DEPARTMENT OF STATISTICS
University of Wisconsin
1300 University Avenue
Madison, WI 53706

TECHNICAL REPORT NO. 1146

December 1, 2008

Revised on January 27, 2009

Causal Graphical Models in System Genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes[1]

Elias Chaibub Neto[2]

Mark P Keller[3]

Alan D Attie[3]

Brian S Yandell[2,4]

[1]This work was supported by CNPq Brazil (ECN); NIDDK grants DK66369, DK58037 and DK06639 (ADA, MPK, BSY, ECN); and by NIGMS grants PA02110 and GM069430-01A2 (BSY).

[2]Department of Statistics, University of Wisconsin - Madison.

[3]Department of Biochemistry, University of Wisconsin - Madison.

[4]Department of Horticulture, University of Wisconsin - Madison.

Causal Graphical Models in Systems Genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes

Elias Chaibub Neto, Mark P. Keller, Alan D. Attie and Brian S. Yandell

January 27, 2009

Abstract

Causal inference approaches in systems genetics exploit quantitative trait loci (QTL) genotypes to infer causal relationships among phenotypes. The genetic architecture of each phenotype may be complex, and poorly estimated genetic architectures may compromise the inference of causal relationships among phenotypes. Existing methods assume QTLs are known or inferred without regard to the phenotype network structure. In this paper we develop a QTL-driven phenotype network method (QTLnet) to jointly infer a causal phenotype network and associated genetic architecture for sets of correlated phenotypes. Randomization of alleles during meiosis and the unidirectional influence of genotype on phenotype allow the inference of QTLs causal to phenotypes. Causal relationships among phenotypes can be inferred using these QTL nodes, enabling us to distinguish among phenotype networks that would otherwise be distribution equivalent. We jointly model phenotypes and QTLs using homogeneous conditional Gaussian regression models, and we derive a graphical criterion for distribution equivalence. We validate the QTLnet approach in a simulation study. Finally, we illustrate with simulated data and a real example how QTLnet can be used to infer both direct and indirect effects of QTLs and phenotypes that co-map to a genomic region.

1 Introduction

In the past few years it has been recognized that genetics can be used to establish causal relationships among phenotypes organized in networks (Schadt et al. 2005, Kulp and Jagalur 2006, Li et al. 2006, Chen et al. 2007, Liu et al. 2008, Aten et al. 2008, Chaibub Neto et al. 2008). These approaches aim to generate hypothesis about causal relationships among phenotypes involved in biological pathways underlying complex diseases such as diabetes. A key element in these methods is the identification of quantitative trait loci (QTLs) that are causal for each phenotype. The genetic architecture of each phenotype, which consists of the locations and effects of detectable QTLs, may be complex. Poorly estimated genetic

architectures may compromise the inference of causal relationships among phenotypes. Existing methods that estimate QTLs from genome scans that ignore causal phenotypes bias the genetic architecture by incorrectly inferring QTLs that have indirect effects.

In this paper we propose a novel framework for the joint inference of phenotype network structure and genetic architecture (QTLnet). We model phenotypes and QTL genotypes jointly using homogeneous conditional Gaussian regression (HCGR) models (Lauritzen 1996). The genetic architecture for each phenotype is inferred conditional on the phenotype network. Because the phenotype network structure is itself unknown, the algorithm iterates between network structure and genetic architecture using a Markov chain Monte Carlo approach. The posterior sample of network structures is summarized by Bayesian model averaging. To the best of our knowledge, no other proposed method explicitly uses an inferred network structure among phenotypes when performing QTL mapping. Tailoring QTL mapping to network structure avoids the false detection of QTLs with indirect effects and improves phenotype network structure inference.

We employ a causal inference framework with components of both randomized experiments and conditional probability. Randomization of alleles during meiosis and the unidirectional influence of genotype on phenotype allow the inference of causal QTLs for phenotypes. Causal relationships among phenotypes can be inferred using these QTL nodes, enabling us to distinguish between networks that would otherwise be distribution equivalent.

We are particularly interested in inferring causal networks relating sets of phenotypes mapping to coincident genomic regions. It is widely asserted that alleged "hotspots" may have a "master regulator" and that most co-mapping is due to indirect effects (Breitling et al. 2008). That is, such a hotspot QTL could influence a single phenotype that is upstream of many others in a causal network; ignoring the phenotype network would result in a perceived hotspot. One objective of our QTLnet method is to sort out the direct and indirect effects of QTLs and phenotypes in such situations.

Section 2 describes the homogeneous conditional Gaussian regression (HCGR) model adopted in this work, while Section 3 derives our causal inference framework and a graphical criterion to determine distribution equivalence. Section 4 shows how a conditional LOD score can formally measure conditional dependence among phenotypes and QTLs and how QTL mapping can be embed in a graphical models framework. Section 5 presents the MCMC approach for QTLnet. Simulation studies in Section 6 validate the QTLnet approach and provide a explanation for some QTL hotspots. Section 7 uses real data to illustrate how QTLnet can be used to infer direct and indirect effects of QTLs and phenotypes that co-map to a genomic region. The discussion in Section 9 puts this work in context of open questions. Proofs of formal results are given in the Supplementary materials.

2 HCGR Genetic Model

In this section, we recast the genetical model for QTL studies as a homogeneous conditional Gaussian regression (HCGR) model that jointly models phenotypes and QTL genotypes. Conditional on the QTL genotypes and covariates, the phenotypes are distributed according

to a multivariate normal distribution. The QTLs and covariates enter the HCGR model through the mean in a similar fashion to the seemingly unrelated regression (SUR) model (Banerjee et al. 2008). However, the correlation structure among phenotypes is explicitly modeled according to the directed graph representation of the phenotype network. We derive the genetic model from a system of linear regression equations and show that it corresponds to a homogenous conditional Gaussian regression model.

In QTL studies, the observed data consist of phenotypic trait values, $\mathbf{y}$, and marker genotypes, $\mathbf{m}$, on n individuals derived from an inbred line cross. Following Sen and Churchill (2001) we condition on unobserved QTL genotypes, $\mathbf{q}$, to partition our model into genetic and recombination components, respectively relating phenotypes to QTLs and QTLs to observed markers across the genome,

$$p(\mathbf{y}\,,\,\mathbf{q}\mid\mathbf{m}) \;=\; p(\mathbf{y}\mid\mathbf{q}\,,\,\mathbf{m})\,p(\mathbf{q}\mid\mathbf{m}) \;=\; p(\mathbf{y}\mid\mathbf{q})\,p(\mathbf{q}\mid\mathbf{m})\,,$$

where the second equality follows from conditional independence, $\mathbf{y} \perp\!\!\!\perp \mathbf{m} \mid \mathbf{q}$. That is, given that QTL genotypes, the marker genotypes provide no additional information about the phenotypes. The recombination model, $p(\mathbf{q}\mid\mathbf{m})$, has been extensively studied in the past and is a well solved problem, with many efficient algorithms now available (Broman et al. 2003).

We assume that T phenotype traits are correlated in such a way that

$$y_{ti} \;=\; \mu^{\star}_{ti} \;+\; \sum_{v\in pa(y_t)} \beta_{tv}\, y_{vi} \;+\; \epsilon_{ti}\,, \tag{1}$$

with means $\mu^{\star}_{ti} = \mu_t + \mathbf{X}_{ti}\,\boldsymbol{\theta}_t$, $i = 1,\ldots,n$, $t = 1,\ldots,T$, and independent error terms ϵ_{ti}. μ_t is the overall mean for trait t. $\boldsymbol{\theta}_t$ is a column vector of all genetic effects constituting the genetic architecture of trait t plus any additional additive or interactive covariates. $pa(y_t)$ represent the set parent nodes of y_t, that is, the set of nodes that directly affect y_t. $\mathbf{X}_{ti}$ is the row vector of genetic effects predictors derived from the QTL genotypes along with any covariates. Genetic effects may follow Cockerham's genetic model, but need not be restricted to this form (Zeng et al. 2005).

Combine together the phenotypes and independent error terms for the ith individual as $\boldsymbol{\epsilon}_i = (\epsilon_{1i},\ldots,\epsilon_{Ti})'$ and $\mathbf{y}_i = (y_{1i},\ldots,y_{Ti})'$, respectively. The Jacobian transformation from $\boldsymbol{\epsilon}_i \to \mathbf{y}_i$ allows us to represent the joint density of the phenotype traits conditional on the respective genetic architectures as multivariate normal with the following mean vector and covariance matrix.

Result 1. *The set of structural equations*

$$y_{ti} \;=\; \mu^{\star}_{ti} \;+\; \sum_{v\in pa(y_t)} \beta_{tv}\, y_{vi} \;+\; \epsilon_{ti}, \qquad \epsilon_{ti} \;\sim\; N(0,\sigma^2_t)$$

with $i = 1,\ldots,n$, $t = 1,\ldots,T$ and ϵ_{ti} independent error terms has distribution

$$\mathbf{y}_i \mid \boldsymbol{\mu}^{\star}_i,\, \boldsymbol{\beta},\, \boldsymbol{\sigma}^2 \;\sim\; N_T\left(\boldsymbol{\Omega}^{-1}\,\boldsymbol{\gamma}_i\,,\,\boldsymbol{\Omega}^{-1}\right) \tag{2}$$

where $\boldsymbol{\mu}_i^\star = (\mu_{1i}^\star, \ldots, \mu_{Ti}^\star)'$, $\boldsymbol{\beta} = \{\beta_{tv} : v \in pa(y_t),\, t = 1, \ldots, T\}$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_T^2)'$, $\boldsymbol{\Omega}$ *is the concentration matrix with entries given by*

$$\omega_{tv} = \begin{cases} \dfrac{1}{\sigma_t^2} + \sum_s \dfrac{\beta_{st}^2}{\sigma_s^2} \mathbb{1}_{\{t \to s\}}, & \text{for} \quad t = v \\ \\ -\dfrac{\beta_{vt}}{\sigma_v^2} \mathbb{1}_{\{t \to v\}} - \dfrac{\beta_{tv}}{\sigma_t^2} \mathbb{1}_{\{v \to t\}} + \sum_s \dfrac{\beta_{sv}\, \beta_{st}}{\sigma_v^2} \mathbb{1}_{\{v \to s\,,\, t \to s\}}, & \text{for} \quad t \neq v \end{cases}$$

$\boldsymbol{\gamma}_i$ *is a vector with entries* $\frac{\mu_{ti}^\star}{\sigma_t^2} - \sum_{s \neq t} \frac{\beta_{st}\, \mu_{si}^\star}{\sigma_s^2} \mathbb{1}_{\{t \to s\}}$, *and* $\mathbb{1}_{\{t \to s\}}$ *is the indicator function that trait* t *affects trait* s.

Remarks: (1) The model allows different genetic architectures for each phenotype. (2) The covariance structure depends exclusively in the relationships among phenotypes since $\boldsymbol{\Omega}$ depends only on the partial regression coefficients relating phenotypes (βs) and variances of error terms (σ^2s), and not on the genetic architectures defined by the $\boldsymbol{\theta}$s. (3) When the correlation between two phenotypes arises exclusively because of a pleoitropic QTL, conditioning on the QTL genotypes makes the phenotypes independent; thus the concentration matrix of the conditional model does not depend on the genetic architecture. (4) This model can represent acyclic and cyclic networks. However, we focus on acyclic networks in this paper.

We now show that our model corresponds to a homogeneous conditional Gaussian regression model. The conditional Gaussian (CG) parametric family models the covariation of discrete and continuous random variables. Continuous random variables conditional on discrete variables are multivariate normal (Lauritzen 1996). The joint distribution of the vectors of discrete ($\mathbf{q}_i$) and continuous ($\mathbf{y}_i$) variables have a density f such that

$$\log f(\mathbf{q}_i\,,\, \mathbf{y}_i) \;=\; g(\mathbf{q}_i) \;+\; \mathbf{h}'(\mathbf{q}_i)\, \mathbf{y}_i \;-\; \mathbf{y}_i'\, \mathbf{K}(\mathbf{q}_i)\, \mathbf{y}_i/2 \tag{3}$$

where $g(\mathbf{q}_i)$ is a scalar, $\mathbf{h}(\mathbf{q}_i)$ is a vector and $\mathbf{K}(\mathbf{q}_i)$ is a positive definite matrix. The density f depends on observed markers $\mathbf{m}_i$ as

$$\log f(\mathbf{q}_i\,,\, \mathbf{y}_i) \;=\; \log p(\mathbf{y}_i\,,\, \mathbf{q}_i \mid \mathbf{m}_i) \tag{4}$$

where $g(\mathbf{q}_i) = \log p(\mathbf{q}_i \mid \mathbf{m}_i) - \frac{1}{2}\,(T \log 2\pi \;-\; \log det(\boldsymbol{\Omega}) \;+\; \sum_{t=1}^T \mu_{ti}^{\star\,2}/\sigma_t^2)$. Observe that the linear coefficients $\mathbf{h}(\mathbf{q}_i) = \boldsymbol{\gamma}_i$ depends on $\mathbf{q}_i$, through $\mathbf{X}_{ti}$, while the concentration matrix $\mathbf{K}(\mathbf{q}_i) = \boldsymbol{\Omega}$ does not. Thus, our model is a homogeneous CG model (Lauritzen 1996, page 160). Furthermore, since our genetic model was derived from a set of regression equations with normal errors, our model is in the homogeneous conditional Gaussian regression (HCGR) parametric family.

