Inferring Causal Phenotype Networks From Segregating Populations

Elias Chaibub Neto,* Christine T. Ferrara,†,‡ Alan D. Attie‡ and Brian S. Yandell*,§,1

*Department of Statistics, ‡Biochemistry Department and §Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706 and †Sarah W. Stedman Nutrition and Metabolism Center and Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27704

ABSTRACT

A major goal in the study of complex traits is to decipher the causal interrelationships among correlated phenotypes. Current methods mostly yield undirected networks that connect phenotypes without causal orientation. Some of these connections may be spurious due to partial correlation that is not causal. We show how to build causal direction into an undirected network of phenotypes by including causal QTL for each phenotype. We evaluate causal direction for each edge connecting two phenotypes, using a LOD score. This new approach can be applied to many different population structures, including inbred and outbred crosses as well as natural populations, and can accommodate feedback loops. We assess its performance in simulation studies and show that our method recovers network edges and infers causal direction correctly at a high rate. Finally, we illustrate our method with an example involving gene expression and metabolite traits from experimental crosses.

NETWORK models derived from microarray experiments have shed light on the manner and extent of connectedness among expressed genes. However, these networks mostly summarize association, connecting phenotypes without causal direction. Genetical genomic studies, with microarray data in a segregating population, have shown evidence for *cis*-acting and *trans*-acting genomic regions (BREM *et al.* 2002; SCHADT *et al.* 2003). These studies yield expression quantitative trait loci (eQTL) (KENDZIORSKI and WANG 2006) that provide a causal experimental system where genotype drives phenotype. Directed graphs inferred from such a system (JANSEN and NAP 2001; ZHU *et al.* 2004; BING and HOESCHELE 2005; KULP and JAGALUR 2006; KEURENTJES *et al.* 2007) can yield causal gene networks that predict biochemical pathways, generating new hypotheses about functional relationships among expressed genes. These methods need a network node that consists of a gene physically located within the QTL support interval of one or more transcripts. This gene thus becomes a candidate regulator for the other transcripts. Unfortunately, phenotypes such as metabolites and physiological traits cannot serve as causal nodes in these networks. SCHADT *et al.* (2005) allowed for causal clinical traits by using pairs of phenotypes that share multiple QTL. However, their method cannot infer causal direction between phenotypes without a common QTL.

In this article, we introduce a novel method for building causal networks among phenotypes that may not have common QTL, allowing the use of phenotypes other than gene expression. We first build an undirected graph that infers associations among phenotypes, using either an undirected dependency graph (UDG) (SHIPLEY 2002; DE LA FUENTE *et al.* 2004) or a skeleton derived from the PC (Peter-Clark) algorithm of SPIRTES *et al.* (2000). We next use a LOD score to determine causal direction for every edge that connects a pair of phenotypes, conditional on connected QTL. Our method differs from the PC algorithm, which first infers a graph skeleton (with the "PC-skeleton algorithm") and then uses partial correlations among phenotypes to infer the direction of some, but not all, edges of the inferred skeleton. Our QTL-directed dependency graph (QDG) can include feedback loops, which are common in biology. This integrated QDG approach overcomes serious limitations of both the genetical genomics network methods and the PC algorithm.

Our goal, in short, is to causally orient every edge connecting a pair of phenotypes in an undirected network. That is, does phenotype 1 drive phenotype 2, or vice versa?:

$$M_1: \text{①} \rightarrow \text{②} \qquad M_2: \text{①} \leftarrow \text{②}$$

Model selection procedures cannot distinguish between M_1 and M_2 because they are distribution or likelihood equivalent. That is, variation in the level of phenotype 1, y_1, causing a variation in the level of phenotype 2, y_2, yields the same joint density as the reverse situation,

$$f(y_1)f(y_2 \mid y_1) = f(y_1, y_2) = f(y_2)f(y_1 \mid y_2).$$

[1]*Corresponding author:* Department of Statistics, University of Wisconsin, 1300 University Ave., MSC-1239, Madison, WI 53706.
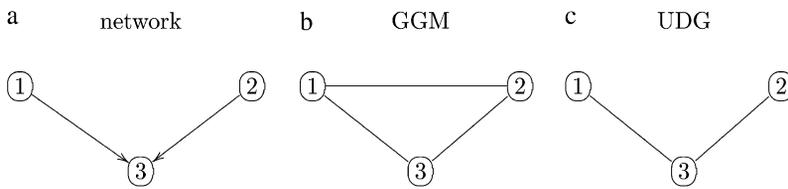E-mail: byandell@wisc.edu

FIGURE 1.—The causal network for three nodes (a) yields a Gaussian graphical model (GGM) using partial correlations (b). The edge between phenotypes 1 and 2 reflects partial correlation given 3, even though these two variables are causally independent. The undirected dependency graph (UDG) infers the correct edges (c) by first using the fact that 1 and 2 are uncorrelated to remove the 1–2 edge.

However, if in addition to the two phenotypes, we have measured genotypes, $\mathbf{q}_1 = \{q_{11}, \ldots, q_{1k}\}$ on k QTL affecting y_1 and $\mathbf{q}_2 = \{q_{21}, \ldots, q_{2l}\}$ on l QTL affecting y_2, we can now resolve direction. The new models
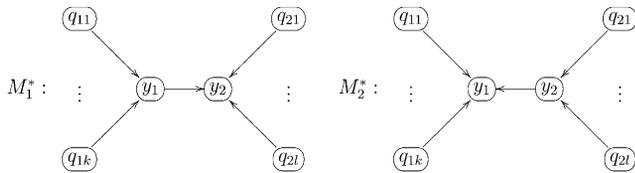
are *not* likelihood equivalent because the predictive densities disagree,

$$f(\mathbf{q}_1)f(y_1 \mid \mathbf{q}_1)f(y_2 \mid y_1, \mathbf{q}_2)f(\mathbf{q}_2) \neq f(\mathbf{q}_2)f(y_2 \mid \mathbf{q}_2)f(y_1 \mid y_2, \mathbf{q}_1)f(\mathbf{q}_1). \quad (1)$$

Therefore, we distinguish between models M_1^* and M_2^* by inferring direction of causation between phenotypes using a LOD score that conditions on genotypes at multiple QTL. We assume distinct QTL for different phenotypes and consider allegedly pleiotropic QTL in the DISCUSSION.

The methods of this article are widely applicable to human studies and outbred populations. Here, we confine attention to experimental crosses, such as a backcross or intercross derived from two disparate inbred lines. We focus on a mouse intercross with a set of already identified phenotypes of interest, ideally involved in or related to a common pathway. Further, we assume that multiple QTL associated with these traits had previously been determined.

We present a series of simulation studies where we evaluate the recovery rate of undirected edges and (1) compare the performance of the QDG and the PC algorithm for causal orientation in a network containing 100 phenotypes; (2) assess the accuracy of the QDG method for the inference of small cyclic graphs; and (3) study a causal model describing the relationship of 20 QTL, five gene expression traits, and one metabolite, constructed with the QDG method and presented in FERRARA *et al.* (2008).

METHODS

Undirected graph: We build an association network using UDGs (SHIPLEY 2002) and PC skeletons (SPIRTES *et al.* 2000). These methods are better suited to causal

inference than other association networks such as Gaussian graphical models (GGMs) (SCHAFER and STRIMMER 2005a,b), which include edges between any pair of nodes that has significant partial correlation. However, partial correlation can exist even when two nodes are uncorrelated, leading to spurious edges (Figure 1). In directed acyclic graphs, UDGs and PC skeletons avoid this problem by first removing edges where there is no significant correlation. In directed cyclic graphs, it may not be possible to remove all spurious edges. The PC-skeleton algorithm performs fewer computations than the UDG algorithm for sparse graphs, but it is less accurate in small sample sizes (SHIPLEY 2002). An efficient implementation of the PC algorithm for sparse high-dimensional directed graphs (KALISCH and BUHLMANN 2007) is available in the R package pcalg (R DEVELOPMENT CORE TEAM 2006).

