait Gene Mapping Framework

Simultaneous identification of heritable mRNA expression: Our first step involves simultaneous mapping of all mRNAs to identify a subset of mRNAs that show evidence for genetic control (KENDZIORSKI *et al.* 2004). Suppose H signifies the event that the expression of a particular mRNA is heritable at some set of loci for an experimental cross. We want to estimate $\text{pr}(H \mid Y, Q)$, the probability that the mRNA is heritable given the observed expression values Y and genotypes Q at those loci. We find all mRNA that for some Q have $\text{pr}(H \mid Y, Q) > 0.95$, which controls the expected posterior false discovery rate (NEWTON *et al.* 2004) at 5%.

Bayes theorem yields the posterior probability of linkage,

$$\text{pr}(H \mid Y, Q) = \text{pr}(Y \mid Q, H)\ \text{pr}(H \mid Q) / \text{pr}(Y \mid Q).$$

where pr($H \mid Q$) can be estimated from the data. The marginal probability of expression is a weighted average over the cases of heritable versus non-heritable traits,

$$\text{pr}(Y \mid Q) = (1 - \text{pr}(H|Q))\ \text{pr}(Y \mid Q, \text{not } H) + \text{pr}(H|Q)\ \text{pr}(Y \mid Q, H).$$

An mRNA trait with no heritability has marginal probability of

$$\text{pr}(Y \mid Q, \text{not } H) = f_0(Y) = \int f(Y \mid \mu)\ \text{pr}(\mu)\ d\mu,$$

the average of a probability of expression, $f(Y|\mu)$, weighted by a prior probability for the mean expression level, pr(μ). Heritable traits have marginal probability of

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

with the product over all possible genotypes q. Here, $Y_q = \{Y_i \mid Q_i = q\}$ are the mRNA expression values for the subset of n_q individuals with QTL genotype q, and $n = \text{sum}_q\ n_q$.

Functional groups and meta-traits: We focus on modeling genetic architecture for a few meta-traits that are highly correlated with many mRNA expression traits. A meta-trait (DOBRA *et al.* 2004) is a weighted average of heritable mRNA expression traits that approximates a putative transcription factor or regulator driving expression of a set of mRNA transcripts. That is, $Z = YW$, with Y a matrix of all heritable trait values and the weights W determined to capture the correlation across expression mRNA. We use principal components or discriminant analysis to determine different weights, W, but other dimension-reduction methods could be used.

Principal component (PC) meta-traits can be created for functional groups of heritable mRNA with strong putative co-regulation to capture the major phenotypic variation in one or two dimensions. Roughly half of the genes corresponding to mRNA expression traits can be organized using gene ontology principles (www.geneontology.org). We followed a strategy analogous to MOOTHA *et al.* (MOOTHA *et al.* 2003) on a subset of heritable traits to develop functional groups. We then constructed the first two principal components (meta-traits) for each

functional group and prioritized analysis based on the percent variation explained by the first PC meta-trait.

Discriminant analysis (DA) meta-traits are weighted averages of *all* mRNA that show evidence of heritability, yielding the highest heritability relative to a particular genetic architecture. We scan over a class of possible genetic architectures to find the combinations of QTL and DA meta-traits with the highest heritability. Due to computational considerations, this was limited to models with two QTLs.

Mapping meta-traits to infer genetic architecture: Gene mapping (LANDER and BOTSTEIN 1989) associates phenotype with genotype via models of genetic architecture, including putative genetic loci and their associated gene action. Here, we propose a model search for multiple QTL that act together to influence gene expression. We gene-map a modest number of meta-traits, then use their inferred genetic architectures to map expression traits most closely associated with each meta-trait.