3 A causal framework for systems genetics

This section formalizes our causal inference framework for systems genetics that combines ideas from randomized experiments with a purely probabilistic approach to causal inference.

We argue that while causal claims about the relationship between QTLs and phenotypes are justified by randomization of alleles during meiosis and the unidirectional influence of genotype on phenotype, causal claims about the relationships between phenotypes follow from conditional probability. In a nutshell: by adding QTL nodes to phenotype networks we can distinguish between phenotype networks that would, otherwise, be distribution equivalent.

In order to formalize our approach we first show that adding causal QTL nodes can break Markov-equivalence among phenotype networks by creating new conditional independence relationships among nodes. Secondly, we note that two models in the HCGR parametric family are distribution equivalent if and only if they are Markov equivalent. The last result together with Theorem 1 (see below) provide a simple graphical criterion to determine whether two DAGs belonging to the HCGR parametric family are distribution equivalent.

Current literature in systems genetics (Li et al 2006, Kulp and Jagalur 2006, Aten et al 2008) claim that causality in systems genetics follows solely from the fact that recombination during meiosis mimics the process of randomization of treatments in a designed experiment. More correctly, randomization of alleles allows us to detect causal QTLs, but does not justify causal claims among phenotypes. Causality can be unambiguously inferred from a randomized experiment for two reasons (Dawid 2007) : (1) because application of the treatment to the experimental unit precedes the measurement of the outcomes, the direction of the causality must go from the treatment to measured outcome; (2) by randomly allocating treatments to experimental units, we guarantee that common causes to the outcome (other than the experimental treatment) get averaged out, that is, we eliminate the effects of confounding.

Pursuing the analogy of QTL mapping and a randomized experiment, let the different QTL genotypes represent treatment levels and the phenotypes measurements represent the measured outcomes of the experiment, then: (1) the central dogma of molecular biology (Crick 1958) guarantees that the application of the treatment (QTL genotypes) precedes the measurement of the outcomes (phenotype measurements). Therefore the direction of the causality must go from the QTL to the phenotype; (2) by randomly allocating treatments (QTL genotypes), we guarantee that other common causes to the phenotype get averaged out, that is, we eliminate the effects of confounding in the form of other genetic and environmental effects.

However, this analogy does not carry over to the relationships between phenotypes. Suppose we have a QTL, Q, and two phenotypes Y_1 and Y_2 mapping to Q, and their true causal relationship is $Q \rightarrow Y_1 \rightarrow Y_2$. One could argue that randomization of genotypes in Q leads to a randomization of Y_1 itself, so that we could think of Y_1 as a new randomized treatment affecting a second phenotype Y_2 (Chen et al. 2007). Because the randomization would average out confounding affects, we would be one step closer to claiming that $Y_1 \rightarrow Y_2$. However, contrary to the QTL/phenotype case, we have no *a priori* reason to believe that the "randomized treatment" Y_1 precedes the measurement of the "outcome" Y_2; it is possible that instead $Y_1 \leftarrow Y_2$. That is, causal claims need both randomization and precedence.

We next formalize a simple probabilistic causal framework that allows us to make causal claims about relationships among phenotypes. But first we need to present some graphical

model definitions and results. A path is any unbroken, nonintersecting sequence of edges in a graph, which may go along or against the direction of the arrows.

Definition 1. (d-separation) *A path p is said to be d-separated (or blocked) by a set of nodes Z if and only if*

1. *p contains a chain* $i \to m \to j$ *or a fork* $i \leftarrow m \to j$ *such that the middle node m is in Z, or*

2. *p contains an inverted fork (or collider)* $i \to m \leftarrow j$ *such that the middle node m is not in Z and such that no descendant of m is in Z.*

A set Z is said to d-separate X from Y if and only if Z blocks every path from a node in X to a node in Y. X and Y are d-connected if they are not d-separated (Pearl, 1988, 2000).

Two graphs are Markov equivalent (or faithful indistinguishable) if they have the same set of d-separation relations (Spirtes et. al. 2000). The skeleton of a causal graph is the undirected graph obtained by replacing its arrows by undirected edges. A v-structure is composed by two converging arrows whose tails are not connected by an arrow.

Theorem 1. (Detecting Markov equivalence) *Two directed acyclic graphs (DAGs) are Markov equivalent if and only if they have the same skeletons and the same set of v-structures. (Verma and Pearl 1990).*

Two models are likelihood equivalent if $f(y \mid M_1) = f(y \mid M_2)$ for any data set y, where $f(y \mid M)$ represent the prior predictive density of the data, y, conditional on model M (Heckerman et. al. 1995). In this paper we extend the definition of likelihood equivalence to predictive densities obtained by plugging-in maximum likelihood estimates in the respective sampling models. A closely related concept, distribution equivalence, states that two models are distribution equivalent if one is a re-parametrization of the other. While likelihood equivalence is defined in terms of predictive densities (prior predictive density or sampling model evaluated on the maximum likelihood estimates), distribution equivalence is defined in terms of the sampling model directly. Because of the invariance property of maximum likelihood estimates, distribution and likelihood equivalence are equivalent concepts in the frequentist setting. This is also true in the Bayesian setting with proper priors invariant to model re-parameterizations.

Suppose that for each pair of connected phenotypes in a graph there exists at least one QTL affecting one but not the other phenotype. Denote this new graph with QTLs included by the "extended graph". The next result shows that, in this particular situation, we can distinguish between causal models belonging to a Markov equivalent class of phenotype networks.

Result 2. *Consider a class of Markov equivalent DAGs* $\mathcal{G}$*. Let* Y_1 *and* Y_2 *be any two adjacent nodes in the graphs in* $\mathcal{G}$*. Assume that for each such pair there exists at least one variable, Q, directly affecting* Y_1 *but not* Y_2*. Let* $\mathcal{G}_E$ *represent the class of extended graphs. Then the graphs in* $\mathcal{G}_E$ *are not Markov equivalent.*

As an illustrative example consider the following class of Markov equivalent models: $\mathcal{G} = \{Y_1 \rightarrow Y_2 \rightarrow Y_3\,, Y_1 \leftarrow Y_2 \leftarrow Y_3\,, Y_1 \leftarrow Y_2 \rightarrow Y_3\}$. These causal models are Markov equivalent because they have the same set of conditional independence relations, namely $Y_1 \perp\!\!\!\perp Y_3 \mid Y_2$. In accordance with Theorem 1 the three models have the same skeleton, $Y_1 - Y_2 - Y_3$, and the same set of v-structures (no v-structures). Now consider one QTL, Q, affecting Y_2 but not Y_1 and Y_3. Then $\mathcal{G}_E$ is composed by

$$\begin{array}{ccc} Q & Q & Q \\ \downarrow & \downarrow & \downarrow \\ Y_1 \rightarrow Y_2 \rightarrow Y_3 & Y_1 \leftarrow Y_2 \leftarrow Y_3 & Y_1 \leftarrow Y_2 \rightarrow Y_3 \end{array}$$

Observe that these models still have the same skeleton but different sets of v-structures: $Y_1 \rightarrow Y_2 \leftarrow Q$, $Q \rightarrow Y_2 \leftarrow Y_3$ and $\emptyset$, respectively.

The next result guarantees that for the HCGR parametric family Markov equivalence implies distribution equivalence and vice-versa.

Result 3. *For the HCGR parametric family, two DAGs are distribution equivalent if and only if they are Markov equivalent.*

It follows from Results 2 and 3 that by extending the phenotype network to include QTLs we are able to infer a single network, instead of an equivalence class of graphs. Furthermore, if we consider Theorem 1 and Result 3 together, we have

Result 4. *For the HCGR parametric family, two DAGs are distribution equivalent if and only if they have the same skeletons and same sets of v-structures.*

Result 4 provides a simple graphical criterion to determine whether two HCGR models are distribution equivalent. This allows us to determine distribution equivalence by inspection of graph structures without the need to go through algebraic manipulations of joint probability distributions as in Chaibub Neto et al. (2008).

4 QTL mapping and phenotype network structure

In this paper, we rely heavily on the d-separation criterion to interpret the results of unconditional and conditional mapping analysis. d-separation (Pearl 1988, 2000) is a graphical criterion that allows us to predict conditional independence relationships in the data by simple inspection of the graph structure underlying the data generation process. In this section, we show that the conditional LOD score can be used as a formal measure of conditional independence relationships between phenotypes and QTLs. Contrary to partial correlations, the conditional LOD score does not require the assumption of multi-normality of the data in order to formally test for independence (recall that only in the Gaussian case, a zero (partial) correlation implies statistical (conditional) independence), and it can handle interactive covariates.

The conditional LOD score is defined as

$$\begin{aligned} LOD(y\,,\,q \mid x) &= LOD(y\,,\,q\,,\,x) - LOD(y\,,\,x) \\ &= log_{10}\left\{\frac{f(y \mid q\,,\,x)}{f(y)}\right\} - log_{10}\left\{\frac{f(y \mid x)}{f(y)}\right\}\,, \end{aligned} \tag{5}$$

where $f()$ represent a predicting density (a maximized likelihood or the prior predictive density in a Bayesian setting). It follows directly from this definition that

$$LOD(y\,,\,q \mid x) = 0 \quad \Leftrightarrow \quad f(y \mid q\,,\,x) = f(y \mid x) \quad \Leftrightarrow \quad Y \perp\!\!\!\perp Q \mid X. \tag{6}$$

Therefore we can use conditional LOD scores as a formal measure of independence between continuous (Y) and discrete (Q) random variables, conditional on any set of variables X, that could be either continuous, discrete or both. Observe that the unconditional LOD score, defined as $log_{10}\{f(y \mid q)/f(y)\}$, will be zero if and only if $Y \perp\!\!\!\perp Q$.

Furthermore, the conditional LOD score can be used to formally test for conditional independence in the presence of interacting covariates (denoted by $X \cdot Q$) since

$$LOD(y\,,\,q \mid x\,,\,x\cdot q) = log_{10}\left\{\frac{f(y \mid q\,,\,x\,,\,x\cdot q)}{f(y)}\right\} - log_{10}\left\{\frac{f(y \mid x)}{f(y)}\right\} = 0 \tag{7}$$

if and only if $Y \perp\!\!\!\perp \{Q\,,\,X\cdot Q\} \mid X$. This is a very desirable property since, in general, testing for conditional independence in the presence of interactions is not strait forward. For example, Andrei and Kendziorski (2008) point that in the presence of interactions, there is no one-to-one correspondence between zero partial correlations and conditional independencies, even when we assume normality of the full conditional distributions.

Traditional QTL mapping focuses on single trait analysis, where the network structure among the phenotypes is not taken into consideration in the analysis. Thus, single-trait analysis may detect QTLs that directly affect the phenotype under investigation, as well as QTLs with indirect effects, affecting phenotypes upstream to the phenotype under study. Consider, for example, the causal graph in Figure 1. The outputs of single trait analysis when Figure 1 represents the true network are given in Figure 2.

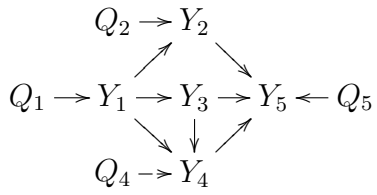


Figure 1: Example network with five phenotypes and four QTLs.

Now let's consider QTL mapping according to the phenotype network structure. When the phenotype structure corresponds to the true causal model, we avoid detecting indirect

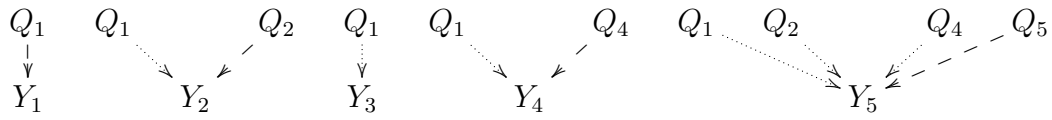


Figure 2: Output of a single trait QTL mapping analysis for the phenotypes in Figure 1. Dashed and pointed arrows represent direct and indirect QTL/phenotype causal relationships, respectively.

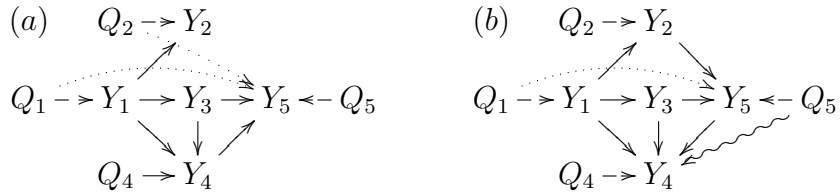


Figure 3: QTL mapping tailored to the network structure. (a) and (b) display the results of QTL mapping according to slightly altered network structures from Figure 1. Dashed, pointed and wiggled arrows represent, respectively, direct, indirect and incorrect QTL/phenotype causal relationships.