Distinguishing QTL with direct and indirect effects: The edge orientation procedure requires the inclusion of QTL that directly affect the phenotypes. Since it is possible that a strong QTL affecting an upstream trait may also be incorrectly detected as a QTL for a downstream phenotype, we first employ the following two-step preprocessing procedure to distinguish QTL with direct and indirect effects: (1) identify all pairs of connected phenotypes that share common QTL and (2) apply a generalization of the method proposed by SCHADT *et al.* (2005) to each of these pairs to determine if the common QTL directly affects both traits or if it has an indirect effect in one of the phenotypes. If the pair of phenotypes is affected by a single common QTL (in addition to the QTL particular to each trait) we score the three graphical models presented in Figure 2, using the Bayesian information criterion (BIC) or Akaike's information criterion (AIC), and select the model with the smallest value. Figure 2a has both phenotypes directly affected by the common QTL, while in Figure 2b q has an indirect effect on phenotype y_2 and no direct effect on y_2. Figure 2c represents the reverse situation. Supplemental Figure S1 illustrates the case where a pair of connected phenotypes map to two common QTL.

We point out that the original goal of SCHADT *et al.* (2005) was to determine the causal direction between a pair of phenotypes. Here, we test the direct *vs.* the indirect effect of a common QTL on two phenotypes. We note in the DISCUSSION how causal direction inference for our method can be extended to orient edges that have only common QTL.
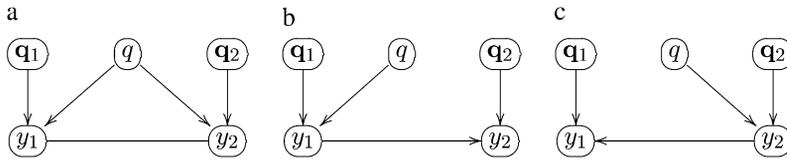
a   b   c

$\mathbf{q}_1$ q $\mathbf{q}_2$ $\mathbf{q}_1$ q $\mathbf{q}_2$ $\mathbf{q}_1$ q $\mathbf{q}_2$

y_1 —— y_2 y_1 ——→ y_2 y_1 ←—— y_2

FIGURE 2.—Distinguishing direct and indirect effects of a common QTL. Suppose that QTL mapping of phenotype y_1 detected QTL q and $\mathbf{q}_1$ and mapping of phenotype y_2 detected the common QTL q plus QTL $\mathbf{q}_2$. A strong QTL directly affecting an upstream trait may also be (incorrectly) detected as a QTL for a downstream phenotype. To resolve this situation we apply a generalization of SCHADT *et al.* (2005), allowing for multiple QTL, where we score the three models above using BIC or AIC. Model a supports both traits being directly affected by the common QTL q. Model b implies that q directly affects y_1 but should not be included as a QTL of phenotype y_2. Model c supports the reverse situation. Observe that the assumption behind model a is that the correlation between y_1 and y_2 can be explained by the common QTL q, in addition to common environmental influences or other shared loci.

Preprocessing the undirected graph: Whenever a pair of connected phenotypes have a common QTL, as in the situation described in Figure 2a, we further test the null hypothesis of no partial correlation between the phenotypes conditional on the common QTL. If we accept the null, we drop the edge between the phenotypes since their correlation is caused by the pleiotropic effect of the common QTL.

Edge orientation: We orient causal edges between all pairs of connected phenotypes in an undirected network using associated multiple QTL to break likelihood equivalence. The QTL are assumed to come from earlier gene mapping of phenotypes. Edge orientation among pairs of phenotypes is based on a linear regression model with phenotypes regressed on QTL genotypes and on additive or interacting covariates such as sex, age, and other phenotypes. We orient each edge, conditioning on all other nodes (phenotypic, genotypic, or covariate) that are connected to that edge. For each edge, we evaluate a LOD score comparing the two possible orientations. We orient the edge in favor of the direction with the higher likelihood in the ratio. For the toy network presented in the Introducton this ratio is

$$LOD = \log_{10}\left\{\frac{\prod_{i=1}^{n} f(y_{1i}\,|\,\mathbf{q}_{1i})f(y_{2i}\,|\,y_{1i},\mathbf{q}_{2i})}{\prod_{i=1}^{n} f(y_{2i}\,|\,\mathbf{q}_{2i})f(y_{1i}\,|\,y_{2i},\mathbf{q}_{1i})}\right\}$$
$$= LOD_1 + LOD_{2|1} - LOD_2 - LOD_{1|2},$$

where $f()$ represents the predictive density, that is, the sampling model with parameters replaced by the corresponding maximum-likelihood estimates, and

$$LOD_1 = \log_{10}\left\{\prod_{i=1}^{n}\frac{f(y_{1i}\,|\,\mathbf{q}_{1i})}{f(y_{1i})}\right\},$$

$$LOD_{2|1} = \log_{10}\left\{\prod_{i=1}^{n}\frac{f(y_{2i}\,|\,y_{1i},\mathbf{q}_{2i})}{f(y_{2i})}\right\},$$

$$LOD_2 = \log_{10}\left\{\prod_{i=1}^{n}\frac{f(y_{2i}\,|\,\mathbf{q}_{2i})}{f(y_{2i})}\right\},$$

$$LOD_{1|2} = \log_{10}\left\{\prod_{i=1}^{n}\frac{f(y_{1i}\,|\,y_{2i},\mathbf{q}_{1i})}{f(y_{1i})}\right\},$$

represent standard LOD scores comparing the "full" model (a multiple-QTL model possibly containing covariates and interactions) with the "reduced" model (no QTL or covariates).

In more complex networks we orient edges in two steps: (1) build an initial directed network orienting each edge as above and (2) recompute the LOD score for each edge, connecting a pair of phenotype nodes by conditioning on all other phenotypes causative to either or both nodes. We repeat the second step for all edges until no edge switches direction and this involves some iteration to find the best orientation across the entire graph. With four phenotypes, an initial directed network might look like ① ——→ ② ——→ ③ ←—— ④. The second step recomputes the LOD score for the edge between nodes 2 and 3, conditioning, respectively, on nodes 1 and 4. Similarly, we recompute the LOD score for the 3–4 edge, conditioning on node 2 for phenotype 3. If the direction of the 2–3 edge switches, the graph becomes ① ——→ ② ←—— ③ ←—— ④. We then recompute the LOD scores for the 1–2 edge, using node 3 as a covariate for 2, and for the 2–3 edge conditioning, respectively, on 1 and 4.

In complex networks this algorithm may find more than one solution. That is, starting the algorithm from a different edge ordering may yield a different graph. To get the best solution we (1) rerun this algorithm using different initial edges to get all possible solutions, (2) score each solution using a likelihood-based overall measure of fit, and (3) select the graph with the highest score. This algorithm may oscillate between two graphs changing directions on some edges. In this case we choose the graph with the highest measure of fit.