The genetic architecture of a meta-trait, Z, involves a model, M, for the gene action and epistatic interactions of genetic loci, λ. For instance, with three loci, $\lambda = (\lambda_1, \lambda_2, \lambda_3)$, the model $M = \{\lambda_1, \lambda_2, \lambda_3, (\lambda_1, \lambda_2)\}$ includes the main effects of loci λ_1, λ_2, and λ_3 as well as the epistatic interaction of pair λ_1 and λ_2. The main effects could be further partitioned into additive and dominance effects, but need not be. Those individuals with genotype $Q = q$ at the loci, λ, have phenotypes with probability $f(Z \mid \mu_q)$ that depends on the genotype q only through the genotypic mean, μ_q. This mean expression can be partitioned into genotypic effects for each model component,

$$\mu_q = \mathrm{E}(Z \mid M, Q = q) = \beta_0 + \Sigma_{\{k \text{ in } M\}} \beta_{qk} .$$

In the absence of data, the genotypic means, μ_q, follow the same prior probability assumed for non-heritable traits, pr(μ), discussed in the previous section. This prior can be partitioned into priors for each model effect. For instance, we can give β_0 and β_{qk} independent priors $p_0(\mu) = \text{pr}(\mu / \tau) / \tau$ and $p_{1m}(\mu) = \text{pr}(\mu / \tau_{1m}) / \tau_{1m}$, respectively, with $m = |M|$, the number of effects in the model, τ an arbitrary value between 0 and 1, and $(\tau_{1m})^2 = (1-\tau^2) / m$. Ideally $(1-\tau^2)$ is proportional to heritability (see GAFFNEY 2001).

With a dense genetic map and completely informative markers, Q would be known exactly for every set of loci. In practice, a map is incomplete, and we infer Q based on flanking markers, X, to loci, λ, using a recombination model $f(Q = q \mid X, \lambda)$. The interval mapping problem formally involves finding the genetic architecture, M, that maximizes the probability

$$f(Z \mid X, M) = \Sigma_q \; f(Z \mid \mu_q) \; f(Q = q \mid X, \lambda).$$

This was developed as single QTL interval mapping (Lander and Botstein 1989), and later extended to multiple QTL (KAO *et al.* 1999; SATAGOPAN *et al.* 1996; SEN and CHURCHILL 2001; YI and XU 2002; YI *et al.* 2003). Multiple QTL interval mapping includes searching over possible loci, λ, and possible models, M, and assessing the best model according to Bayes factors or another criterion. Additional issues arise when the meta-trait depends on the model, as with discriminant analysis (GILBERT and LE ROY 2004).

A meta-trait, Z, represents multiple traits that are correlated, while its inferred genetic architecture, M, identifies genomic regions, or loci, driving the major heritable aspects of those correlated traits. We approximate each meta-trait by a subset of heritable traits, $Z \approx \underline{YW}$, in a parsimonious way (*cf.* TIBSHIRANI 1996) to balance the fit and keep the subset of traits conveniently small. We can then extend M to a multiple trait genetic architecture, $\underline{M}$, using multiple trait mapping (*cf.* JIANG and ZENG 1995).

Causal Bayesian networks to infer biochemical pathways: Now we wish to find the best graphical model, G, to explain the relationship among a set of expression traits, $\underline{Y}$, with respect to the genetic architecture, $\underline{M}$. This graphical model then estimates the key biochemical pathways important to the meta-trait and its associated expression phenotypes.

We model their genetic correlation structure using a graphical model that connects traits in a directed graph along putative biochemical pathways. We wish to infer the posterior probability of a graph, G, given the genetic architecture, $\underline{M}$; the genotypes, Q, at the loci of $\underline{M}$; and the expression traits, $\underline{Y}$:

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

Here $\text{pr}(G \mid \underline{M})$ is a prior probability on valid graphs given the genetic architecture, and the denominator is the marginal probability for traits averaged across all possible graphs. The marginal probability for the correlated traits, $\underline{Y}$, given a particular graph, G, is

$$\text{pr}(\underline{Y} \mid Q, G) = f_1(\underline{Y} \mid Q, G) = \prod_q f_0(\underline{Y}_q \mid G).$$

The trait probabilities, $f_0(\underline{Y} \mid G)$, now are multivariate,

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

with prior, $\text{pr}(\underline{\mu}) = \prod_t \text{pr}(\mu_t)$, assumed independent across the traits.

The multivariate trait probability, $f_1(\underline{Y} \mid \underline{Q}, G)$, depends on the correlation structure through the graph G. The phenotype model for a graph follows the chaining implied by the graphical model. For instance, with two traits and a graph g such that $Y_1 \rightarrow Y_2$, this probability becomes the product of two univariate trait probabilities,

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

Essentially, the second probability models the genetic architecture of the QTL on the residuals after regressing Y_2 on Y_1. We focus graphical models on the QTLs identified by the meta-trait,

although individual expression mRNA in the identified subset of heritable traits, $\underline{Y}$, need not be influenced by all these QTLs. Details of inference across graphical models will be reported elsewhere.