QTLs by simply performing mapping analysis of the phenotypes conditional on their parents. For example, in Figure 1, if we perform a mapping analysis of Y_5 conditional on Y_2, Y_3 and Y_4 we do not detect Q_1, Q_2 and Q_4 because $Y_5 \perp\!\!\!\perp Q_1 \mid Y_2, Y_3, Y_4$, $Y_5 \perp\!\!\!\perp Q_2 \mid Y_2, Y_3, Y_4$ and $Y_5 \perp\!\!\!\perp Q_4 \mid Y_2, Y_3, Y_4$. We only detect Q_5 since $Y_5 \not\perp\!\!\!\perp Q_5 \mid Y_2, Y_3, Y_4$.

Now consider Figure 3 (a). If we perform a mapping analysis of Y_5 conditional on Y_3 and Y_4 we still detect Q_1 and Q_2 as QTLs for Y_5, since failing to condition on Y_2 leaves the paths $Q_1 \rightarrow Y_1 \rightarrow Y_2 \rightarrow Y_5$ and $Q_2 \rightarrow Y_2 \rightarrow Y_5$ in Figure 1 open. In other words, Q_1 and Q_2 are d-connected to Y_5 conditional on (Y_3, Y_4) in the true causal graph.

Furthermore, if we perform mapping analysis of a phenotype conditional on phenotypes located downstream in the true network, we induce dependencies between the phenotype under study and QTLs affecting downstream phenotypes, and end up incorrectly detecting these QTLs. Consider, for example, Figure 3 (b). If we perform a mapping analysis of Y_4 conditional on Y_1, Y_3 and Y_5 we incorrectly detect Q_5 as a QTL for Y_4 because in the true network the paths $Y_4 \rightarrow Y_5 \leftarrow Q_5$ and $Y_4 \leftarrow Y_3 \rightarrow Y_5 \leftarrow Q_5$ in Figure 1 are open when we condition on Y_5.

In the next section we present a Metropolis-Hastings algorithm that at each step proposes a slightly modified phenotype network and fits the genetic architecture conditional on this proposed network. The proposal is accepted or rejected relative to the current network of phenotypes and QTLs. Models with better inferred genetic architectures should generally lead to higher marginal likelihood scores. Therefore, accounting for network structure in the mapping analysis should allow more efficient determination of the phenotype network structure.

5 QTLnet algorithm

In this section we propose a statistical framework (QTLnet) for the joint inference of phenotype network structure and genetic architecture in systems genetics studies. Work to date in genetical network reconstruction has treated the problems of QTL inference and phenotype network reconstruction separately, generally performing genetic architecture inference first, and then using QTLs to help in the determination of the phenotype network structure (Chaibub et al 2008, Zhu et al 2008). As indicated in the previous section, such strategy can incorporate QTLs with indirect effects into the genetic architecture of phenotypes.

The great challenge in the reconstruction of networks is that the graph space grows super-exponentially with the number of nodes, so that exhaustive searches are impractical even for small networks, and heuristic approaches are needed to efficiently traverse the graph space. The Metropolis-Hastings algorithm below integrates the sampling of network structures (Madigan and York 1995, Husmeier 2003) and QTL mapping.

Let $\mathcal{M}$ represent the structure of a phenotype network composed of T nodes. The posterior probability of a specified structure is given by

$$p(\mathcal{M} \mid \mathbf{y}\,,\, \mathbf{q}) = \frac{p(\mathbf{y} \mid \mathbf{q}\,,\, \mathcal{M})\, p(\mathcal{M})}{\sum_{\mathcal{M}} p(\mathbf{y} \mid \mathbf{q}\,,\, \mathcal{M})\, p(\mathcal{M})} \tag{8}$$

where the marginal likelihood

$$p(\mathbf{y} \mid \mathbf{q}\,,\, \mathcal{M}) = \int_{\mathbf{\Gamma}} p(\mathbf{y} \mid \mathbf{q}\,,\, \mathbf{\Gamma}\,,\, \mathcal{M})\, p(\mathbf{\Gamma} \mid \mathcal{M})\, d\mathbf{\Gamma} \tag{9}$$

is obtained by integrating the product of the prior and likelihood of the HCGR model with respect to all parameters $\mathbf{\Gamma}$ in the model. Assuming that the phenotype network is a DAG, the likelihood function factors according to $\mathcal{M}$ as

$$p(\mathbf{y}_i \mid \mathbf{q}_i\,,\, \mathbf{\Gamma}\,,\, \mathcal{M}) = \prod_t p\big(y_{ti} \mid \mathbf{q}_{ti}\,,\, pa(y_t)\big) \tag{10}$$

where

$$p\big(y_{ti} \mid \mathbf{q}_{ti}\,,\, pa(y_t)\big) = N\left(\mu^{\star}_{ti} + \sum_{y_k \in pa(y_t)} \beta_{tk}(y_{ki} - \mu^{\star}_{ki})\,,\, \sigma^2_t\right) \tag{11}$$

and the problem factors out as a series of linear regression models. (Note that QTL genotypes $\mathbf{q}_{ti}$ enters the model through $\mu^{\star}_{ti}$).

A Metropolis-Hastings step involves proposing a new network with a single modification from the last accepted network. Allowed modifications include single arrow deletions, additions or reversions that do not result in a cyclic network. (See Supplementary Figure S1 for details of the mechanics for the proposal scheme and for the computation of the proposal probabilities.) The Metropolis-Hastings proposal ratio $q(\mathcal{M}_{old} \mid \mathcal{M}_{new})/q(\mathcal{M}_{new} \mid \mathcal{M}_{old})$ is the ratio of the neighborhood sizes of the two networks involved in the proposal modification.

The neighborhood of a DAG, $ne(\mathcal{M})$ is defined as the number of DAGs that can be obtained from the present DAG in a single proposal modification. Specifically, the QTLnet algorithm proceeds as follows:

1. Given a phenotype network structure $\mathcal{M}_{old}$, propose a new network, $\mathcal{M}_{new}$, with a proposal probability $q(\mathcal{M}_{new} \mid \mathcal{M}_{old}) = 1/|ne(\mathcal{M}_{old})|$, $\mathcal{M}_{new} \in ne(\mathcal{M}_{old})$.

2. Factor the likelihood according to $\mathcal{M}_{new}$ and perform mapping analysis of y_t conditional on its parents, $pa(y_t)$, for each phenotype y_t separately.

3. Estimate each piece of the marginal likelihood, $p\big(y_t \mid \mathbf{q}_t\,,\, pa(y_t)\big)$, and combine then to obtain $\hat{p}(\mathbf{y} \mid \mathbf{q}\,,\, \mathcal{M}_{new})$.

4. Accept $\mathcal{M}_{new}$ with probability

$$\begin{aligned}\alpha &= min\left\{1\,,\, \frac{p(\mathcal{M}_{new} \mid \mathbf{y}\,,\, \mathbf{q})}{p(\mathcal{M}_{old} \mid \mathbf{y}\,,\, \mathbf{q})}\,\frac{q(\mathcal{M}_{old} \mid \mathcal{M}_{new})}{q(\mathcal{M}_{new} \mid \mathcal{M}_{old})}\right\} \\ &= min\left\{1\,,\, \frac{\hat{p}(\mathbf{y} \mid \mathbf{q}\,,\, \mathcal{M}_{new})\,p(\mathcal{M}_{new})}{\hat{p}(\mathbf{y} \mid \mathbf{q}\,,\, \mathcal{M}_{old})\,p(\mathcal{M}_{old})}\,\frac{q(\mathcal{M}_{old} \mid \mathcal{M}_{new})}{q(\mathcal{M}_{new} \mid \mathcal{M}_{old})}\right\}\,.\end{aligned}$$

Mapping analysis and the estimation of each component, $p(y_t \mid \mathbf{q}_t, pa(y_t))$, of the marginal likelihood can be done in two different ways: (1) perform Bayesian interval mapping (Yi et al. 2005) and use the posterior samples generated in the mapping analysis to estimate the marginal likelihood pieces using the stabilized harmonic mean estimator (Raftery et al 2008); (2) perform classical interval mapping, sum the BIC score for each piece of the sampling model to obtain the BIC score of the whole network, and then use the asymptotic approximation of the Bayes factors by BIC scores. It is a well known (*cf.* Kass and Raftery 1995) that when sample size is big enough, we can approximate the Bayes factor comparing old and new models by, in this case,

$$\frac{p(\mathbf{y} \mid \mathbf{q}\,,\, \mathcal{M}_{new})}{p(\mathbf{y} \mid \mathbf{q}\,,\, \mathcal{M}_{old})} \approx \exp\left\{-\frac{1}{2}(BIC_{\mathcal{M}_{new}} - BIC_{\mathcal{M}_{old}})\right\}\,. \tag{12}$$

Step 2 in the QTLnet algorithm is the most computationally demanding. Given the big size of the graph space even for small networks, efficient computation is key. The first obvious gain in efficiency is to store the estimates of the $p(y_t \mid \mathbf{q}_t\,,\, pa(y_t))$ as we proceed. Because $\mathcal{M}_{new}$ will only slightly differ from $\mathcal{M}_{old}$, most of the factor pieces of the network will not need to be recomputed. Actually, we will only need to perform new mapping analysis to the pair of nodes involved in an arrow reversal or to the child node involved in an arrow addition or deletion. In the first case, we need to redo the mapping analysis for both nodes when we reverse the arrow $y_1 \rightarrow y_2$ to $y_1 \leftarrow y_2$, since we need to include y_2 as a covariate in the mapping of y_1 and we need to exclude y_1 as a covariate in the mapping of y_2. For a child node, when we add (or delete) an arrow from y_1 to y_2 we need to include (or exclude) y_1 as a covariate in the mapping of y_2.

Another way to improve computational efficiency is to first perform a multiple-trait multiple-QTL analysis to all phenotypes with the SUR model (Banerjee et al 2008) and then only consider the QTLs detected by this analysis in the rest of the analysis. In this way, instead of considering all locations in the genome as potential QTLs every time we need to redo a mapping analysis, we simply consider the locations that were detected by the SUR analysis, greatly decreasing the total amount of necessary computations. Still another way to decrease computational time, in Step 2, is to perform classical interval mapping instead of Bayesian interval mapping, and then use the asymptotic approximation of the Bayes factors by BIC scores. We should point out, however, that since classical interval mapping is less powered to detect multiple QTL and misses epistasis altogether, the gain in computational efficiency provided by this approach comes at the expense of performing an inferior job in terms of genetic architecture inference.

Because the graph space grows rapidly with the number of phenotype nodes, the network structure with the highest posterior probability may still have a very low probability. Therefore, instead of selecting the network structure with the highest posterior probability, we perform Bayesian model averaging (Hoeting et al 1999) for the causal relationships between phenotypes and infer an averaged network. Explicitly, let Δ_{uv} represents a causal relationship between phenotypes u and v, that is, $\Delta_{uv} = \{Y_u \rightarrow Y_v\,,\, Y_u \leftarrow Y_v\,,\, Y_u \not\rightarrow Y_v$ and $Y_u \not\leftarrow Y_v\}$. Then

$$\begin{aligned} p(\Delta_{uv} \mid \mathbf{y}) &= \sum_k p(\Delta_{uv} \mid \mathcal{M}_k\,,\, \mathbf{y}\,,\, \mathbf{q})\, p(\mathcal{M}_k \mid \mathbf{y}\,,\, \mathbf{q}) \\ &= \sum_k \mathbb{1}\{\Delta_{uv} \in \mathcal{M}_k\}\, p(\mathcal{M}_k \mid \mathbf{y}\,,\, \mathbf{q})\ . \end{aligned} \tag{13}$$

Note that $p(Y_u \rightarrow Y_v \mid \mathcal{M}_k\,,\, \mathbf{y}) = \mathbb{1}\{Y_u \rightarrow Y_v \in \mathcal{M}_k\}$ since the probability that Y_u directly affects Y_v conditional on a graph structure $\mathcal{M}_k$ is either one or zero depending on whether Y_u does, or does not, directly affects Y_v in $\mathcal{M}_k$. The averaged network is constructed by putting together all causal relationships with maximum posterior probability or with posterior probability above a predetermined threshold.

6 Simulations

In this section we evaluate the performance of the QTLnet approach in simulation studies of a causal network with five phenotypes and four causal QTLs. We consider situations with strong or weak causal signals, leading respectively to high or low phenotype correlations. We show that important features of the causal network can be recovered. Further, this simulation illustrates how an alleged hotspot could be explained by sorting out direct and indirect effects of the QTLs.

We generated 1,000 data sets according to Figure 1. Each simulated cross object (Broman et al. 2003) had 5 phenotypes simulated for an F2 population with 500 individuals. The genome had 5 chromosomes of length 100cM with 10 equally spaced markers per chromosome.

We simulated one QTL per phenotype, except for phenotype Y_3 with no QTLs. The QTLs Q_t, $t = 1, 2, 4, 5$ were unlinked and placed at a the middle marker on chromosomes t.

Each simulated cross object had different sampled parameter value combinations for each realization. In the strong signal simulation, we sampled the additive and dominance effects according to $U[0.5, 1]$ and $U[0, 0.5]$, respectively. The partial regression coefficients for the phenotypes were sampled according to $\beta_{uv} \sim 0.5\,U[-1.5, -0.5] + 0.5\,U[0.5, 1.5]$. In the weak signal simulation, we generated data sets with additive and dominance effects from $U[0, 0.5]$ and $U[0, 0.25]$, respectively, and partial regression coefficients sampled with $\beta_{uv} \sim U[-0.5, 0.5]$. The residual phenotypic variance was fixed at 1 in both settings.