QDG algorithm:
1. Construct an UDG (PC skeleton or an association graph).
2. Use QTL genotypes to direct each edge in the UDG. Call it DG.
3. Randomly choose an ordering of all edges in the DG. Call this list ODG.
4. For an edge in the ODG recompute the direction LOD score using all other causative phenotypes to either or both nodes, according to the DG, in addition to the QTL genotypes. If the direction changes, update the DG and move to the next edge in the ODG.
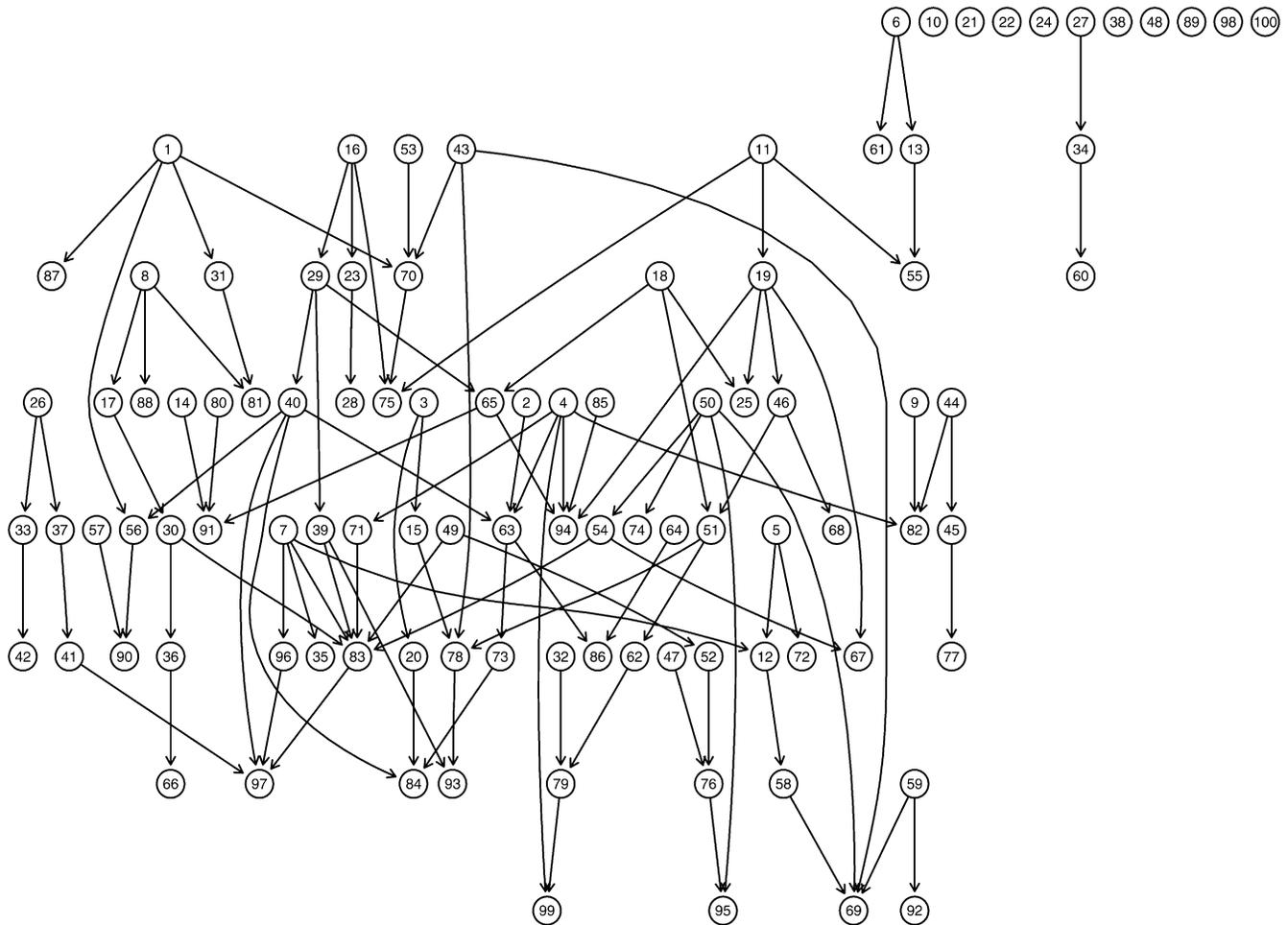
FIGURE 3.—Randomly sampled sparse directed acyclic graph (DAG) composed of 100 nodes (phenotypes) connected by 107 edges, using the randomDAG function from the pcalg R package (R DEVELOPMENT CORE TEAM 2006). We generated data according to this network, adopting two or three QTL (not shown) per phenotype. This network demonstrates many features that can be inferred with varying degrees of difficulty by the PC algorithm. For instance, nodes organized in an unshielded collider pattern (SHIPLEY 2002) such as $x \to z \leftarrow y$ are easier to direct than nodes organized in a bifurcation or line pattern such as $x \leftarrow z \to y$ and $x \to z \to y$. Figure 8 compares the performances of the QDG and PC algorithms for nodes involved in unshielded collider structures and all other remaining patterns pooled together. In supplemental Figures S2 and S3 we highlighted all nodes involved in unshielded collider patterns.

5. Repeat step 4 until no more edges change direction. The corresponding directed graph is one solution. If step 4 keeps oscillating between different graphs without converging to a solution in 30 steps, we include all distinct graphs as solutions.
6. Repeat steps 3, 4, and 5 1000 times and store all different solutions.
7. Score all solutions in step 6, using a likelihood-based measure of fit of the whole graph, and choose the graph with the highest score.

We point out that for each solution of this algorithm, the direction of each edge connecting two nodes is computed conditional on the parents of the nodes. While we start the QDG as a pairwise algorithm (step 2), the last steps actually represent a multivariate algorithm, where we perform local computations of direction on subsets of the graph. An R package (R DEVELOPMENT

CORE TEAM 2006) implementing our procedure is under development and we intend to make it available at Bioconductor.

Simulations: *One hundred phenotypes network:* In each simulation experiment we generated synthetic data according to the network in Figure 3, using linear regression models. Each phenotype node was affected by two or three QTL (not shown in Figure 3), and we allowed only additive genetic effects. The QTL for each phenotype were randomly selected among 200 markers, with 10 markers unevenly distributed on each of 20 autosomes. We allowed different phenotypes to potentially share common QTL. For each phenotype, the regression coefficients with other phenotypes were chosen uniformly between 0.5 and 1; residual standard deviations were chosen between 0.1 and 0.5. The regression coefficient of the phenotype on the QTL ranged from 0.2 to 0.6. More specifically, we assumed
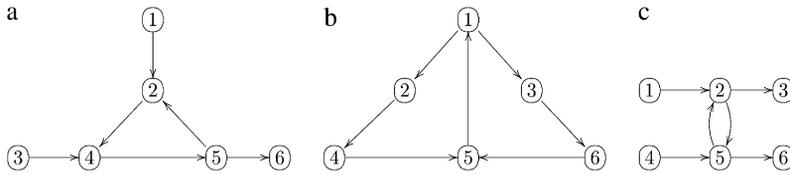
FIGURE 4.—Cyclic graphs. (a) A single three-node cycle. (b) Two neighboring four-node cycles. (c) A reciprocal interaction cycle between nodes 2 and 5. This graph has two pairs of nodes $(1, 5)$ and $(2, 4)$ that are not directly connected but are d-connected (PEARL 1988).

intralocus additivity (no dominance) and an increment of 0.1 per allele. The coefficients for genotype *aa* were drawn uniformly between 0.2 and 0.4. The coefficients for genotypes *Aa* and *AA* were calculated by adding 0.1 and 0.2, respectively. For each simulation experiment, we generated 100 realizations and applied the QDG and PC algorithms to infer causal directions for the edges of the skeleton obtained by the PC-skeleton algorithm.

Cyclic networks: For Figure 4, a and b, we generated data sets with 100 and 200 individuals, fixing regression coefficients at 0.5, error variances at 0.025, and the QTL effects at 0.2, 0.3, and 0.4 for the three F_2 genotypes. We used a Gibbs sampling scheme to generate 100 data sets in each simulation study. For Figure 4c, we adopted a sample size of 100 and parameter values as above with the exception of the reciprocal parameters β_{25} and β_{52}. We used a burn-in of 2000 for the Gibbs sampler and inspection of Markov chains showed good mixing.

Permutation tests for directions: *P*-values for the LOD scores are computed using 10,000 permutations. To get the null distribution for the direction we permute blocks of data in such a fashion that we break all the connections that affect the direction, keeping the others intact. As an example, consider Figure 5 and suppose we are interested in computing a permutation *P*-value for the arrow pointing from phenotype y_1 to phenotype y_2 conditional on covariates x and z. Here, x represents a covariate of both phenotypes (could be another phenotype, age, sex, etc.). z is a covariate only of y_1. $\mathbf{q}_1$ and $\mathbf{q}_2$ correspond to sets of QTL affecting y_1 and y_2, respectively. To break the connections (brk) that affect the direction of an edge, we permute the corresponding pair of nodes (and their common covariates) as a block. Thus in Figure 5a we permute (y_1, y_2, x) as a block breaking the connections with z, $\mathbf{q}_1$, and $\mathbf{q}_2$. If instead we keep the connection between z and y_1 by permuting only

$\mathbf{q}_1$ and $\mathbf{q}_2$ (Figure 5b), we would not break all the connections affecting the directionality [resulting models $f(y_1 \mid z, x)f(y_2 \mid y_1, x)$ and $f(y_1 \mid y_2, z, x)f(y_2 \mid x)$ are not likelihood equivalent and can be used to infer the direction between y_1 and y_2]. Finally, the common covariate x to y_1 and y_2 improves the fit of the linear model, even though it does not directly affect the direction between y_1 and y_2 since $f(y_1 \mid x)f(y_2 \mid y_1, x) = f(y_1, y_2 \mid x) = f(y_1 \mid y_2, x)f(y_2 \mid x)$.