Details of Implementation

We use an empirical Bayes histogram estimate of the prior pr(μ) by combining all mRNA measurements (KENDZIORSKI *et al.* 2004). We control the expected posterior false discovery rate by directly estimating the heritability via single marker regression for all mRNA traits. For convenience, we consider traits Y as log base 2 expression (IRIZARRY *et al.* 2003) and assume the phenotype model f is normal.

We report two-QTL gene mapping using the Haley-Knott (HALEY and KNOTT 1992) regression, which is fast and fairly close to the Bayesian evaluation. Even with two QTL, the computational costs were substantial. Further, with 60 mice in our study (JIN *et al.* 2004), it seemed prudent to restrict attention to at most two QTL to avoid over-interpreting our data. We used R/qtl (BROMAN *et al.* 2003) to scan the mouse genome in a variety of ways sketched below. Genetic architectures with highest LODs for meta-traits were tested using 500 permutations under the null model of no heritability.

Functional groups and meta-traits: Roughly 4000 mRNAs were identified as heritable by one of several methods (KENDZIORSKI *et al.* 2004). A subset of over 1500 was organized by the Attie laboratory by molecular function into 84 functional groups using gene ontology.

Functional groups with the PC meta-trait explaining over 65% of the variation were considered promising candidates for gene mapping for one or two QTL. The PC meta-trait was mapped using one-dimensional and two-dimensional genome scans with the Haley-Knott (HALEY and KNOTT 1992) regression approximation. The chromosome pair with the peak LOD

for 2-D scans was evaluated using a permutation test with 500 replications. Factor loadings, or correlations of mRNA expression with PC meta-trait, were plotted to show patterns of correlation within a functional group. Groups with high correlation between the first PC meta-trait and many mRNAs show evidence for coregulation. The mRNAs in these groups were then fit separately to the same genetic architecture.

Discriminant analysis (DA) meta-traits are based on genotype under a model *M.* Gilbert and le Roy (GILBERT and LE ROY 2003; GILBERT and LE ROY 2004) gene-mapped using DA at 2 cM spacing for single QTL in a backcross. We considered all possible pairs of the 194 markers. The DA meta-trait for a marker pair is that weighted average of phenotypes with highest heritability relative to two QTL at the marker pair. That is, the DA meta-trait produces the largest possible F statistic from a two-factor ANOVA on the marker pair. The nominal *p*-value is artificially deflated, but provides a relative basis for comparison. Locus pairs with the best fits (smallest *p*-values) across the genome scan were evaluated using 500 permutations: QTL genotypes were permuted and new DA meta-traits were computed for each permutation. Each DA meta-trait was mapped using a 2-D scan with Haley-Knott (HALEY and KNOTT 1992) regression to verify that it peaked only near the marker pair.

RESULTS

Principal components on functional groups: Eight functional groups had two or more unique mRNA and at least 65% of variation in the PC meta-trait (Figure 1a). Single QTL scans had highest LODs of 2 to 4, barely significant. Strong evidence emerged for two QTL on several of the functional groups, with joint LODs above 8, significant at the 1-5% level based on permutation tests (not shown). Permutation tests for epistatic interactions were generally not significant.

"Translation machinery" and "RNA binding" functional groups had strong correlation of mRNA expression with their PC meta-trait, suggesting coregulation (Figure 1b-c). "Chromosome organization" had small factor loadings in general, with the exception of one mRNA trait (not shown). The other five groups showed intermediate patterns and were dropped.

The translation machinery PC meta-trait LOD peak of 9.55 occurred at Chr 4 at 37 cM and Chr 15 at 15 cM. One gene, EIF 3 subunit 6, appears to be *cis*-regulated as it resides in the Chr 15 locus support interval. Seven of 11 unique mRNA in this group had peaks coinciding with that of the PC meta-trait, and five of these seven had joint LODs above 8. Thus two of the seven mRNA with correlated transcription in this group would have been missed by individual analysis.