We first show the accuracy of the mapping analysis in our simulated data sets. We used interval mapping with a LOD score threshold of 5 to detect significant QTLs. Table 1 shows the results of both unconditional and QTL mapping according to the phenotype network in Figure 1. In the strong signal setting, the unconditional mapping often detected indirect QTLs, but the mapping of phenotypes conditional on their parent nodes increased detection of the true genetic architectures. In the weak signal simulation, the unconditional mapping did not detect indirect QTLs in most cases, but we still observe improvement in detection of the correct genetic architecture when we condition on the parents.

The expected architecture contains the d-connected QTLs when conditioning (or not) on other phenotypes as indicated in the first column of Table 1. For instance, Q_1 and Q_2 are d-connected to Y_2, but only Q_2 is d-connected to Y_2 when properly conditioning on Y_1. Supplementary Tables S1, S2, S3, S4 and S5 show the simulation results for all possible conditional mapping combinations.

	strong signal				weak signal				
Phenotypes	Q_1	Q_2	Q_4	Q_5	Q_1	Q_2	Q_4	Q_5	Expected architecture
Y_1	0.997	0.000	0.000	0.000	0.431	0.000	0.000	0.000	$\{Q_1\}$
Y_2	0.884	0.930	0.000	0.000	0.001	0.384	0.000	0.000	$\{Q_1, Q_2\}$
Y_3	0.941	0.000	0.000	0.000	0.003	0.000	0.000	0.000	$\{Q_1\}$
Y_4	0.603	0.000	0.690	0.000	0.003	0.000	0.370	0.000	$\{Q_1, Q_4\}$
Y_5	0.637	0.321	0.321	0.340	0.000	0.000	0.001	0.336	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_2 \mid Y_1$	0.001	0.999	0.000	0.000	0.000	0.424	0.000	0.000	$\{Q_2\}$
$Y_3 \mid Y_1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	$\emptyset$
$Y_4 \mid Y_1, Y_3$	0.000	0.000	0.999	0.000	0.000	0.000	0.422	0.000	$\{Q_4\}$
$Y_5 \mid Y_2, Y_3, Y_4$	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.415	$\{Q_5\}$

Table 1: Frequencies of QTL detection for both unconditional (top half) and conditional (bottom half) QTL mapping according with the true phenotype network structure in Figure 1. Results, for each simulation, based on 1,000 simulated data sets described in the text. The expected architecture is the set of d-connected QTLs for the phenotype conditioning with respect to the network in Figure 1.

.

For each simulated data set we applied the QTLnet algorithm using simple interval mapping for QTL detection. The ratio of marginal likelihoods in the Metropolis-Hastings algorithm was computed using the BIC asymptotic approximation to the Bayes factor. We adopted uniform priors over network structures. We ran each Markov chain for 30,000 iterations, sampled a structure every 10 iterations, and discarded the first 300 (burnin)

network structures producing posterior samples of size 2,700. Posterior probabilities for each causal relationship were computed via Bayesian model averaging.

Table 2 shows the frequency, out of the 1,000 simulations, the true model was the most probable, second most probable, etc. The results show that in the strong signal setting the true model got the highest posterior probability in most of the simulations. In the weak signal setting the range of rankings was very widespread.

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	≥15th
Strong	842	100	21	11	3	4	3	2	1	1	3	1	1	1	6
Weak	21	33	18	19	19	16	17	15	8	12	13	5	9	13	782

Table 2: Frequencies that the posterior probability of the true model was the highest, second highest, etc. Results based on 1,000 simulated data sets described in the text.

Table 3 shows the proportion of times that each possible causal direction ($Y_u \rightarrow Y_v$, $Y_u \leftarrow Y_v$) or no causal relation ($\{Y_u \not\rightarrow Y_v \ , \ Y_u \not\leftarrow Y_v\}$) had the highest posterior probability for all pairs of phenotypes. The results show that in the strong signal simulations, the correct causal relationships were recovered with high probability. The results are weaker but in the correct direction in the weak signal setting.

	strong signal			weak signal		
Phenotypes	$\rightarrow$	$\leftarrow$	$\not\rightarrow, \not\leftarrow$	$\rightarrow$	$\leftarrow$	$\not\rightarrow, \not\leftarrow$
(1,2)	0.996	0.002	0.002	0.594	0.177	0.229
(1,3)	0.990	0.002	0.008	0.471	0.263	0.266
(1,4)	0.990	0.001	0.009	0.541	0.196	0.263
(1,5)	0.054	0.002	0.944	0.028	0.005	0.967
(2,3)	0.016	0.022	0.962	0.018	0.017	0.965
(2,4)	0.037	0.012	0.951	0.018	0.015	0.967
(2,5)	0.997	0.003	0.000	0.712	0.075	0.213
(3,4)	0.967	0.031	0.002	0.482	0.253	0.265
(3,5)	0.997	0.003	0.000	0.653	0.116	0.231
(4,5)	0.996	0.004	0.000	0.670	0.115	0.215

Table 3: Frequencies that each possible causal relationships had the highest posterior probability (computed via Bayesian model averaging). Results based on 1,000 simulated data sets described in the text.

Interestingly, single trait analysis with strong signal showed that Y_5 mapped to Q_1 more frequently than to Q_5 (Table 1). This result can be understood using a path analysis (Wright 1934) argument. In path analysis, we decompose the correlation between two variables among all paths connecting the two variables in a graph. Let $\mathcal{D}_{uv}$ represent the set of all direct and indirect directed paths connecting u and v (a directed path is a path with all arrows pointing in the same direction). Then the correlation between these nodes can be decomposed as

$$cor(y_u, y_v) = \sum_{P \in \mathcal{D}_{uv}} \phi_{p_2 u} \, \phi_{p_3 p_2} \ldots \phi_{v p_{m-1}} = \left\{ \frac{var(y_u)}{var(y_v)} \right\}^{1/2} \sum_{P \in \mathcal{D}_{uv}} \beta_{p_2 u} \, \beta_{p_3 p_2} \ldots \beta_{v p_{m-1}} \quad (14)$$

where $\phi_{ij} = \beta_{ij} \left\{ \frac{var(y_j)}{var(y_i)} \right\}^{1/2}$ is a standardized path coefficient. Assuming intra-locus additivity and encoding the genotypes as 0, 1 and 2 (for the sake of easy computation), we have

from equation (14) that

$$cor(y_5, q_1) = \beta_{1,q_1}(\beta_{52}\,\beta_{21} + \beta_{53}\,\beta_{31} + \beta_{54}\,\beta_{41} + \beta_{54}\,\beta_{43}\,\beta_{31}) \left\{ \frac{var(q_1)}{var(y_5)} \right\}^{1/2},$$
$$cor(y_5, q_5) = \beta_{1,q_5} \left\{ \frac{var(q_5)}{var(y_5)} \right\}^{1/2}.$$

We therefore see that if the partial regression coefficients between phenotypes are high, and the QTL effects β_{1,q_1} and β_{5,q_5} and QTL variances are close (as in the strong signal simulation), then $cor(y_5, q_1)$ will be higher than $cor(y_5, q_5)$ and Y_5 will map to Q_1 with stronger signal than to Q_5.

This result suggests a possible scenario for the appearance of eQTL hotspots when phenotypes are highly correlated. A set of correlated phenotypes may be better modeled in a causal network with one upstream phenotype that in turn has a causal QTL. Ignoring the phenotype network can result in an apparent hotspot for the correlated phenotypes. Here, all phenotypes detect QTL Q_1 with high probability when mapped unconditionally (Table 1). In the weak signal setting, the phenotypes map mostly to their respective QTLs and do not show evidence for a hotspot. No hotspot was found in additional simulations having strong QTL/phenotype relationships and weak phenotype/phenotype relations (results not shown). Thus, our QTLnet approach can effectively explain a hotspot found with unconditional mapping when phenotypes show strong causal structure.

7 Real data example

In this section we illustrate the application of QTLnet to a subset of gene expression data derived from a F2 intercross between inbred lines $C3H/HeJ$ and $C57BL/6J$ (Ghazalpour et al. 2006, Wang et al. 2006). The data set is composed of genotype data on 1,065 markers and liver expression data on the 3421 available transcripts from 135 female mice. Interval mapping indicates that 14 transcripts map to the same region on chromosome 2 with a LOD score higher than 5.3 (permutation p-value <0.001). Only one transcript, *Pscdbp*, is located on chromosome 2 near the hotspot locus. The 14 transcripts show a strong correlation structure, and the correlation structure adjusting for the peak marker on chromosome 2, $rs3707138$, is still strong (see Supplementary Table S6). This co-mapping suggest all transcripts are under the regulation of a common factor. Causal relationships among phenotypes could explain the strong correlation structure that we observe, although other possibilities are environmental factors or a latent factor that is not included.

We applied the QTLnet algorithm on the 129 mice that had no missing transcript data using Haley-Knott (1992) regression (and assuming genotyping error rate of 0.01) for the detection of QTLs conditional on the network structure, and approximated the marginal likelihood ratio in the Metropolis-Hastings algorithm according to equation (12). We adopted uniform priors over network structures. We ran a Markov chain for 1,000,000 iterations and sampled a network structure every 100 iterations, discarding the first 1,000 and basing

inference on a posterior sample with 9,000 network structures. Inspection of the trace plots suggest good mixing of the Markov chain (Supplementary Figure S2).

We performed Bayesian model averaging and, for each of 91 possible pairs (Y_u, Y_v), we obtained the posterior probabilities of $Y_u \to Y_v$, $Y_u \leftarrow Y_v$ and of no direct causal connection. The results are shown in Supplementary Table S7. Figure 4 show a model-averaged network.

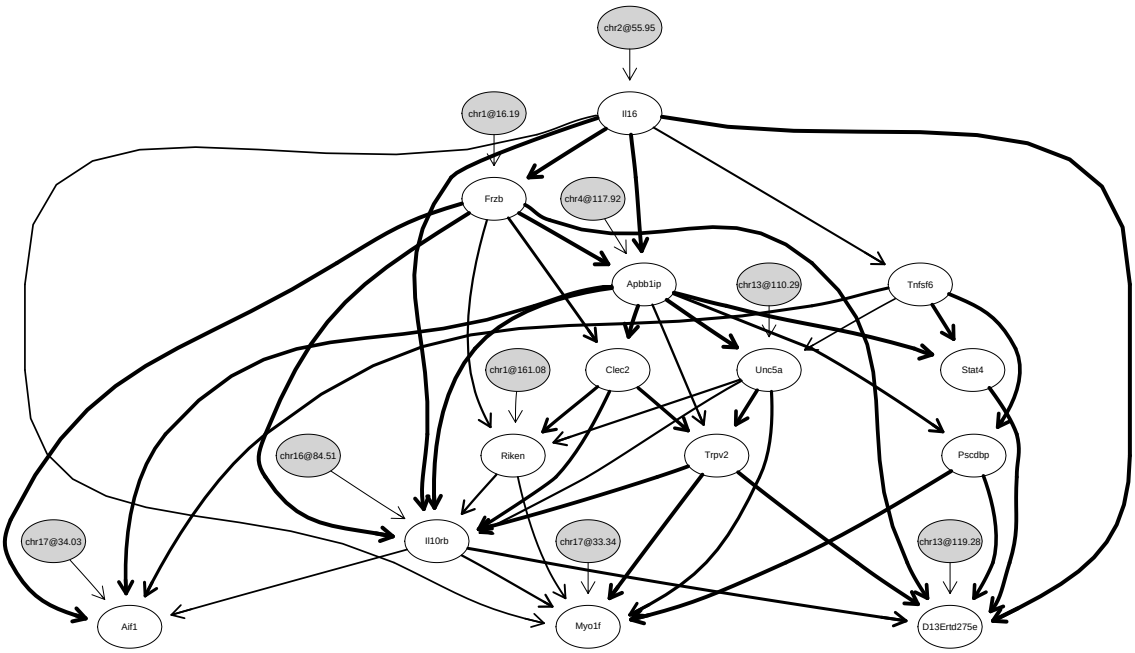


Figure 4: Model-averaged posterior network. Arrow thickness is proportional to the posterior probability of the causal relation computed via Bayesian model averaging. For each pair of phenotypes, the figure displays the causal relationship (presence or absence of an arrow) with highest posterior probability. Light grey nodes represent QTLs and show their chromosome number and position in centimorgans. Riken represents the riken gene $6530401C20Rik$.

This network suggests a key role of $Il16$ in the regulation of the other transcripts in the network. $Il16$ is upstream to all other transcripts, and is the only one directly mapping to the locus of chromosome 2. We would have expected the *cis* transcript, $Pscdbp$, to be the upstream phenotype in this network. However, the data suggests $Pscdbp$ is causal to only two other transcripts and that some other genetic factor on chromosome 2 may be driving this pathway. This estimated QTLnet causal network provides new hypotheses that could be tested in future mouse experiments.