Simulating data from cyclic networks: Consider the set of structural equations

$$y_k = \sum_u \theta_{k,q_u} + \sum_{v:y_v \in pa(y_k)} \beta_{kv} y_v + \epsilon_k, \qquad (2)$$

where $pa(y_k)$ represents the set of parental nodes of y_k, that is, the set of phenotypes that directly affect y_k; $\epsilon_k \sim N(0, \sigma_k^2)$ are independent error terms; and $q_u = 1, 2, 3$, where u indexes all QTL associated with phenotype y_k.

Let $\boldsymbol{\epsilon}$ represent the vector of error terms. Using the Jacobian transformation from $\boldsymbol{\epsilon} \rightarrow \mathbf{y}$, we get the joint density of the phenotypes conditional on the QTL genotypes.

We simulate a data set of n individuals from a cyclic network in two steps. First, we simulate the QTL genotype data. Next, we generate phenotype data, conditional on the QTL genotype data, using a Gibbs sampling scheme. The distribution of each phenotype conditional on the remaining phenotypes, QTL genotypes, and parameters is univariate normal. Because each individual may have a unique multilocus genotype and the phenotype distributions are conditional on the QTL genotypes, we run separate Markov chains to generate each of the n data points.

Distinguishing graphs inside an equivalence class: In the Introduction we demonstrated how the incorporation of driving QTL allows for causal direction inference in a network composed by two phenotypes linked by an edge. This result readily extends to more complex networks belonging to an equivalent class. Suppose,
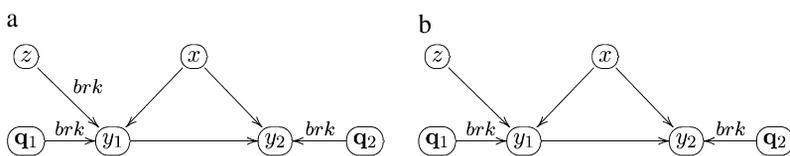

FIGURE 5.—Preserved neighborhoods for permutation *P*-value computation. The *P*-value for the arrow pointing from phenotype y_1 to phenotype y_2 depends on edges to neighboring nodes. x represents a covariate of both phenotypes (could be another phenotype, age, sex, etc.). z is a covariate of y_1. $\mathbf{q}_1$ and $\mathbf{q}_2$ correspond to sets of QTL affecting y_1 and y_2, respectively. To break the connections (brk) that affect direction of an edge, we permute the corresponding pair of nodes (and their common covariates) as a block. In a we permute (y_1, y_2, x) as a block breaking the connections with z, $\mathbf{q}_1$, and $\mathbf{q}_2$; in b we incorrectly keep z in the permutation block. We keep the connections to a common covariate x to y_1 and y_2 in because of its role in improving the fit of the linear model.

for example, that two *N*-node directed acyclic graphs belong to the same equivalence class. Then

$$\prod_{k=1}^{N} f(y_k \mid pa(y_k)) = f(y_1, \ldots, y_N)$$

$$= \prod_{v=1}^{N} f(y_v \mid pa(y_v)),$$

where the indexes k and v represent different factorizations of the same joint predictive density $f(y_1, \ldots, y_N)$. By incorporating the driving QTL we are able to break the likelihood equivalence between these two phenotype networks since

$$\prod_{k=1}^{N} f(y_k \mid \mathbf{q}_k, pa(y_k)) \neq \prod_{v=1}^{N} f(y_v \mid \mathbf{q}_v, pa(y_v)).$$

This result is also true for directed cyclic graphs. Consider, for example, two phenotype networks composed by a cycle with three nodes, ① ⟶ ② ⟶ ③ ⟶ ① and ① ⟵ ② ⟵ ③ ⟵ ①. The predictive densities of the phenotypes agree by application of Bayes' formula:

$$\frac{f(y_1 \mid y_3) f(y_2 \mid y_1) f(y_3 \mid y_2)}{\int_{\mathbf{y}} f(y_1 \mid y_3) f(y_2 \mid y_1) f(y_3 \mid y_2) \, d\mathbf{y}}$$

$$= \frac{f(y_1 \mid y_2) f(y_2 \mid y_3) f(y_3 \mid y_1)}{\int_{\mathbf{y}} f(y_1 \mid y_2) f(y_2 \mid y_3) f(y_3 \mid y_1) \, d\mathbf{y}}.$$

However, when we include the respective driving QTL,

$$f(y_1 \mid y_3, \mathbf{q}_1) f(y_2 \mid y_1, \mathbf{q}_2) f(y_3 \mid y_2, \mathbf{q}_3)$$

$$= \frac{f(y_3 \mid y_1, \mathbf{q}_1) f(y_1 \mid \mathbf{q}_1)}{f(y_3 \mid \mathbf{q}_1)} \frac{f(y_1 \mid y_2, \mathbf{q}_2) f(y_2 \mid \mathbf{q}_2)}{f(y_1 \mid \mathbf{q}_2)}$$

$$\times \frac{f(y_2 \mid y_3, \mathbf{q}_3) f(y_3 \mid \mathbf{q}_3)}{f(y_2 \mid \mathbf{q}_3)}$$

$$\neq \frac{f(y_2 \mid y_1, \mathbf{q}_1) f(y_1 \mid \mathbf{q}_1)}{f(y_2 \mid \mathbf{q}_1)} \frac{f(y_3 \mid y_2, \mathbf{q}_2) f(y_2 \mid \mathbf{q}_2)}{f(y_3 \mid \mathbf{q}_2)}$$

$$\times \frac{f(y_1 \mid y_3, \mathbf{q}_3) f(y_3 \mid \mathbf{q}_3)}{f(y_1 \mid \mathbf{q}_3)}$$

$$= f(y_1 \mid y_2, \mathbf{q}_1) f(y_2 \mid y_3, \mathbf{q}_2) f(y_3 \mid y_1, \mathbf{q}_3),$$

and the graphs are no longer likelihood equivalent.

Thus incorporating driving QTL in the analysis allows us to distinguish between phenotype networks that would otherwise be undistinguishable. Therefore, if the QDG algorithm converges to solutions belonging to the same equivalence class (considering the phenotypes only), it is still able to distinguish between then and reconstructs a single network.

Likelihood-based overall measure of fit: For directed acyclic graphs the joint likelihood factors nicely and maximum-likelihood analysis is straightforward. Therefore, we adopted the maximized log likelihood of the whole graph as an overall measure of fit in this case.

For directed cyclic graphs, the likelihood cannot be completely factored (because of the normalization con-
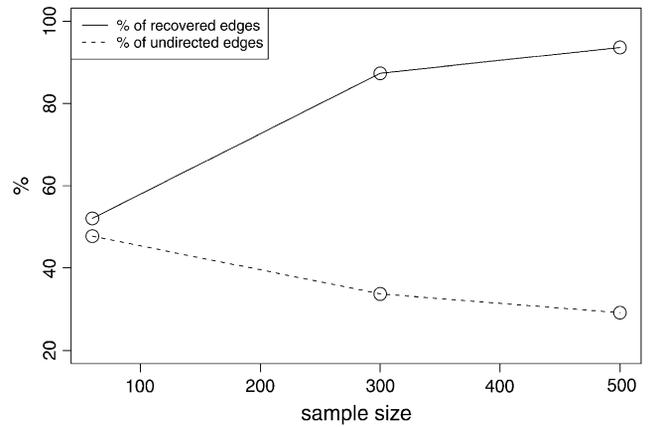

Figure 6.—Average percentage of discovered edges (PC skeleton) and average percentage of edges that could not be directed by the PC algorithm relative to the discovered edges for the network in Figure 3. Averages were computed across all edges in the network.

stant) and maximum-likelihood estimators do not have closed analytical forms, requiring numerical techniques for their evaluation. When comparing cyclic networks with reversed directions, we can avoid these computations because the normalization constants agree and we can simply compare the unnormalized likelihoods. To see why that is the case, recall that the QTL enter as a location parameter (θ_{k,q_u}) in the structural Equations 2 defining the cyclic graphs, so that the hypervolume under the *N*-dimensional integral (the inverse of the normalization constant) is not affected by a shifting of the *N*-dimensional curve.