The RNA binding PC meta-trait had its LOD peak of 8.62 coincident with that of translation machinery. Six of 12 unique mRNA in this group had LOD peaks at this loci pair between 4.9 and 8.0, and would otherwise have been missed. Across all heritable traits in functional groups, we found over 600 mRNA with absolute correlations to the translation machinery PC meta-trait above 0.5. This implied a much broader function for our locus pair.

Discriminant analysis genome scan: A discriminant analysis genome scan of the 1500+ heritable traits with known function turned up three promising pairs of loci showing epistatic interaction. These all had nominal *p*-values just below 10^{-10}, or $-\log_{10}(p\text{-value})>10$, for ANOVA of the DA meta-trait on the nine genotypes determined by pairs of loci.

One of these (DA1) was the same Chr 4 at 37 cM and Chr 15 at 15 cM loci pair identified by principal component analysis on functional groups. This is not surprising, given that we found strong correlation between that PC meta-trait and many of the heritable mRNA traits. The other two loci pairs are Chr 2 at 25 cM and Chr 13 at 0 cM (DA2) and Chr 5 at 49 cM and Chr 9 at 45 cM (DA3). Permutation *p*-values for DA meta-traits at these three pairs are 0.008, 0.036, and 0.036, respectively. All three DA meta-traits had genome-wide maximum LODs at these locations.

Representation of functional groups (Table 1) strongly suggests association of DA meta-traits with signal transduction (DA1), apoptosis (DA2) and lipid metabolism (DA3). The fatty acid synthesis functional group, including SCD1, figures prominently in DA2 and DA3 meta-traits.

Figure 2 shows the epistatic interactions of the DA meta-traits. Additive effect of Chr 4 locus apparent only when heterozygous at Chr 15 locus; (b) B6/B6 genotype is significantly different from other genotypes Chr 2 and Chr 13 loci; (c) additive effect of Chr 5 and Chr 9. Note the similarities in epistatic interaction patterns for SCD1 at these same loci, shown in Figure 3. Reanalysis using the 4000 traits identified as heritable confirmed these pairs as interesting; although the nominal *p*-values were somewhat larger (near 10^{-8}), the permutation *p*-values were similar (0.004, 0.036, 0.092, respectively). In addition, a few other promising marker pairs were identified (not reported here).

Causal biochemical networks: At this point we are tracking down the *cis*- and *trans*- acting loci that influence mRNA abundance. Complementary but independent biochemical investigations in the Attie lab are converging on the same regions in Chr 2 and Chr 9. These results will be reported elsewhere.

DISCUSSION

Our main interest is building causal biochemical networks with a coherent statistical model of multiple QTL tied directly to observed expression phenotypes. We present a data-driven theoretical framework and show preliminary results that will allow us to focus attention on a few important regulatory pathways. We believe this can sharpen the search for key network components, reducing the rate of false discovery while allowing directly for the multidimensional nature of such networks.

No research team to our knowledge has attempted DA-based multiple QTL interval mapping, especially with thousands of traits. Gilbert and le Roy (GILBERT and LE ROY 2003; GILBERT and LE ROY 2004) use exact methods for one QTL in a backcross. We found certain anomalies with multiple QTL, in particular when one possible genotype $Q=q$ had only one individual who differed from the others for a small subset of heritable traits. We had to remove one mouse because it created 29 spurious associations at loci pairs where it had a unique genotype.

While we focus in this development on mRNA expression, the methods naturally extend to other high throughput "omic" data on the horizon. In addition, other massive data technologies, such as brain and body scans, could be used in gene mapping with modest adjustments.

ACKNOWLEDGEMENTS

Karl Broman initially suggested to BY that discriminant analysis might be useful in this situation. The authors wish to acknowledge the insights and feedback of Jessica Byers and other members of the Attie Laboratory. BSY and ADA were supported in part by United States Department of Agriculture CSREES. BSY, ADA and HL were supported in part by National Institutes of Health/NIDDK, 5803701 and 66369-01, and American Diabetes Association Innovation Award, 7-03-IG-01. CMK was supported in part by Howard Hughes Medical Institute, 133-ES29. ENC was supported by the National Council for Scientific and Technological Development (CNPq), Brazil.