8 Discussion

We have developed a statistical framework for causal inference in systems genetics. Causal relationships between QTLs and phenotypes are justified by the randomization of alleles during meiosis together with the unidirectional influence of genotypes on phenotypes. Causal relationships between phenotypes follows from breakage of distribution equivalence due to QTL nodes augmenting the phenotype network. We have proposed a novel approach to jointly infer genetic architecture and causal phenotype network structure using HCGR models. We argue in this paper that failing to properly account for phenotype network structure for mapping analysis can yield QTLs with indirect effects in the genetic architecture, which can decrease the power to detect the correct causal relationships between phenotypes.

Current literature in systems genetics (Chaibub Neto et al. 2008, Zhu et al 2008) has considered the problems of genetic architecture and phenotype network structure inference separately. Chaibub Neto et al. (2008) used the PC-algorithm (Spirtes et al 2000) to first infer the skeleton of the phenotype network and then use QTLs to determine the directions of the edges in the phenotype network. Zhu et al. (2008) reconstructed networks from a consensus of Bayesian phenotype networks with a prior distribution based on causal tests of Schadt et al. (2005). Their prior was computed with QTLs determined by single trait analysis.

Liu et al. (2008) presented an approach based in structural equation models (and applicable to species where sequence information is available) that partially accounts for the phenotype network structure when selecting the QTLs to be incorporated in the network. They perform eQTL mapping using *cis*, *cis-trans* and *trans*-regulation (Doss et al. 2005, Kulp and Jagalur 2006) and then use local structural models to identify regulator-target pairs for each eQTL. The identified relationships are then used to construct an encompassing directed network (EDN) with nodes composed by transcripts and eQTLs and arrows from (1) eQTls to *cis*-regulated target transcripts; (2) *cis*-regulated transcripts to *cis-trans*-regulated target transcripts; and (3) *trans*-regulator transcripts to target transcripts, and from *trans*-eQTL to target transcripts. The EDN defines a network search space for network inference with model selection based on penalized likelihood scores and an adaptation of Occam's window (Madigan and Raftery 1994). Their local structural models, which fit at most two candidate regulators per target transcript, can include indirect eQTLs in the genetic architecture of target trancripts when there are multiple paths connecting a *cis*-regulator to a *cis-trans*-target transcript. In other words, some transcripts identified as *cis*-regulated targets may actually be *cis-trans*.

Current interest in the eQTL literature centers on understanding the relationships among expression traits that co-map to a genomic region. It is often suggested that these eQTL "hotspots" result from a master regulator affecting the expression levels of many other genes (see Breitling et al. 2008). A path analysis argument suggest that if the correlation structure between the phenotypes is strong because of a strong causal signal, a well defined hotspot pattern will likely appear when we perform single trait analysis. Our simulations and real data example suggest that this is the situation where the QTLnet algorithm is expected to

be most fruitful.

The QTLnet approach is based on a Metropolis-Hastings algorithm that at each step proposes a slightly modified phenotype network and fits the genetic architecture conditional on this proposed network. Conditioning on the phenotype network structure should generally lead to a better inferred genetic architecture. Likewise, a better inferred genetic architecture should lead to a better inferred phenotype structure (models with better inferred genetic architectures should have higher marginal likelihood scores. A poorly inferred genetic architecture may compromise the marginal likelihood of a network with phenotype structure close to the true network).

Because the proposal mechanism of the Metropolis-Hastings algorithm is based in small modification of the last accepted network (addition, deletion or reversion of a single edge), the mixing of the Markov chain is generally slow and it is necessary to run long chains and use big thinning windows in order to achieve good mixing. This is a bottle-neck to the scalability of this approach. We therefore plan to investigate more efficient versions of the Metropolis-Hastings algorithm for network structure inference. In particular a new and more extensive edge reversal move proposed by Grzegorczyk and Husmeier (2008) and an approach based in a Markov blanket decomposition of the network (Riggelsen 2005).

One of the most attractive features of a Bayesian framework is its ability to formally incorporate prior information in the analysis. Given the complexity of biological processes and the many limitations associated with the partial pictures provided by any of the "omic" data sets now available, incorporation of external information is highly desirable. We are currently working in the development of priors for network structures.

The QTLnet approach can be seen as a method to infer causal Bayesian networks composed of phenotype and QTL nodes. Standard Bayesian networks provide a compact representation of the conditional dependency and independencies associated with a joint probability distribution. The main criticism of a causal interpretation of such networks is that different structures may be likelihood equivalent while representing totally different causal process. In other words, we can only infer a class of likelihood equivalent networks. We have formally shown how to break likelihood equivalence by incorporating causal QTLs.

We have focussed on experimental crosses with inbred founders as the recombination model and genetic architecture are relatively straightforward. However, this approach might be extended to outbred populations with some additional work. The genotypic effects are random, and the problem needs to be recast in terms of variance components.

In this paper we only consider directed acyclic graphs. We point out, however, that the HCGR parametric family accommodates cyclic networks. In the QTLnet approach we work under the assumption of no latent variables and we do not explicitly model measurement error. These complications can impact network reconstruction. We are currently investigating extensions of the proposed framework along these lines.

9 References

1. Andrei A, Kendziorski, C (2008) An efficient method for identifying statistical interactors in graphical models. Technical Report 200, Department of Biostatistics and Medical Informatics, University of Wisconsin - Madison.
2. Banerjee S, Yandell BS, Yi N (2008) Bayesian QTL mapping for multiple traits. it Genetics **179**: 2275-2289.
3. Breitling R, Li Y, Tesson BM, Fu J, Wu C, Wiltshire T, Gerrits A, Bystrykh LV, de Haan G, Su AI, Jansen RC (2008) Genetical genomics: spotlight on QTL hotspots. PLoS Genet 4: e1000232.
4. Broman K, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889-890.
5. Chaibub Neto E, Ferrara C, Attie AD, Yandell BS (2008) Inferring causal phenotype networks from segregating populations. *Genetics* **179**: 1089-1100.
6. Chen LS, Emmert-Streib F, Storey JD (2007) Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biology* **8**: R219.
7. Chickering D (1995) Search operators for learning equivalent classes of Bayesian network structures. Techinical Report R231, Cognitive Systems Laboratory, University of California, Los Angeles.
8. Crick FHC (1958) On Protein Synthesis. *Symp. Soc. Exp. Biol.* **XII**: 139-163.
9. Dawid P (2007) Fundamentals of statistical causality. Research Report 279, Department of Statistical Science, University College London.
10. Doss S, Schadt EE, Drake TA, Lusis AJ (2005) Cis-acting expression quantitative trait loci in mice. *Genome Research* **15**: 681-691.
11. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S (2006) Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genetics* **2**: e130.
12. Grzegorczyk M, Husmeier D (2008) Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. *Machine Learning* **71**: 265-305.
13. Haley C, Knott S (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315-324.
14. Heckerman D and Geiger D (1996) Likelihoods and parameter priors for Bayesian networks. Technical Report MSR-TR-95-94, Microsoft Research.
15. Heckerman D, Geiger D, Chickering D (1995) Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning* **20**: 197-243.
16. Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. *Statistical Science* **14**: 382-417.
17. Husmeier D (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics* **19**: 2271-2282.
18. Jiang C, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**: 1111-1127.
19. Kass RE, Raftery A (1995) Bayes factors. Journal of the American Statistical Association **90**: 773-795.
20. Kulp DC, Jagalur M (2006) Causal inference of regulator-target pairs by gene mapping of expression phenotypes. *BMC Genomics* **7**: 125.
21. Lauritzen S (1996) Graphical models. Oxford University Press.
22. Li R, Tsaih SW, Shockley K, Stylianou IM, Wergedal J, Paigen B, Churchill GA (2006) Structural model analysis of multiple quantitative traits. *PLoS Genetics* **2**: e114.
23. Liu B, de la Fuente A, Hoeschele I (2008) Gene network inference via structural equation modeling in genetical genomics experiments. *Genetics* **178**: 1763-1776.
24. Madigan D, York J (1995) Bayesian graphical models for discrete data. *International Statistical Review* **63**: 215-232.
25. Madigan D, Raftery J (1994) Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association* **89**: 1535-1546.

26. Meek C (1995) Strong completeness and faithfulness in Bayesian networks. *Proceedings of the Eleventh Conference on Uncertainty in AI*. San Francisco: Morgan and Kaufmann, 411-418.
27. Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann.
28. Pearl J (2000) Causality: models, reasoning and inference. Cambridge University Press.
29. Raftery AE, Newton MA, Satagopan JM, Krivitsky PN (2007) Estimating the integrated likelihhod via posterior simulation using the harmonic mean identity. *Bayesian Statistics* **8**: 1-45.
30. Riggelsen C (2005) MCMC learning of Bayesian network models by Markov blanket decomposition. *Lecture Notes in Computer Science* 329-340.
31. Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. *Genetics* **159**: 371-387.
32. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, Lusis AJ (2005) An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics* **37**: 710-717.
33. Shachter RD and Kenley R (1989) Gaussian influence diagrams. *Management Science*, **35**: 527-550.
34. Spirtes P, Glymour C, Scheines R (2000) Causation, prediction and search. MIT Press.
35. Verma T, Pearl J (1990) Equivalence and synthesis of causal models. *Proceedings of the Sixth Conference on Uncertainty in AI*, Boston, Morgan Kaufmann, 220-227.
36. Yi N, Yandell BS, Churchill GA, Allison DB, Eisen EJ, Pomp D (2005) Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* **170**: 1333-1344.
37. Wang S, Yehya N, Schadt EE, Wang H, Drake TA, Lusis AJ (2006) Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genetics* **2**: e15.
38. Wright S (1934) The method of path coefficients. *Annals of Mathematical Statistics*, **5**, 161-215.
39. Zeng ZB, Wang T, Zou W (2005) Modeling quantitative trait loci and interpretation of models. *Genetics* **169**: 1711-1725.
40. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics* **40**: 854-861.

E. Chaibub Neto
Department of Statistics
University of Wisconsin - Madison
1300 University Avenue
Madison, Wisconsin 53706 USA
E-mail: chaibub@stat.wisc.edu

M. P. Keller
Department of Biochemistry
University of Wisconsin - Madison
433 Babcock Drive
Madison, Wisconsin 53706 USA
E-mail: mkeller3@wisc.edu

A. D. Attie
Department of Biochemistry
University of Wisconsin - Madison
433 Babcock Drive
Madison, Wisconsin 53706 USA
E-mail: adattie@wisc.edu

B. S. Yandell
Department of Statistics and Horticulture
University of Wisconsin - Madison
1300 University Avenue
Madison, Wisconsin 53706 USA
E-mail: byandell@wisc.edu

Supplementary materials for:

Causal Graphical Models in System Genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes

Elias Chaibub Neto

Mark P Keller

Alan D Attie

Brian S Yandell

Supplementary figures and tables

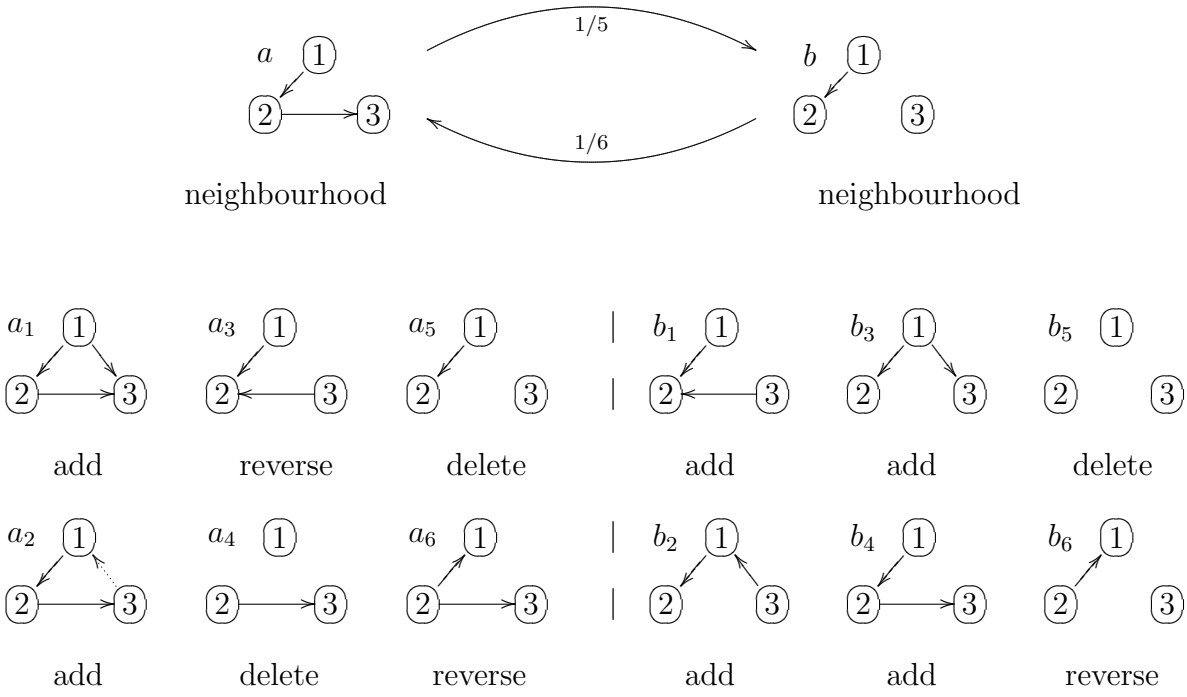


Figure S 1: Adapted from Husmeier 2003. Metropolis-Hastings proposal moves and ratio. This figure illustrates the three proposal modifications: adding, deleting or reversing one arrow in the network structure. When adding or reversing an arrow we need to check if the proposed modification leads to a cyclic network such (as in graph a_2). Cyclic structures are not valid and need to be discarded. The Metropolis-Hastings proposal ratio $q(\mathcal{M}_{old} \mid \mathcal{M}_{new})/q(\mathcal{M}_{new} \mid \mathcal{M}_{old})$ is given by the ratio of the neighborhood sizes of the two networks involved in the proposal modification. The neighborhood of a DAG, $ne(\mathcal{M})$ is defined as the number of DAGs that can be obtained from the present DAG in a single proposal modification. As an example, let structure a be $\mathcal{M}_{old}$. Since structure a_2 is cyclic, $ne(\mathcal{M}_{old}) = 5$ and the $q(\mathcal{M}_{new} \mid \mathcal{M}_{old}) = 1/5$. Suppose we sampled structure a_5 so that $a_5 = b = \mathcal{M}_{new}$. Then $ne(\mathcal{M}_{new}) = 6$ and the $q(\mathcal{M}_{old} \mid \mathcal{M}_{new}) = 1/6$. In this case the Metropolis-Hastings proposal ratio is 5/6.