If we need to compare acyclic and cyclic graphs, the normalization constant of the cyclic graph cannot be ignored. But since an acyclic graph cannot be likelihood equivalent to a cyclic one, we can simply fit a path model to the respective phenotype networks (using any structural equation models package) and use the respective maximized log likelihood as an overall measure of fit.

RESULTS

In simulations, we contrast our QDG method with the PC algorithm (Spirtes *et al.* 2000), which can orient some but not all edges in an undirected network using only partial correlations.

Network topology: In all simulation studies, we assessed the performance of the PC skeleton or the UDG algorithm in terms of the average (across all edges in the respective graphs) of the percentage of times the algorithm recovered each of the true edges. Edge recovery rates were strong, increasing with sample size.

One hundred phenotypes network: A complicated network with 100 phenotypes and 107 edges (Figure 3) illustrates how the percentage of edges recovered using the PC skeleton increases with sample size (Figure 6; supplemental Figures S2 and S3). Similar improve-

TABLE 1

Measures of overall performance of the PC skeleton

Sample size	True positive rate	False positive rate	True discovery rate
60	0.5207	0.0007	0.9453
300	0.8733	0.0010	0.9518
500	0.9364	0.0020	0.9122

True positive rate is the number of correctly inferred edges divided by the number of "true edges." False positive rate is the number of incorrectly inferred edges divided by the number of "true gaps." True discovery rate is the number of correctly inferred edges divided by the number of inferred edges. The averages across the 100 simulations are shown.

ments were found for the true positive rate, while false positive rates were uniformly low and true discovery rates were uniformly high (Table 1).

Cyclic networks: We investigated the properties of three cyclic graphs (Figure 4). As sample size increased from 100 to 200, the average percentage of true recovered edges using the PC skeleton improved from 83 to 99% for Figure 4a, and from 78 to 99% for Figure 4b (Tables 2 and 3). Figure 4c had two pairs of nodes, (1, 5) and (2, 4), that had no direct connection but were not d-separated (PEARL 1988). We expected the PC skeleton to detect spurious edges between them, but we found only 2 realizations in 100 with an edge between nodes 1 and 5 and none between 2 and 4 (Tables 4 and 5). Therefore, the PC skeleton successfully recovered important features of cyclic graph topology.

Causal direction inference: *One hundred phenotypes network:* Both QDG and PC algorithms improved in terms of correct causal direction as sample size increased. In all cases, the QDG method got more correct directions (Figure 7a) than the PC algorithm and, except for a sample size of 60, fewer incorrect directions (Figure 7b). Supplemental Figures S2 and S3 show the results per edge.

TABLE 3

Proportion of recovered edges (using PC skeleton) and correct edge direction (using QDG) for Figure 4b

	Recovered edges		Correct direction	
Edge	$n = 100$	$n = 200$	$n = 100$	$n = 200$
$1 \rightarrow 2$	0.50	0.99	0.90	1.00
$1 \rightarrow 3$	0.74	0.99	0.99	1.00
$2 \rightarrow 4$	1.00	1.00	0.95	1.00
$5 \rightarrow 1$	0.64	0.98	1.00	1.00
$3 \rightarrow 6$	0.85	1.00	0.99	1.00
$4 \rightarrow 5$	0.83	1.00	1.00	1.00
$6 \rightarrow 5$	0.96	1.00	0.97	1.00

False positive edges between nodes 2 and 3, nodes 2 and 5, and nodes 3 and 5 were detected, respectively, six, eight, and three times in the simulation study with sample size 100. No false positives were detected with sample size 200.

The network in Figure 3 demonstrates many features that can be inferred with varying degrees of difficulty by the PC algorithm. For instance, nodes organized in an unshielded collider pattern (SHIPLEY 2002) such as $x \rightarrow z \leftarrow y$ are easier to direct than nodes organized in a bifurcation or line pattern such as $x \leftarrow z \rightarrow y$ or $x \rightarrow z \rightarrow y$. In Figure 8, we present the average performances of the QDG and PC algorithms for nodes involved in unshielded collider structures separately from the performances of all other remaining patterns (that we pooled together). As expected, the PC algorithm performed much better in the unshielded collider pattern, relative to all other patterns, in terms of correct causal direction inference and of proportion of undirected edges. The QDG algorithm performed equally well for all patterns and was better than the PC. All nodes involved in unshielded collider patterns were highlighted in supplemental Figures S2 and S3.

The QDG method is robust with respect to the number of QTL per phenotype. We randomly selected only one QTL per phenotype (from the two or three

TABLE 2

Proportion of recovered edges (using the PC skeleton) and correct edge direction (using QDG) for Figure 4a

	Recovered edges		Correct direction	
Edge	$n = 100$	$n = 200$	$n = 100$	$n = 200$
$1 \rightarrow 2$	0.57	1.00	0.96	1.00
$2 \rightarrow 4$	0.95	1.00	0.98	1.00
$5 \rightarrow 2$	0.96	1.00	0.90	0.98
$3 \rightarrow 4$	0.77	1.00	0.97	1.00
$4 \rightarrow 5$	0.78	0.98	0.99	1.00
$5 \rightarrow 6$	1.00	1.00	1.00	1.00

A false positive edge between nodes 1 and 3 was detected in the simulation study with sample size 100. One and two false positive edges between nodes 1 and 3 and 4 and 6, respectively, were detected with sample size 200.

TABLE 4

Proportion of recovered edges (using PC skeleton) and correct edge direction (using QDG) for Figure 4c

	Recovered edges			Correct direction		
Edge	$\beta_{25} = \beta_{52}$	$\beta_{25} < \beta_{52}$	$\beta_{25} > \beta_{52}$	$\beta_{25} = \beta_{52}$	$\beta_{25} < \beta_{52}$	$\beta_{25} > \beta_{52}$
$1 \rightarrow 2$	0.41	0.37	0.23	1.00	1.00	1.00
$2 \rightarrow 3$	1.00	0.99	1.00	0.97	0.99	1.00
$4 \rightarrow 5$	0.62	0.37	0.48	0.94	0.97	0.96
$5 \rightarrow 6$	0.99	1.00	1.00	0.99	0.99	0.99

When $\beta_{25} < \beta_{52}$, the net effect is from phenotype 2 toward phenotype 5, whereas when $\beta_{25} > \beta_{52}$, the net effect is toward phenotype 2. There is no net effect when $\beta_{25} = \beta_{52}$. False positive edges between nodes 2 and 6, 1 and 5, and 3 and 5 were detected six, two, and one times, respectively.

TABLE 5

Proportion of correct direction (using QDG) for the reciprocal nodes 2 and 5 from Figure 4c for different regression coefficient patterns

	Inferred direction	
	$2 \leftarrow 5$	$5 \leftarrow 2$
$\beta_{25} = \beta_{52}$	0.61	0.39
$\beta_{25} < \beta_{52}$	0.13	0.87
$\beta_{25} > \beta_{52}$	0.74	0.26

The proportion of recovered edges (using PC skeleton) was 1 for the three patterns. When $\beta_{25} < \beta_{52}$ the net effect is from phenotype 2 toward phenotype 5, whereas when $\beta_{25} > \beta_{52}$ the net effect is toward phenotype 2. There is no net effect when $\beta_{25} = \beta_{52}$.