Table 1. Top ten functional groups associated with each DA meta-trait. Functional groups were ordered by the average absolute correlation of mRNA within each group with the DA meta-trait after dropping mRNA with absolute correlations below 0.5. Number of unique mRNA and 100 * average absolute correlation (100 = perfect correlation) included in parenthesis. For endocytosis, 3 unique mRNA have average correlation of 0.84 with DA1.

DA1 meta-trait at Chr 4 at 37 cM & Chr 15 at 15 cM (signal transduction): endocytosis (3, 84); RNA binding (11, 79); cholesterol/sterol biosynthesis (5, 79); translation machinery (8, 78); RAS oncogene superfamily (14, 77); mRNA processing (5, 76); calcium binding protein (6, 76); intracellular signal transduction (15, 75); transcription factor (24, 75); transcription machinery (12, 74); plus 60 others with 2 or more.

DA2 meta-trait at Chr 2 at 25 cM & Chr 13 at 0 cM (lipid metabolism): nerve signal transduction (2, 64); fatty acid synthesis (10, 63); membrane protein (3, 62); Krebs-TCA cycle (2, 61); triglyceride/phospholipid biosynthesis (2, 61); retinol metabolism (2, 61); cholesterol/sterol biosynthesis (3, 61); digestion (3, 60); cytoskeleton (2, 59); protein degradation/ubiquitin (2. 58); plus 15 others with 2 or more.

DA3 meta-trait at Chr 5 at 49 cM & Chr 9 at 45 cM (apoptosis): apoptosis (2, 66); nuclear acid metabolism (2, 61); cell adhesion (3, 61); cytochrome P450 (4, 60); calcium binding protein (2, 58); fatty acid synthesis (3, 58); solute carrier (3, 57); small molecule transporter (3, 57); urea cycle (2, 55); amino acid metabolism (5, 54); steroid hormone metabolism (3, 53).

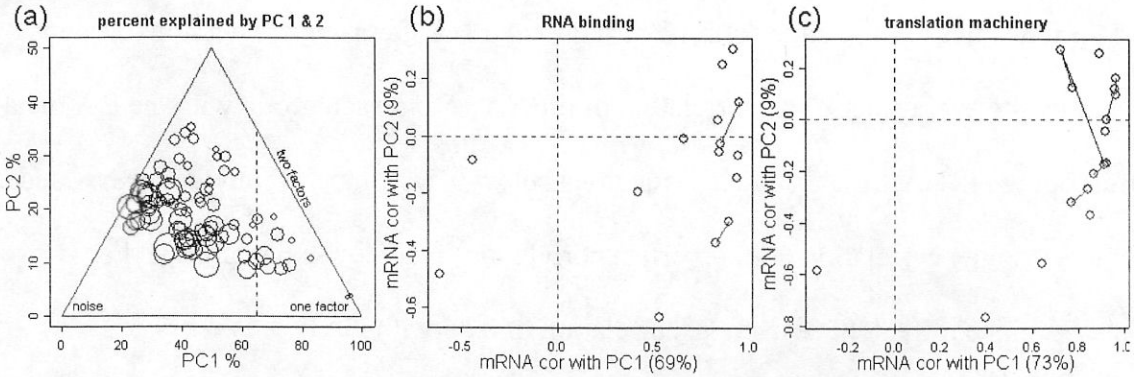

Figure 1. PC meta-trait filtering. (a) Percent of variation in functional groups explained by first two principal components. Circle for functional group have area proportional to the number of unique mRNAs. Vertical dashed line at PC1 = 65% indicates first cut for analysis. (b-c) Factor loadings (correlations) with first two principal components for two functional groups with PC1>65%. Points joined by lines are different probe sets for the same mRNA. (b) "RNA binding" (13 unique mRNA); (c) "translation machinery" (11 unique mRNA).