Phenotypes	strong signal Q_1	Q_2	Q_4	Q_5	weak signal Q_1	Q_2	Q_4	Q_5	Expected architecture
Y_1	0.997	0.000	0.000	0.000	0.431	0.000	0.000	0.000	$\{Q_1\}$
$Y_1 \mid Y_2$	0.969	0.948	0.000	0.000	0.405	0.006	0.000	0.000	$\{Q_1, Q_2\}$
$Y_1 \mid Y_3$	0.956	0.000	0.000	0.000	0.406	0.000	0.000	0.000	$\{Q_1\}$
$Y_1 \mid Y_4$	0.945	0.000	0.494	0.000	0.408	0.000	0.003	0.000	$\{Q_1, Q_4\}$
$Y_1 \mid Y_5$	0.972	0.158	0.151	0.149	0.432	0.000	0.000	0.000	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_1 \mid Y_2, Y_3$	0.862	0.759	0.000	0.000	0.376	0.004	0.000	0.000	$\{Q_1, Q_2\}$
$Y_1 \mid Y_2, Y_4$	0.855	0.798	0.395	0.000	0.382	0.004	0.002	0.000	$\{Q_1, Q_2, Q_4\}$
$Y_1 \mid Y_2, Y_5$	0.911	0.857	0.119	0.117	0.400	0.003	0.000	0.000	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_1 \mid Y_3, Y_4$	0.867	0.000	0.766	0.000	0.382	0.000	0.003	0.000	$\{Q_1, Q_4\}$
$Y_1 \mid Y_3, Y_5$	0.903	0.116	0.143	0.106	0.406	0.000	0.000	0.000	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_1 \mid Y_4, Y_5$	0.908	0.191	0.452	0.178	0.412	0.000	0.003	0.000	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_1 \mid Y_2, Y_3, Y_4$	0.731	0.651	0.638	0.000	0.360	0.004	0.003	0.000	$\{Q_1, Q_2, Q_4\}$
$Y_1 \mid Y_2, Y_3, Y_5$	0.810	0.700	0.110	0.067	0.372	0.003	0.000	0.000	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_1 \mid Y_2, Y_4, Y_5$	0.831	0.752	0.407	0.175	0.378	0.004	0.002	0.000	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_1 \mid Y_3, Y_4, Y_5$	0.817	0.107	0.714	0.084	0.379	0.000	0.003	0.000	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_1 \mid Y_2, Y_3, Y_4, Y_5$	0.734	0.650	0.636	0.000	0.354	0.003	0.003	0.000	$\{Q_1, Q_2, Q_4\}$

Table S 1: Frequencies of QTL detection for all possible conditional mapping analysis of Y_1. Results based on 1,000 simulated data sets described in the text. The expected architecture is the set of d-connected QTLs for the phenotype conditioning with respect to the network in Figure 1.

Phenotypes	strong signal Q_1	Q_2	Q_4	Q_5	weak signal Q_1	Q_2	Q_4	Q_5	Expected architecture
Y_2	0.884	0.930	0.000	0.000	0.001	0.384	0.000	0.000	$\{Q_1, Q_2\}$
$Y_2 \mid Y_1$	0.000	0.999	0.000	0.000	0.000	0.424	0.000	0.000	$\{Q_2\}$
$Y_2 \mid Y_3$	0.375	0.986	0.000	0.000	0.001	0.390	0.000	0.000	$\{Q_1, Q_2\}$
$Y_2 \mid Y_4$	0.498	0.968	0.096	0.000	0.001	0.392	0.000	0.000	$\{Q_1, Q_2, Q_4\}$
$Y_2 \mid Y_5$	0.500	0.850	0.192	0.232	0.001	0.360	0.000	0.000	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_2 \mid Y_1, Y_3$	0.000	0.999	0.000	0.000	0.000	0.424	0.000	0.000	$\{Q_2\}$
$Y_2 \mid Y_1, Y_4$	0.000	0.999	0.000	0.000	0.000	0.426	0.000	0.000	$\{Q_2\}$
$Y_2 \mid Y_1, Y_5$	0.000	0.995	0.134	0.206	0.000	0.405	0.000	0.003	$\{Q_2, Q_4, Q_5\}$
$Y_2 \mid Y_3, Y_4$	0.144	0.994	0.115	0.000	0.001	0.397	0.000	0.000	$\{Q_1, Q_2, Q_4\}$
$Y_2 \mid Y_3, Y_5$	0.223	0.910	0.347	0.427	0.002	0.364	0.000	0.003	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_2 \mid Y_4, Y_5$	0.253	0.826	0.180	0.686	0.001	0.357	0.000	0.003	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_2 \mid Y_1, Y_3, Y_4$	0.000	0.999	0.000	0.000	0.000	0.426	0.000	0.000	$\{Q_2\}$
$Y_2 \mid Y_1, Y_3, Y_5$	0.000	0.991	0.336	0.421	0.000	0.404	0.000	0.004	$\{Q_2, Q_4, Q_5\}$
$Y_2 \mid Y_1, Y_4, Y_5$	0.000	0.988	0.035	0.698	0.000	0.405	0.000	0.004	$\{Q_2, Q_4, Q_5\}$
$Y_2 \mid Y_3, Y_4, Y_5$	0.020	0.827	0.027	0.963	0.002	0.361	0.000	0.006	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_2 \mid Y_1, Y_3, Y_4, Y_5$	0.000	0.969	0.000	0.937	0.000	0.402	0.000	0.007	$\{Q_2, Q_5\}$

Table S 2: Frequencies of QTL detection for all possible conditional mapping analysis of Y_2. Results based on 1,000 simulated data sets described in the text. The expected architecture is the set of d-connected QTLs for the phenotype conditioning with respect to the network in Figure 1.

Phenotypes	strong signal Q_1	Q_2	Q_4	Q_5	weak signal Q_1	Q_2	Q_4	Q_5	Expected architecture
Y_3	0.941	0.000	0.000	0.000	0.003	0.000	0.000	0.000	$\{Q_1\}$
$Y_3 \mid Y_1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	$\emptyset$
$Y_3 \mid Y_2$	0.525	0.503	0.000	0.000	0.003	0.000	0.000	0.000	$\{Q_1, Q_2\}$
$Y_3 \mid Y_4$	0.478	0.000	0.688	0.000	0.005	0.000	0.003	0.000	$\{Q_1, Q_4\}$
$Y_3 \mid Y_5$	0.546	0.212	0.228	0.258	0.003	0.000	0.000	0.001	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_3 \mid Y_1, Y_2$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	$\emptyset$
$Y_3 \mid Y_1, Y_4$	0.000	0.000	0.916	0.000	0.000	0.000	0.001	0.000	$\{Q_4\}$
$Y_3 \mid Y_1, Y_5$	0.000	0.199	0.214	0.218	0.000	0.000	0.000	0.001	$\{Q_2, Q_4, Q_5\}$
$Y_3 \mid Y_2, Y_4$	0.318	0.315	0.747	0.000	0.003	0.000	0.003	0.000	$\{Q_1, Q_2, Q_4\}$
$Y_3 \mid Y_2, Y_5$	0.290	0.275	0.379	0.421	0.003	0.000	0.000	0.001	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_3 \mid Y_4, Y_5$	0.309	0.228	0.631	0.283	0.002	0.000	0.002	0.001	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_3 \mid Y_1, Y_2, Y_4$	0.000	0.000	0.914	0.000	0.000	0.000	0.001	0.000	$\{Q_4\}$
$Y_3 \mid Y_1, Y_2, Y_5$	0.000	0.000	0.377	0.413	0.000	0.000	0.000	0.001	$\{Q_4, Q_5\}$
$Y_3 \mid Y_1, Y_4, Y_5$	0.000	0.135	0.840	0.217	0.000	0.000	0.001	0.002	$\{Q_2, Q_4, Q_5\}$
$Y_3 \mid Y_2, Y_4, Y_5$	0.156	0.157	0.568	0.857	0.001	0.000	0.002	0.002	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_3 \mid Y_1, Y_2, Y_4, Y_5$	0.000	0.000	0.764	0.762	0.000	0.000	0.001	0.002	$\{Q_4, Q_5\}$

Table S 3: Frequencies of QTL detection for all possible conditional mapping analysis of Y_3. Results based on 1,000 simulated data sets described in the text. The expected architecture is the set of d-connected QTLs for the phenotype conditioning with respect to the network in Figure 1.

Phenotypes	strong signal Q_1	Q_2	Q_4	Q_5	weak signal Q_1	Q_2	Q_4	Q_5	Expected architecture
Y_4	0.603	0.000	0.690	0.000	0.003	0.000	0.370	0.000	$\{Q_1, Q_4\}$
$Y_4 \mid Y_1$	0.000	0.000	0.960	0.000	0.000	0.000	0.392	0.000	$\{Q_4\}$
$Y_4 \mid Y_2$	0.373	0.383	0.808	0.000	0.000	0.000	0.373	0.000	$\{Q_1, Q_2, Q_4\}$
$Y_4 \mid Y_3$	0.372	0.000	0.979	0.000	0.002	0.000	0.401	0.000	$\{Q_1, Q_4\}$
$Y_4 \mid Y_5$	0.383	0.330	0.607	0.385	0.002	0.000	0.348	0.005	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_4 \mid Y_1, Y_2$	0.000	0.000	0.960	0.000	0.000	0.000	0.392	0.000	$\{Q_4\}$
$Y_4 \mid Y_1, Y_3$	0.000	0.000	0.999	0.000	0.000	0.000	0.422	0.000	$\{Q_4\}$
$Y_4 \mid Y_1, Y_5$	0.000	0.357	0.827	0.411	0.000	0.000	0.374	0.003	$\{Q_2, Q_4, Q_5\}$
$Y_4 \mid Y_2, Y_3$	0.125	0.114	0.988	0.000	0.001	0.000	0.402	0.000	$\{Q_1, Q_2, Q_4\}$
$Y_4 \mid Y_2, Y_5$	0.174	0.153	0.626	0.672	0.000	0.000	0.352	0.007	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_4 \mid Y_3, Y_5$	0.205	0.357	0.905	0.450	0.001	0.000	0.376	0.001	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_4 \mid Y_1, Y_2, Y_3$	0.000	0.000	0.999	0.000	0.000	0.000	0.420	0.000	$\{Q_4\}$
$Y_4 \mid Y_1, Y_2, Y_5$	0.000	0.000	0.759	0.728	0.000	0.000	0.373	0.004	$\{Q_4, Q_5\}$
$Y_4 \mid Y_1, Y_3, Y_5$	0.000	0.359	0.987	0.456	0.000	0.000	0.402	0.001	$\{Q_2, Q_4, Q_5\}$
$Y_4 \mid Y_2, Y_3, Y_5$	0.013	0.017	0.836	0.973	0.000	0.000	0.377	0.003	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_4 \mid Y_1, Y_2, Y_3, Y_5$	0.000	0.000	0.971	0.950	0.000	0.000	0.401	0.002	$\{Q_4, Q_5\}$

Table S 4: Frequencies of QTL detection for all possible conditional mapping analysis of Y_4. Results based on 1,000 simulated data sets described in the text. The expected architecture is the set of d-connected QTLs for the phenotype conditioning with respect to the network in Figure 1.