QTL used to generate the data) for causal orientation. We found a relatively small decrease in performance when only one QTL was used (Figure 9). While the QDG algorithm required more iterations to converge on a solution when using only one driving QTL per phenotype, there was little loss in percentage of correct directions.
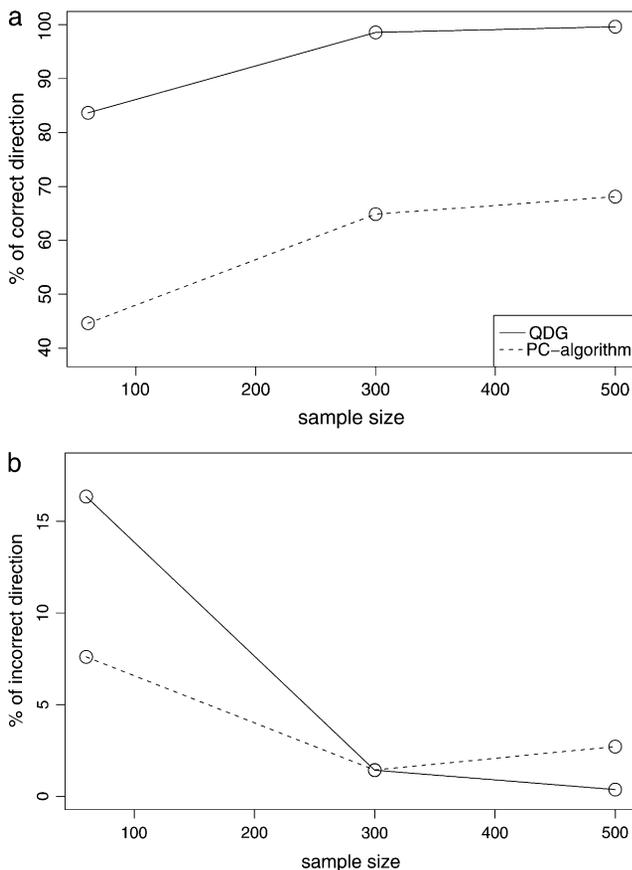

FIGURE 7.—Average percentage of correctly inferred directions (a) and average percentage of incorrectly inferred directions (b) for the network in Figure 3. Averages were computed across all edges in the network.
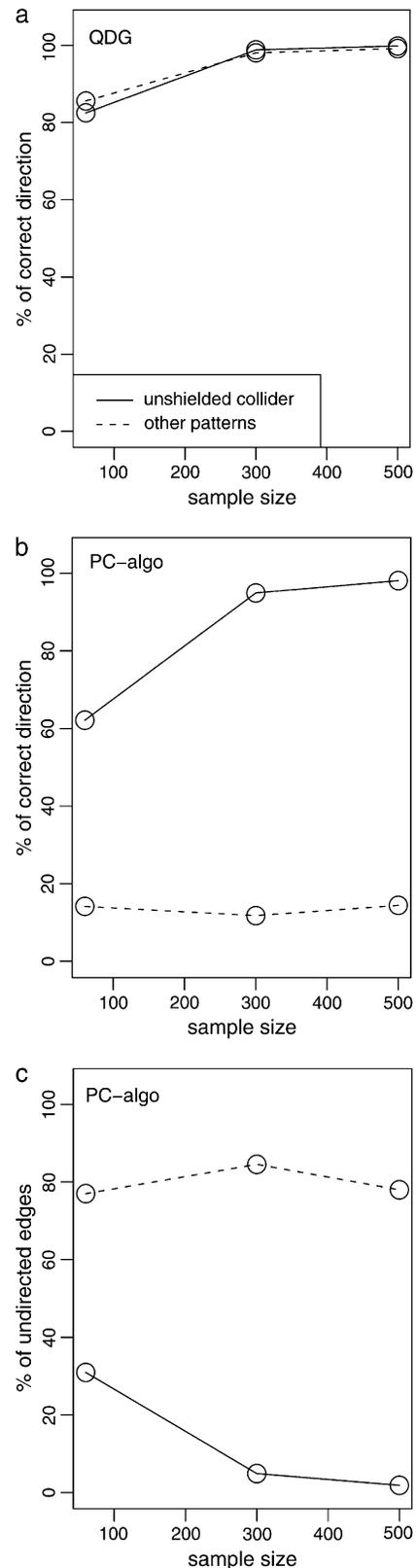

FIGURE 8.—Comparison of the QDG and PC algorithms relative to the 69 edges involved in unshielded colliders and the other 38 edges involved in other patterns. (a and b) The averaged percentage of correct directions (across all nodes belonging to the respective pattern). (c) The averaged percentage of undirected edges in PC alone.
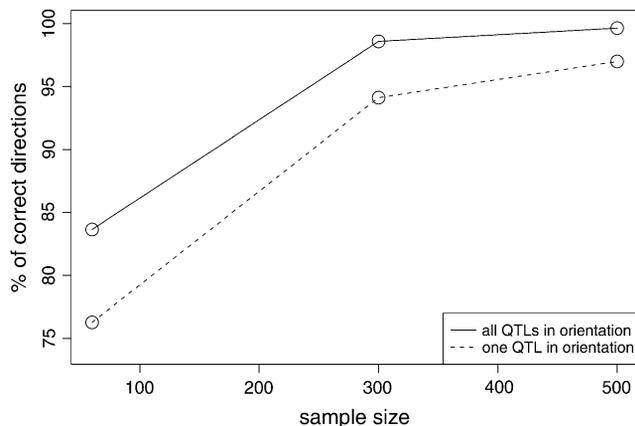
FIGURE 9.—Comparison of the performance of the QDG method when all QTL (two or three per phenotype) were used to infer the directions *vs.* the performance of the QDG method using only one QTL (randomly selected from the two or three QTL used to generate the data). Averages were computed across all edges in the network in Figure 3.

Cyclic networks: For the three small cyclic graphs shown in Figure 4, the QDG inferred the correct direction most of the time. As sample size went from 100 to 200, the average percentage of correct direction relative to the proportion of recovered edges increased from 96 to 99% for Figure 4a (Table 2) and from 97 to 100% for Figure 4b (Table 3).

The simulations for Figure 4c illustrate that even though our method cannot detect reciprocal interactions (*e.g.*, between phenotypes 2 and 5), it can still infer the stronger direction, that is, the direction corresponding to the higher regression coefficient. Tables 4 and 5 present the results of three simulation studies where we adopted the following regression coefficients: (1) $\beta_{25} = \beta_{52} = 0.5$; (2) $\beta_{25} = 0.5$, $\beta_{52} = 0.8$; and (3) $\beta_{25} = 0.8$, $\beta_{52} = 0.5$.

METABOLITE AND GENE EXPRESSION NETWORK

We inferred a causal network among expression and metabolite traits and validated the network with *in vitro* experiments (FERRARA *et al.* 2008). Here we simulate from a connected subset of this inferred network to study its statistical properties.

The causal network was constructed using *in vivo* gene expression and metabolite data from a sample of 58 mice from an $F_2 - ob/ob$ population generated from the C57BL/6J (B6) and BTBR founder strains (LAN *et al.* 2006). The B6 and BTBR strains, when made obese, differ in diabetes susceptibility (STOEHR *et al.* 2000); B6-*ob/ob* mice are diabetes resistant, whereas BTBR-*ob/ob* mice develop severe diabetes. The included transcripts had a high positive correlation with glutamine plus glutamate (Glx) and/or comapped eQTL and mQTL (FERRARA *et al.* 2008). Our QDG algorithm with random-

edge orderings converged to two solutions. Figure 10 shows the major connected subset of this network.

We simulated 1000 realizations from the network depicted in Figure 10, inferring the correct edge directions for recovered edges on average 83% of the time. The average percentage of true recovered edges was 75%. All edges except $Slc1a2 \rightarrow Glx$ were recovered at least 70% of the time, with correct direction at least 72% of the time (Figure 10). False edges were detected at rates <2% (Table 6). In the construction and simulations, age was used as an additive covariate and we allowed for sex-by-genotype interactions. Estimates from the real data were used as the parameter values in the simulations and we adopted $\alpha = 0.05$ for the UDG algorithm.

Our network predicted six edge directions (Figure 10), which we tested by isolating hepatocytes from the B6 and BTBR parental strains. We used quantitative real-time PCR to measure gene expression changes in response to the 10-mm glutamine supplement *in vitro* (normalization to *Actb*). Our results indicated that glutamine does significantly change expression of *Agxt, Arg1*, and *Pck1* in both strains, as predicted by the network. Glutamine increased expression of *Ass1*, which is consistent with previous studies showing that glutamine increases urea cycle enzyme expression. Glutamine decreased expression of *Slc1a2* in both strains and decreased *Ivd* in the B6 strain, contradicting our best network, but our confidence in these directions was marginal (FERRARA *et al.* 2008). Thus we used our inferred network for hypothesis generation, and we confirmed four of six predictions with subsequent laboratory work.