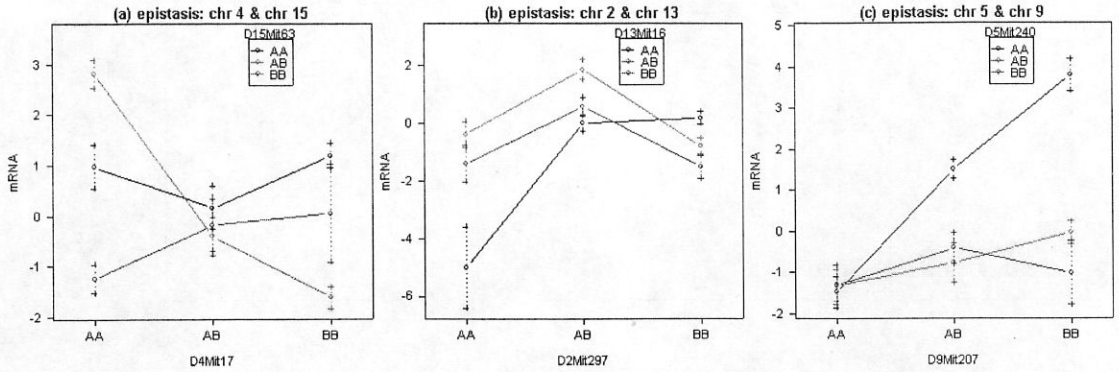

Figure 2. Epistatic interaction plots for three most significant DA meta-traits. Genotypes at markers are AA (B6 type), AB (heterozygous) and BB (BTBR type). Vertical units are standard deviation. (a) DA1: additive mRNA effect of Chr 4 locus only when heterozygous at Chr 15 locus; (b) DA2: B6/B6 genotype is significantly different from other genotypes Chr 2 and Chr 13 loci; (c) DA3: additive effect of Chr 5 and Chr 9.

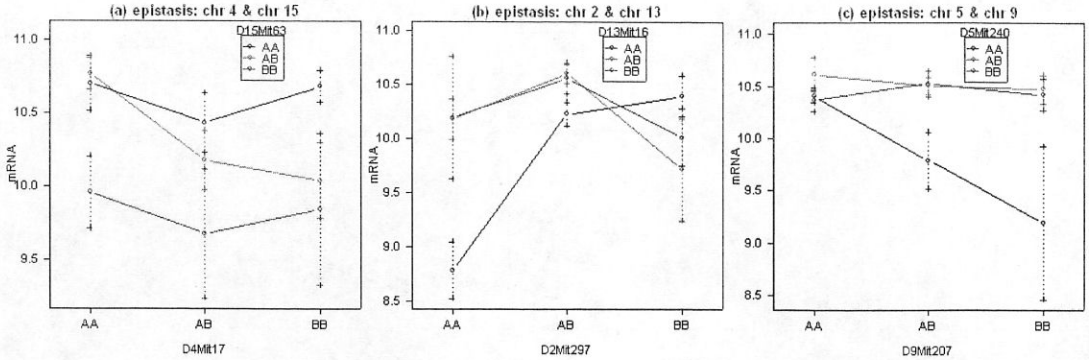

Figure 3. Epistatic interaction plots for stearoyl-Coenzyme A desaturase 1 (SCD1) at marker pairs identified by DA meta-traits. Vertical units are $\log_2$(expression). Note similarity of interaction patterns with Figure 2.

LITERATURE CITED

BREM, R. B., G. YVERT, R. CLINTON and L. KRUGLYAK, 2002 Genetic dissection of transcriptional regulation in budding yeast. Science **296:** 752-755.

BROMAN, K. W., H. WU, S. SEN and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890.

DE LA FUENTE, A., N. BING, I. HOESCHELE and P. MENDES, 2004 Discovery of meaningful associations in genomic data using partial correlation coefficients. Bioinformatics **00:** 000-000.

DOBRA, A., C. HANS, B. JONES, J. R. NEVINS, G. A. YAO *et al.*, 2004 Sparse graphical models for exploring gene expression data. Journal of Multivariate Analysis **90:** 196-212.

FRIEDMAN, N., 2004 Inferring cellular networks using probabilistic graphical models. Science **303:** 799-805.

GAFFNEY, P., 2001 An efficient reversible jump Markov chain Monte Carlo approach to detect multiple loci and their effects in inbred crosses, pp. in *Department of Statistics*. University of Wisconsin-Madison, Madison, WI.

GILBERT, H., and P. LE ROY, 2003 Comparison of three multitrait methods for QTL detection. Genet Sel Evol **35:** 281-304.