Phenotypes	strong signal				weak signal				Expected architecture
	Q_1	Q_2	Q_4	Q_5	Q_1	Q_2	Q_4	Q_5	
Y_5	0.637	0.321	0.320	0.340	0.000	0.000	0.001	0.336	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_1$	0.000	0.540	0.524	0.582	0.000	0.000	0.001	0.338	$\{Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_2$	0.296	0.272	0.546	0.566	0.000	0.000	0.001	0.350	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_3$	0.242	0.584	0.598	0.652	0.000	0.000	0.001	0.362	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_4$	0.280	0.669	0.221	0.744	0.000	0.001	0.000	0.377	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_1, Y_2$	0.000	0.000	0.657	0.708	0.000	0.000	0.001	0.357	$\{Q_4, Q_5\}$
$Y_5 \mid Y_1, Y_3$	0.000	0.701	0.708	0.809	0.000	0.000	0.001	0.367	$\{Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_1, Y_4$	0.000	0.793	0.130	0.893	0.000	0.001	0.000	0.369	$\{Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_2, Y_3$	0.045	0.033	0.853	0.874	0.000	0.000	0.001	0.389	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_2, Y_4$	0.118	0.093	0.283	0.968	0.000	0.000	0.000	0.395	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_3, Y_4$	0.051	0.846	0.026	0.892	0.000	0.001	0.000	0.392	$\{Q_1, Q_2, Q_4, Q_5\}$
$Y_5 \mid Y_1, Y_2, Y_3$	0.000	0.001	0.894	0.928	0.000	0.000	0.001	0.392	$\{Q_4, Q_5\}$
$Y_5 \mid Y_1, Y_2, Y_4$	0.000	0.000	0.271	0.988	0.000	0.000	0.000	0.394	$\{Q_4, Q_5\}$
$Y_5 \mid Y_1, Y_3, Y_4$	0.000	0.893	0.000	0.936	0.000	0.001	0.000	0.392	$\{Q_2, Q_5\}$
$Y_5 \mid Y_2, Y_3, Y_4$	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.415	$\{Q_5\}$
$Y_5 \mid Y_1, Y_2, Y_3, Y_4$	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.417	$\{Q_5\}$

Table S 5: Frequencies of QTL detection for all possible conditional mapping analysis of Y_5. Results based on 1,000 simulated data sets described in the text. The expected architecture is the set of d-connected QTLs for the phenotype conditioning with respect to the network in Figure 1.

	Apbb.	Clec2	Trpv2	D13E.	Pscd.	Frzb	Myo1f	Il10.	Unc5a	Aif1	Riken	Stat4	Il16	Tnfs.
Apbb.	1	0.881	0.854	0.776	0.815	0.604	0.754	0.718	0.789	0.750	0.698	0.657	0.649	0.486
Clec2	0.860	1	0.852	0.739	0.704	0.642	0.736	0.776	0.708	0.651	0.735	0.604	0.614	0.442
Trpv2	0.827	0.825	1	0.782	0.709	0.597	0.783	0.750	0.807	0.693	0.687	0.587	0.587	0.536
D13E.	0.732	0.691	0.739	1	0.764	0.751	0.734	0.663	0.741	0.619	0.678	0.650	0.728	0.571
Pscd.	0.777	0.644	0.645	0.708	1	0.594	0.781	0.595	0.664	0.699	0.591	0.619	0.648	0.554
Frzb	0.519	0.570	0.510	0.693	0.484	1	0.601	0.631	0.500	0.371	0.700	0.462	0.545	0.442
Myo1f	0.705	0.686	0.740	0.676	0.727	0.506	1	0.673	0.764	0.747	0.702	0.566	0.545	0.585
Il10.	0.666	0.737	0.704	0.597	0.506	0.552	0.608	1	0.689	0.587	0.694	0.481	0.670	0.402
Unc5a	0.749	0.656	0.770	0.688	0.588	0.389	0.715	0.630	1	0.732	0.681	0.580	0.567	0.567
Aif1	0.698	0.582	0.629	0.532	0.620	0.212	0.688	0.501	0.675	1	0.528	0.621	0.586	0.636
Riken	0.637	0.685	0.624	0.607	0.488	0.627	0.636	0.633	0.614	0.417	1	0.490	0.525	0.471
Stat4	0.594	0.534	0.512	0.581	0.535	0.346	0.479	0.386	0.501	0.542	0.387	1	0.559	0.624
Il16	0.581	0.543	0.507	0.672	0.566	0.442	0.449	0.607	0.481	0.494	0.423	0.474	1	0.509
Tnfs.	0.377	0.330	0.438	0.473	0.437	0.301	0.489	0.276	0.473	0.547	0.346	0.546	0.401	1

Table S 6: Correlations between phenotypes. The upper triangle gives the pairwise correlations between phenotypes. The lower triangle shows the pairwise correlations between the residuals of the phenotypes after adjustment for the common QTL to which all phenotypes map in the unconditional analysis.

pheno	pheno	$\rightarrow$	$\leftarrow$	$\not\rightarrow, \not\leftarrow$	pheno	pheno	$\rightarrow$	$\leftarrow$	$\not\rightarrow, \not\leftarrow$
Apbb.	Clec2	1.000	0.000	0.000	Pscd.	Frzb	0.000	0.302	0.698
Apbb.	Trpv2	0.581	0.000	0.419	Pscd.	Myo1f	0.969	0.031	0.000
Apbb.	D13E.	0.186	0.005	0.809	Pscd.	Il10.	0.186	0.083	0.731
Apbb.	Pscd.	1.000	0.000	0.000	Pscd.	Unc5a	0.058	0.071	0.871
Apbb.	Frzb	0.000	0.998	0.002	Pscd.	Aif1	0.17	0.162	0.668
Apbb.	Myo1f	0.259	0.000	0.741	Pscd.	Riken	0.075	0.048	0.877
Apbb.	Il10.	0.748	0.000	0.252	Pscd.	Stat4	0.067	0.057	0.876
Apbb.	Unc5a	0.998	0.000	0.002	Pscd.	Il16	0.005	0.488	0.507
Apbb.	Aif1	1.000	0.000	0.000	Pscd.	Tnfs.	0.324	0.409	0.267
Apbb.	Riken	0.250	0.000	0.750	Frzb	Myo1f	0.284	0.000	0.716
Apbb.	Stat4	0.908	0.077	0.015	Frzb	Il10.	0.548	0.000	0.452
Apbb.	Il16	0.006	0.983	0.011	Frzb	Unc5a	0.074	0.000	0.926
Apbb.	Tnfs.	0.144	0.038	0.818	Frzb	Aif1	0.932	0.000	0.068
Clec2	Trpv2	0.874	0.123	0.003	Frzb	Riken	1.000	0.000	0.000
Clec2	D13E.	0.201	0.013	0.786	Frzb	Stat4	0.066	0.000	0.934
Clec2	Pscd.	0.087	0.094	0.819	Frzb	Il16	0.005	0.984	0.011
Clec2	Frzb	0.000	0.718	0.282	Frzb	Tnfs.	0.199	0.066	0.735
Clec2	Myo1f	0.122	0.022	0.856	Myo1f	Il10.	0.089	0.570	0.341
Clec2	Il10.	0.838	0.093	0.069	Myo1f	Unc5a	0.086	0.540	0.374
Clec2	Unc5a	0.083	0.113	0.804	Myo1f	Aif1	0.006	0.152	0.842
Clec2	Aif1	0.089	0.035	0.876	Myo1f	Riken	0.117	0.568	0.315
Clec2	Riken	0.810	0.170	0.020	Myo1f	Stat4	0.028	0.166	0.806
Clec2	Stat4	0.065	0.044	0.891	Myo1f	Il16	0.001	0.511	0.488
Clec2	Il16	0.000	0.051	0.949	Myo1f	Tnfs.	0.054	0.442	0.504
Clec2	Tnfs.	0.048	0.054	0.898	Il10.	Unc5a	0.034	0.552	0.414
Trpv2	D13E.	0.954	0.033	0.013	Il10.	Aif1	0.442	0.127	0.431
Trpv2	Pscd.	0.076	0.114	0.810	Il10.	Riken	0.075	0.475	0.450
Trpv2	Frzb	0.000	0.117	0.883	Il10.	Stat4	0.048	0.132	0.820
Trpv2	Myo1f	0.968	0.015	0.017	Il10.	Il16	0.002	0.998	0.000
Trpv2	Il10.	0.966	0.011	0.023	Il10.	Tnfs.	0.011	0.294	0.695
Trpv2	Unc5a	0.059	0.916	0.025	Unc5a	Aif1	0.214	0.310	0.476
Trpv2	Aif1	0.116	0.054	0.830	Unc5a	Riken	0.789	0.173	0.038
Trpv2	Riken	0.109	0.08	0.811	Unc5a	Stat4	0.044	0.058	0.898
Trpv2	Stat4	0.065	0.089	0.846	Unc5a	Il16	0.001	0.059	0.94
Trpv2	Il16	0.002	0.059	0.939	Unc5a	Tnfs.	0.005	0.931	0.064
Trpv2	Tnfs.	0.026	0.222	0.752	Aif1	Riken	0.101	0.143	0.756
D13E.	Pscd.	0.081	0.893	0.026	Aif1	Stat4	0.042	0.074	0.884
D13E.	Frzb	0.000	1.000	0.000	Aif1	Il16	0.004	0.123	0.873
D13E.	Myo1f	0.275	0.212	0.513	Aif1	Tnfs.	0.074	0.926	0.000
D13E.	Il10.	0.132	0.750	0.118	Riken	Stat4	0.037	0.109	0.854
D13E.	Unc5a	0.005	0.367	0.628	Riken	Il16	0.001	0.097	0.902
D13E.	Aif1	0.096	0.211	0.693	Riken	Tnfs.	0.040	0.099	0.861
D13E.	Riken	0.077	0.182	0.741	Stat4	Il16	0.031	0.210	0.759
D13E.	Stat4	0.010	0.959	0.031	Stat4	Tnfs.	0.239	0.761	0.000
D13E.	Il16	0.007	0.993	0.000	Il16	Tnfs.	0.430	0.192	0.378
D13E.	Tnfs.	0.009	0.144	0.847					

Table S 7: Posterior probabilities for causal relationships computed via Bayesian model averaging.

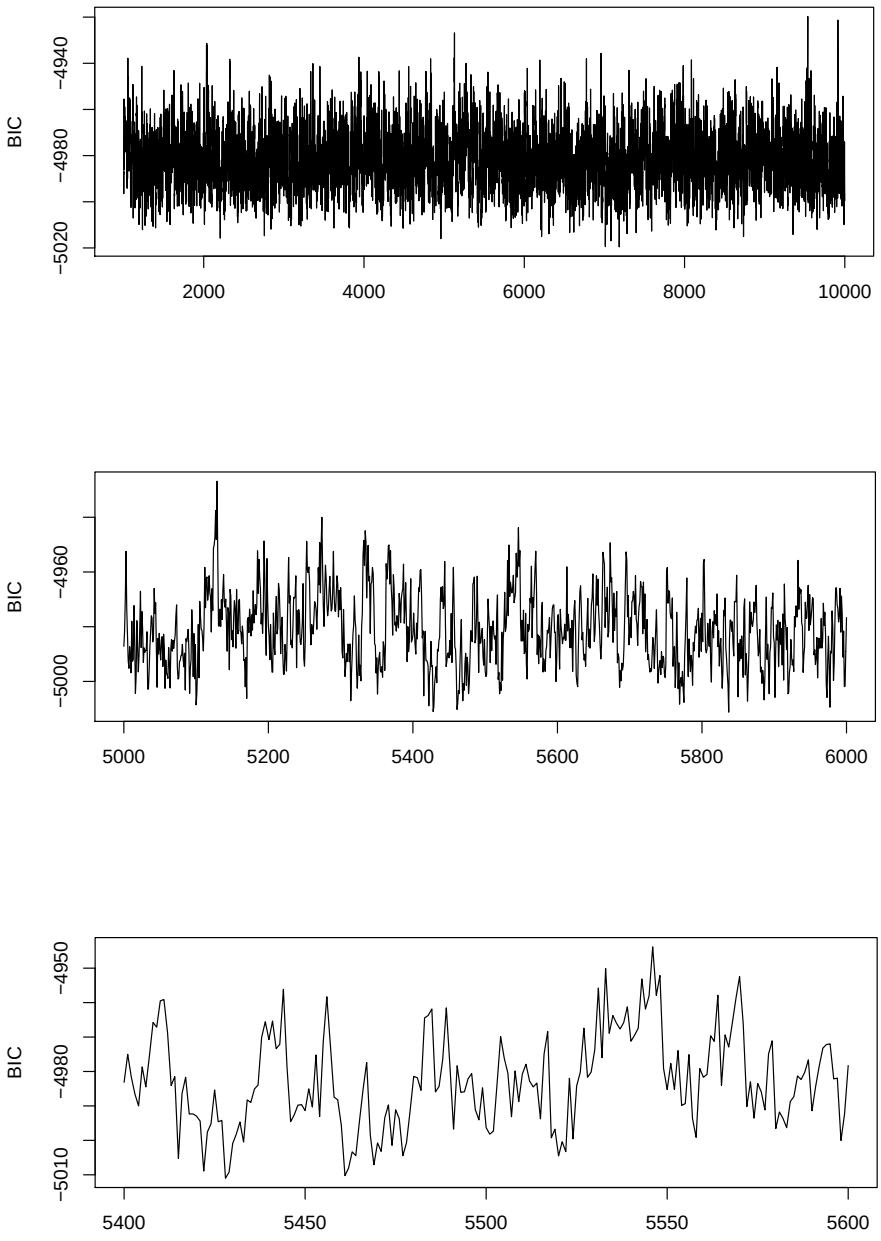


Figure S 2: Trace plots for the Markov chain. The y-axis shows the BIC for each sampled network. The upper plot displays the Markov chain for the 9,000 samples structures. The middle and lower plots show windows of size 1,000 and 200, respectively.