DISCUSSION

In genetical genomics or eQTL studies, we are increasingly interested in inferring causal networks for sets of phenotypes that map to coincident genomic regions. We typically believe that there is a "master regulator" and that most comapping is due to indirect effects. Thus our objective is to untangle the direct and indirect effects of QTL and phenotypes. We provide a novel method that can first infer what phenotypes have proximal relationships (undirected graphs) and then orient those edges to infer causal relationships, leading to a unified causal network.

In recent years, there has been a large effort to model phenotypic networks. Several approaches have been employed to construct association networks (SCHADT and LUM 2006). H. LI *et al.* (2006) based association networks on partial correlations among phenotypes with significant eQTL; ZHANG and HORVATH (2005) clustered phenotypes into modules on the basis of a power function of pairwise correlation; LAN *et al.* (2006) used phenotype correlation with strong eQTL to identify sets of functionally related expression traits.

Much work has been done to infer causal networks. R. LI *et al.* (2006) built causal networks among phenotypes
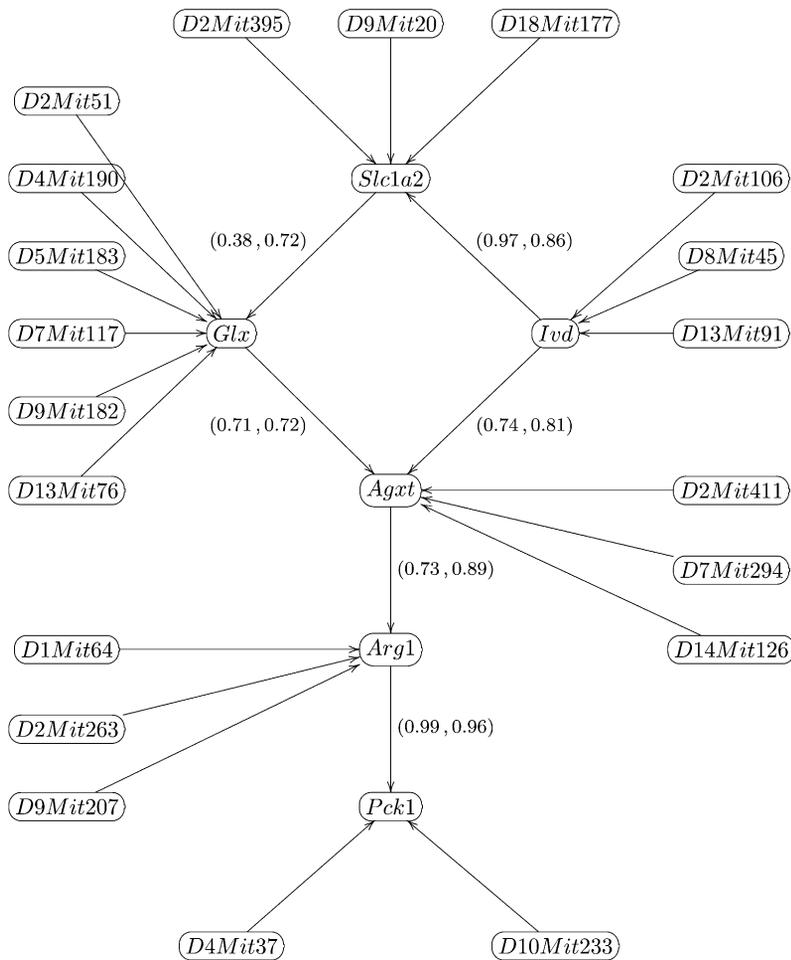
FIGURE 10.—Connected subset of the metabolite and gene expression causal network presented in FERRARA *et al.* (2008). Age and sex (not shown) were used as covariates in the causal orientation. Edges from *Slc1a2* were reversed in the second-best network. Values on edges are the proportion of times the edge was recovered and the proportion of correct direction relative to the recovered edges.

and genotypes, using structural equation modeling involving a hierarchy of interconnected linear models. Their model selection procedure was computationally intensive, searching over a wide class of possible models. In the genetical genomics setting, ZHU *et al.* (2004) extended standard Bayesian network reconstruction methods to include detected major QTL driving gene transcript abundance. Due to computational restrictions, only a few QTL were included for each expression

TABLE 6

Proportion of false positive (PFP) edges detected by the UDG algorithm for the network in Figure 10

Node 1	Node 2	PFP
Slc1a2	Agxt	0.019
Slc1a2	Arg1	0.001
Slc1a2	Pck1	0.006
Glx	Ivd	0.002
Glx	Arg1	0.008
Glx	Pck1	0.015
Ivd	Arg1	0.007
Ivd	Pck1	0.004
Agxt	Pck1	0.006

trait. Further, no feedback cycles were allowed in the model. SCHADT *et al.* (2005) related expression to clinical traits, providing a conditional probability-based approach to distinguish whether expression changes associated with a QTL were causal, reactive, or independent from clinical changes. BING and HOESCHELE (2005) proposed a strategy to build a network extending from candidate genes in eQTL genomic intervals to *cis*- and *trans*-regulated mRNAs. However, their approach required identifying all candidate genes in genomic intervals and did not infer direction between phenotypes (expressed genes) except from a *cis*-acting gene affecting the expression of another gene in *trans*. KULP and JAGALUR (2006) built partially directed networks of expression phenotypes using a gene of interest as a seed and including *trans*-acting regulator–target pairs identified by gene mapping. KEURENTJES *et al.* (2007) used the BING and HOESCHELE (2005) approach in Arabidopsis to rediscover a well-studied regulatory network associated with flowering time. However, none of these genetical genomics methods can infer the causal direction between two phenotypes that do not share a common QTL or when one expressed gene is not physically located within the QTL support interval of other transcripts.

We have proposed a simple method for causal network construction that extends existing causal network methods by accommodating phenotypes that have no common QTL. Our QDG method does not require that any phenotype correspond to a gene in a QTL support interval for another phenotype. We have shown how to infer causal direction in an association network of phenotypes by conditioning on causal QTL, even when graphs include cycles. This method is part of a larger strategy to reveal phenotypic networks. We begin with a feature selection to reduce from thousands of phenotypes to a manageable number (*cf.* ZHANG and HORVATH 2005; LAN *et al.* 2006) followed by a focus on a particular functional group. We then use QTL mapping methods to identify major QTL of phenotypes in this functional group (YI *et al.* 2005, 2007). Finally, we use the approach described in this article to infer a phenotypic network for this functional group.

This proposed method can handle multiple QTL and covariates (additive or interactive) and has computing time proportional to the number of edges. It can be readily extended to include discrete driving factors other than genotypes, such as sex or treatment group. Furthermore, this method can be applied to a wide variety of phenotypes, and here we demonstrate its application to gene expression and metabolite data. The incorporation of genotypes (or discrete driving factors) allows the reconstruction of a single directed network of phenotypes (see METHODS). Previous work in gene expression network reconstruction using Bayesian networks (FRIEDMAN *et al.* 2000) was limited to the inference of equivalence classes, that is, a set of networks that have the same likelihood. Another noteworthy advantage is that the QDG does not require sequence information and can therefore be useful in the discovery process for a broader range of genetical genomics experiments. However, we can use sequence data to improve models in a similar fashion to KEURENTJES *et al.* (2007).

Our QDG method shows superior performance over the PC algorithm (SPIRTES *et al.* 2000) to infer correct direction of edges that bifurcate from a node. That is, the pleiotropic effect of one gene expression phenotype on two other phenotypes is readily detected by our QDG method, but missed by the PC algorithm. Both methods do well in detecting unshielded collider patterns, which correspond to the interactive effect of two gene expression phenotypes on a third phenotype. The enhanced ability of the QDG algorithm to infer both pleiotropy and epistasis will be valuable as eQTL and mQTL studies become more prevalent in experimental crosses.

In preliminary simulation studies (using data generated from Figure 3) we also investigated the performance of the method proposed by OPGEN-RHEIN and STRIMMER (2007) to partially direct graph skeletons. Because their method was unable to recover any of the directions in our preliminary studies, we did not pursue further comparisons.