GILBERT, H., and P. LE ROY, 2004 Power of three multitrait methods for QTL detection in crossbred populations. Genet Sel Evol **36:** 347-361.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315-324.

IRIZARRY, R. A., B. HOBBS, F. COLLIN, Y. D. BEAZER-BARCLAY, K. J. ANTONELLIS *et al.*, 2003 Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:** 249-264.

JIANG, C., and Z.-B. ZENG, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics **140:** 1111-1127.

JIN, C. F., H. LAN, A. D. ATTIE, D. BULUTUGLO, G. A. CHURCHILL *et al.*, 2004 Selective Phenotyping for Increased Efficiency in Genetic Mapping Studies. Genetics **Accepted**.

KAO, C. H., Z. B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203-1216.

KENDZIORSKI, C. M., M. CHEN, M. YUAN, H. LAN and A. D. ATTIE, 2004 Statistical methods for expression trait loci (ETL) mapping. Technical Report 184, Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison. **Submitted**.

LAN, H., M. E. RABAGLIA, J. P. STOEHR, S. T. NADLER, K. L. SCHUELER *et al.*, 2003a Gene expression profiles of nondiabetic and diabetic obese mice suggest a role of hepatic lipogenic capacity in diabetes susceptibility. Diabetes **52:** 688-700.

LAN, H., J. P. STOEHR, S. T. NADLER, K. L. SCHUELER, B. S. YANDELL *et al.*, 2003b Dimension reduction for mapping mRNA abundance as quantitative traits. Genetics **164:** 1607-1614.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185-199.

MOOTHA, V. K., J. BUNKENBORG, J. V. OLSEN, M. HJERRILD, J. R. WISNIEWSKI *et al.*, 2003 Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. Cell **115:** 629-640.

NEWTON, M. A., A. NOUEIRY, D. SARKAR and P. AHLQUIST, 2004 Detecting differential gene expression with a semiparametric hierarchical mixture method. Biostatistics **5:** 155-176.

SATAGOPAN, J. M., B. S. YANDELL, M. A. NEWTON and T. C. OSBORN, 1996 A bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. Genetics **144:** 805-816.

SCHADT, E. E., S. A. MONKS, T. A. DRAKE, A. J. LUSIS, N. CHE *et al.*, 2003a Genetics of gene expression surveyed in maize, mouse and man. Nature **422:** 297-302.

SCHADT, E. E., S. A. MONKS and S. H. FRIEND, 2003b A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. Biochem Soc Trans **31:** 437-443.

SEGAL, E., Y. BARASH, I. SIMON, N. FRIEDMAN and D. KOLLER, 2002 From promoter sequence to expression: a probabilistic framework, pp. in *6th Inter. Conf. on Research in Computational Molecular Biology (RECOMB)*, Washington, DC.

SEGAL, E., M. SHAPIRA, A. REGEV, D. PE'ER, D. BOTSTEIN *et al.*, 2003 Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet **34:** 166-176.

SEN, S., and G. A. CHURCHILL, 2001 A statistical framework for quantitative trait mapping. Genetics **159:** 371-387.

STOREY, J. D., and R. TIBSHIRANI, 2003 Statistical significance for genomewide studies. Proc Natl Acad Sci U S A **100:** 9440-9445.

TIBSHIRANI, R., 1996 Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society Series B-Methodological **58:** 267-288.

YI, N., and S. XU, 2002 Mapping quantitative trait loci with epistatic effects. Genet Res **79:** 185-198.

YI, N., S. XU and D. B. ALLISON, 2003 Bayesian model choice and search strategies for mapping interacting quantitative trait Loci. Genetics **165:** 867-883.

YVERT, G., R. B. BREM, J. WHITTLE, J. M. AKEY, E. FOSS *et al.*, 2003 Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet **35:** 57-64.

ZHOU, X., M. C. KAO and W. H. WONG, 2002 Transitive functional annotation by shortest-path analysis of gene expression data. Proc Natl Acad Sci U S A **99:** 12783-12788.

ZHU, J., P. Y. LUM, J. LAMB, D. GUHATHAKURTA, S. W. EDWARDS *et al.*, 2004 An integrative genomics approach to the reconstruction of gene networks in segregating populations. Cytogenet Genome Res **105:** 363-374.