In this article we assume that a set of multiple QTL affecting each of the phenotypes has been previously determined. With sample sizes in the order of 300–500 we expect to have enough power to detect QTL for a set of functionally related phenotypes appropriate for our causal network methods. Furthermore, phenotypes with no detectable QTL can still be included in the analysis. Edge direction involving these nodes is largely determined by neighboring phenotypes and their QTL. For instance, if we did not detect any QTL for phenotype y_1 we still can orient the edge since the causal models are not likelihood equivalent, $f(y_2 \mid y_1, q_2) f(y_1) \neq f(y_1 \mid y_2) f(y_2 \mid q_2)$. Finally, we point out that previous studies have considered applying relaxed LOD score thresholds to detect minor QTL that would have otherwise been missed and have successfully uncovered interesting pathways (CHESLER *et al.* 2005; LAN *et al.* 2006). We anticipate that such a strategy would work effectively with our proposed method.

Distinct QTL are necessary to infer causal direction in a phenotype network. We can orient an edge between a pair of phenotypes with common QTL if there is at least one additional QTL or causal phenotype that affects one of the pair but not the other. However, two phenotypes with only one common QTL have causal models that are likelihood equivalent, $f(q) f(y_1 \mid q) f(y_2 \mid y_1, q) = f(\mathbf{q}) f(y_2 \mid q) f(y_1 \mid y_2, q)$. In these cases, we can modify the network to determine directionality by breaking genotype-phenotype edges (SCHADT *et al.* 2005), that is, in addition to the pleiotropic causal model, we also consider models $(q) \longrightarrow (y_1) \longrightarrow (y_2)$ and $(q) \longrightarrow (y_2) \longrightarrow (y_1)$. Note, however, that if the data best support the pleiotropic model, we cannot infer a direction. Further, our QDG method determines a *P*-value to orient each edge. If that *P*-value is large, then we have little confidence in the direction. Such edges may be interpreted as undirected.

Our METHODS section describes an application of the SCHADT *et al.* (2005) approach to test whether a QTL has a direct or an indirect effect on a phenotype. Furthermore, we are currently investigating a strategy to integrate QTL as nodes in the undirected dependency graph, allowing the algorithm to automatically test for direct and indirect QTL effects. We feel, however, that these further investigations are out of the scope of this article.

As is the case with any other network inference method, UDGs (and PC skeletons) can show false edges due to latent variables, including other phenotypes and genotypes. The absence of hidden variables is a common assumption in correlation-based network inference methods (*e.g.*, DE LA FUENTE *et al.* 2004; SCHAFER and STRIMMER 2005a,b). Potential problems caused by hidden phenotypes include erroneous inference of direct causation between phenotypes where there is either (a) indirect causation through an intermediate missing phenotype or (b) independence of the phenotypes examined given a hidden variable (DE LA FUENTE *et al.* 2004).

While our approach can detect circular paths involving three or more phenotypes, it merges reciprocal interactions between two phenotypes, $\textcircled{1} \rightleftharpoons \textcircled{2}$. In these cases, a single edge is oriented in the direction of higher strength. Currently we are investigating a Bayesian approach to formally incorporate prior information about network topology and edge orientation.

In conclusion, we have presented a novel, efficient causal network construction method, QDG, with four main advantages: (1) we reduce computation time by avoiding a search over the entire graph space; (2) we screen out edges associated with partial correlation that are not causative in directed acyclic graphs; (3) we can infer feedback loops, which are common in genetic, metabolic, and biochemical networks; and (4) we can infer a single network, instead of an equivalence class of networks. The causal phenotype network method developed here may have broad applicability across a range of population structures. This can include, but is not limited to, inbred crosses, outbred crosses, and natural populations. The causal networks that emerge can form the basis for hypotheses about critical pathways that can be individually tested using knockouts or overexpression of genes.

LITERATURE CITED

BING, N., and I. HOESCHELE, 2005   Genetical genomic analysis of a yeast segregant population for transcription network inference. Genetics **170**: 533–542.

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

CHESLER, E. J., L. LU, S. SHOU, Y. QU, J. GU *et al.*, 2005   Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat. Genet. **37**: 233–242.

DE LA FUENTE, A., N. BING, I. HOESCHELE and P. MENDES, 2004   Discovery of meaningful associations in genomic data using partial correlation coefficients. Bioinformatics **20**: 3565–3574.

FERRARA, C. T., P. WANG, E. CHAIBUB NETO, R. D. STEVENS, J. R. BAIN *et al.*, 2008   Genetic networks of liver metabolism revealed by integration of metabolic and transcriptomic profiling. PLoS Genet. **4**: e1000034.

FRIEDMAN, N., M. LINIAL, I. NACHMAN and D. PE'ER, 2000   Using Bayesian networks to analyze expression data. J. Comput. Biol. **7**: 601–620.

JANSEN, R. C., and J. P. NAP, 2001   Genetical genomics: the added value of segregation. Trends Genet. **17**: 388–391.

KALISCH, M., and P. BUHLMANN, 2007   Estimating high-dimensional directed acyclic graphs with the PC-algorithm. J. Mach. Learn. Res. **8**: 613–636.

KENDZIORSKI, C., and P. WANG, 2006   A review of statistical methods for expression quantitative trait loci mapping. Mamm. Genome **17**: 509–517.

KEURENTJES, J. J. B., J. FU, I. R. TERPSTRA, J. M. GARCIA, G. VAN DEN ACKERVEKEN *et al.*, 2007   Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. Proc. Natl. Acad. Sci. USA **104**: 1708–1713.

KULP, D. C., and M. JAGALUR, 2006   Causal inference of regulator-target pairs by gene mapping of expression phenotypes. BMC Genomics **7**: 125.

LAN, H., M. CHEN, J. E. BYERS, B. S. YANDELL, D. S. STAPLETON *et al.*, 2006   Combined expression trait correlations and expression quantitative trait locus mapping. PLoS Genet. **2**: e6.

LI, H., H. CHEN, L. BAO, K. F. MANLY, E. J. CHESLER *et al.*, 2006   Integrative genetic analysis of transcription modules: towards filling the gap between genetic loci and inherited traits. Hum. Mol. Genet. **15**: 481–492.

LI, R., S. W. TSAIH, K. SHOCKLEY, I. M. STYLIONOU, J. WERGEDAL *et al.*, 2006   Structural model analysis of multiple quantitative traits. PLoS Genet. **2**: 1046–1057.

OPGEN-RHEIN, R., and K. STRIMMER, 2007   From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Syst. Biol. **1**: 37.

PEARL, J., 1988   *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kauffman, San Mateo, CA.

R DEVELOPMENT CORE TEAM, 2006   *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.

SCHADT, E. E. and P. Y. LUM, 2006   Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Reverse engineering gene networks to identify key drivers of complex diseases phenotypes. J. Lipid Res. **47**: 2601–2613.

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

SCHADT, E. E., J. LAMB, X. YANG, J. ZHU, S. EDWARDS *et al.*, 2005   An integrative genomics approach to infer causal associations between gene expression and disease. Nat. Genet. **37**: 710–717.

SCHAFER, J., and K. STRIMMER, 2005a   An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics **21**: 754–764.

SCHAFER, J., and K. STRIMMER, 2005b   A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat. Appl. Genet. Mol. Biol. **4**: 32.

SHIPLEY, B., 2002   *Cause and Correlation in Biology.* Cambridge University Press, Cambridge/London/New York.

SPIRTES, P., C. GLYMOUR and R. SCHEINES, 2000   *Causation, Prediction, and Search,* Ed. 2. MIT Press, Cambridge, MA.

STOEHR, J. P., S. T. NADLER, K. L. SCHUELER, M. E. RABAGLIA, B. S. YANDELL *et al.*, 2000   Genetic obesity unmasks nonlinear interactions between murine type 2 diabetes susceptibility loci. Diabetes **49**: 1946–1954.

YI, N., B. S. YANDELL, G. A. CHURCHILL, D. B. ALLISON, E. J. EISEN *et al.*, 2005   Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. Genetics **170**: 1333–1344.

YI, N., D. SHRINER, S. BANERJEE, T. MEHTA, D. POMP *et al.*, 2007   Efficient strategies for Bayesian mapping of genome-wide interacting QTL. Genetics **176**: 1865–1877.

ZHANG, B., and S. HORVATH, 2005   A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. **4**: Article 17.

ZHU, J., P. Y. LUM, J. LAMB, D. GUHA THAKURTA, 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.

Communicating editor: R. W. DOERGE