Proof of Result 1

Result 1. *The set of structural equations*

$$y_{ti} \;=\; \mu^{\star}_{ti} \;+\; \sum_{v \in pa(y_t)} \beta_{tv}\, y_{vi} \;+\; \epsilon_{ti}, \qquad \epsilon_{ti} \;\sim\; N(0, \sigma^2_t)$$

with $i = 1, \ldots, n$, $t = 1, \ldots, T$ *and* ϵ_{ti} *independent error terms has distribution*

$$\mathbf{y}_i \mid \boldsymbol{\mu}^{\star}_i,\, \boldsymbol{\beta},\, \boldsymbol{\sigma}^2 \;\sim\; N_T\left(\boldsymbol{\Omega}^{-1}\,\boldsymbol{\gamma}_i\,,\, \boldsymbol{\Omega}^{-1}\right)$$

where $\boldsymbol{\mu}^{\star}_i = (\mu^{\star}_{1i}, \ldots, \mu^{\star}_{Ti})'$, $\boldsymbol{\beta} = \{\beta_{tv} : v \in pa(y_t),\, t = 1, \ldots, T\}$, $\boldsymbol{\sigma}^2 = (\sigma^2_1, \ldots, \sigma^2_T)'$, $\boldsymbol{\Omega}$ *is the concentration matrix with entries given by*

$$\omega_{tv} = \begin{cases} \dfrac{1}{\sigma^2_t} + \sum_s \dfrac{\beta^2_{st}}{\sigma^2_s} \mathbb{1}_{\{t \to s\}}, & \text{for} \quad t = v \\[2ex] -\dfrac{\beta_{vt}}{\sigma^2_v} \mathbb{1}_{\{t \to v\}} - \dfrac{\beta_{tv}}{\sigma^2_t} \mathbb{1}_{\{v \to t\}} + \sum_s \dfrac{\beta_{sv}\,\beta_{st}}{\sigma^2_v} \mathbb{1}_{\{v \to s\,,\, t \to s\}}, & \text{for} \quad t \neq v \end{cases}$$

$\boldsymbol{\gamma}_i$ *is a vector with entries*

$$\frac{\mu^{\star}_{ti}}{\sigma^2_t} \;-\; \sum_{s \neq t} \frac{\beta_{st}\,\mu^{\star}_{si}}{\sigma^2_s} \mathbb{1}_{\{t \to s\}}$$

and $\mathbb{1}_{\{t \to s\}}$ *is the indicator function that trait* t *affects trait* s.

We start proving some algebraic equalities needed in the proof of Result 1.

Result 5. *Let* $\mathbb{1}\{j \to k\}$ *represent the indicator function that* $y_j \in pa(y_k)$ *or, conversely, that* $y_k \in ch(y_j)$. *Then*

(a)

$$\sum_{t=1}^{T} \lambda_t\, \mu^{\star}_t \left(y_t \;-\; \sum_{k\,:\, y_k \in pa(y_t)} \beta_{tk}\, y_k \right) \;=\; \sum_{t=1}^{T} \left(\lambda_t\, \mu^{\star}_t - \sum_{s \neq t} \lambda_s\, \beta_{st}\, \mu^{\star}_s\, \mathbb{1}_{\{t \to s\}} \right) y_t$$

(b)

$$\sum_{t=1}^{T} \lambda_t \left(y^2_t \;+\; \sum_{k\,:\, y_k \in pa(y_t)} \beta^2_{tk}\, y^2_k \right) \;=\; \sum_{t=1}^{T} \left(\lambda_t + \sum_{s \neq t} \lambda_s\, \beta^2_{st}\, \mathbb{1}_{\{t \to s\}} \right) y^2_t$$

(c)

$$\sum_{t=1}^{T} \lambda_t \left(- \sum_{k\,:\, y_k \in pa(y_t)} \beta_{tk}\, y_t\, y_k \;+\; \sum_{(k,j)\,:\, y_k\,,\, y_j \in pa(y_t)} \beta_{tk}\, \beta_{tj}\, y_k\, y_j \right) \;=$$

$$= \sum_{v < t} \left(-\lambda_v\, \beta_{vt}\, \mathbb{1}_{\{t \to v\}} - \lambda_t\, \beta_{tv}\, \mathbb{1}_{\{v \to t\}} + \sum_{s \neq v\,,\, s \neq t} \lambda_s\, \beta_{sv}\, \beta_{st}\, \mathbb{1}_{\{v \to s\,,\, t \to s\}} \right) y_v\, y_t$$

The proof is strait-forward, we only need to perform a rearrangement of the terms in the summations. Nonetheless, we present it here for the reader's sake.

Proof. (a) Assume without loss of generality that $T = 3$.

$$\sum_{t=1}^{3} \lambda_t \, \mu_t^\star \left(y_t \; - \sum_{k \, : \, y_k \, \in \, pa(y_t)} \beta_{tk} \, y_k \right) \; = \; \sum_{t=1}^{3} \lambda_t \, \mu_t^\star \left(y_t \; - \sum_{k \neq t} \beta_{tk} \, \mathbb{1}_{\{k \to t\}} \, y_k \right) \; =$$

$$= \; \lambda_1 \, \mu_1^\star \, \left(y_1 - \beta_{12} \, y_2 \, \mathbb{1}_{\{2 \to 1\}} - \beta_{13} \, y_3 \, \mathbb{1}_{\{3 \to 1\}} \right)$$

$$+ \lambda_2 \, \mu_2^\star \, \left(y_2 - \beta_{21} \, y_1 \, \mathbb{1}_{\{1 \to 2\}} - \beta_{23} \, y_3 \, \mathbb{1}_{\{3 \to 2\}} \right)$$

$$+ \lambda_3 \, \mu_3^\star \, \left(y_3 - \beta_{31} \, y_1 \, \mathbb{1}_{\{1 \to 3\}} - \beta_{32} \, y_2 \, \mathbb{1}_{\{2 \to 3\}} \right)$$

$$= \; \left(\lambda_1 \, \mu_1^\star \; - \lambda_2 \, \mu_2^\star \, \beta_{21} \, \mathbb{1}_{\{1 \to 2\}} - \lambda_3 \, \mu_3^\star \, \beta_{31} \, \mathbb{1}_{\{1 \to 3\}} \right) y_1$$

$$+ \; \left(\lambda_2 \, \mu_2^\star \; - \lambda_1 \, \mu_1^\star \, \beta_{12} \, \mathbb{1}_{\{2 \to 1\}} - \lambda_3 \, \mu_3^\star \, \beta_{32} \, \mathbb{1}_{\{2 \to 3\}} \right) y_2$$

$$+ \; \left(\lambda_3 \, \mu_3^\star \; - \lambda_1 \, \mu_1^\star \, \beta_{13} \, \mathbb{1}_{\{3 \to 1\}} - \lambda_2 \, \mu_2^\star \, \beta_{32} \, \mathbb{1}_{\{3 \to 2\}} \right) y_3$$

$$= \; \sum_{t=1}^{3} \left(\lambda_t \, \mu_t^\star - \sum_{s \neq t} \lambda_s \, \beta_{st} \, \mu_s^\star \, \mathbb{1}_{\{t \to s\}} \right) y_t$$

(b) Assume without loss of generality that $T = 3$.

$$\sum_{t=1}^{3} \lambda_t \left(y_t^2 \; + \sum_{k \, : \, y_k \, \in \, pa(y_t)} \beta_{tk}^2 \, y_k^2 \right) \; = \; \sum_{t=1}^{3} \lambda_t \left(y_t^2 \; + \sum_{k \neq t} \beta_{tk}^2 \, \mathbb{1}_{\{k \to t\}} \, y_k^2 \right) \; =$$

$$= \; \lambda_1 \left(y_1^2 + \beta_{12}^2 \, y_2^2 \, \mathbb{1}_{\{2 \to 1\}} + \beta_{13}^2 \, y_3^2 \, \mathbb{1}_{\{3 \to 1\}} \right) + \lambda_2 \left(y_2^2 + \beta_{21}^2 \, y_1^2 \, \mathbb{1}_{\{1 \to 2\}} + \beta_{23}^2 \, y_3^2 \, \mathbb{1}_{\{3 \to 2\}} \right)$$

$$+ \lambda_3 \left(y_3^2 + \beta_{31}^2 \, y_1^2 \, \mathbb{1}_{\{1 \to 3\}} + \beta_{32}^2 \, y_2^2 \, \mathbb{1}_{\{2 \to 3\}} \right) = \left(\lambda_1 + \lambda_2 \, \beta_{21}^2 \, \mathbb{1}_{\{1 \to 2\}} + \lambda_3 \, \beta_{31}^2 \, \mathbb{1}_{\{1 \to 3\}} \right) y_1^2$$

$$+ \left(\lambda_2 + \lambda_1 \, \beta_{12}^2 \, \mathbb{1}_{\{2 \to 1\}} + \lambda_3 \beta_{32}^2 \, \mathbb{1}_{\{2 \to 3\}} \right) y_2^2 + \left(\lambda_3 + \lambda_1 \, \beta_{13}^2 \, \mathbb{1}_{\{3 \to 1\}} + \lambda_2 \beta_{32}^2 \, \mathbb{1}_{\{3 \to 2\}} \right) y_3^2$$

$$= \; \sum_{t=1}^{3} \left(\lambda_t + \sum_{s \neq t} \lambda_s \, \beta_{st}^2 \, \mathbb{1}_{\{t \to s\}} \right) y_t^2$$

(c) Assume without loss of generality that $T = 4$.

$$\sum_{t=1}^{T} \lambda_t \left(- \sum_{k \, : \, y_k \, \in \, pa(y_t)} \beta_{tk} \, y_t \, y_k + \sum_{(k,j) \, : \, y_k \, , \, y_j \, \in \, pa(y_t)} \beta_{tk} \, \beta_{tj} \, y_k \, y_j \right) \; =$$

$$= \sum_{t=1}^{4} \lambda_t \left(- \sum_{k \neq t} \beta_{tk} \, \mathbb{1}_{\{k \to t\}} \, y_t \, y_k + \sum_{k < j \, , \, k \neq t \, , \, j \neq t} \beta_{tk} \, \beta_{tj} \, \mathbb{1}_{\{k \to t \, , \, j \to t\}} \, y_k \, y_j \right)$$

$$
\begin{aligned}
&= -\lambda_1\,\beta_{12}\,\mathbb{1}_{\{2\to1\}}\,y_1\,y_2 - \lambda_1\,\beta_{13}\,\mathbb{1}_{\{3\to1\}}\,y_1\,y_3 - \lambda_1\,\beta_{14}\,\mathbb{1}_{\{4\to1\}}\,y_1\,y_4 \\
&+\lambda_1\,\beta_{12}\,\beta_{13}\mathbb{1}_{\{2\to1\,,\,3\to1\}}\,y_2\,y_3 + \lambda_1\,\beta_{12}\,\beta_{14}\mathbb{1}_{\{2\to1\,,\,4\to1\}}\,y_2\,y_4 + \lambda_1\,\beta_{13}\,\beta_{14}\mathbb{1}_{\{3\to1\,,\,4\to1\}}\,y_3\,y_4 \\
&-\lambda_2\,\beta_{21}\,\mathbb{1}_{\{1\to2\}}\,y_2\,y_1 - \lambda_2\,\beta_{23}\,\mathbb{1}_{\{3\to2\}}\,y_2\,y_3 - \lambda_2\,\beta_{24}\,\mathbb{1}_{\{4\to2\}}\,y_2\,y_4 \\
&+\lambda_2\,\beta_{21}\,\beta_{23}\mathbb{1}_{\{1\to2\,,\,3\to2\}}\,y_1\,y_3 + \lambda_2\,\beta_{21}\,\beta_{24}\mathbb{1}_{\{1\to2\,,\,3\to2\}}\,y_1\,y_4 + \lambda_2\,\beta_{23}\,\beta_{24}\mathbb{1}_{\{3\to2\,,\,4\to2\}}\,y_3\,y_4 \\
&-\lambda_3\,\beta_{31}\,\mathbb{1}_{\{1\to3\}}\,y_3\,y_1 - \lambda_3\,\beta_{32}\,\mathbb{1}_{\{2\to3\}}\,y_3\,y_2 - \lambda_3\,\beta_{34}\,\mathbb{1}_{\{4\to3\}}\,y_3\,y_4 \\
&+\lambda_3\,\beta_{31}\,\beta_{32}\mathbb{1}_{\{1\to3\,,\,2\to3\}}\,y_1\,y_2 + \lambda_3\,\beta_{31}\,\beta_{34}\mathbb{1}_{\{1\to3\,,\,4\to3\}}\,y_1\,y_4 + \lambda_3\,\beta_{32}\,\beta_{34}\mathbb{1}_{\{2\to3\,,\,4\to3\}}\,y_2\,y_4 \\
&-\lambda_4\,\beta_{41}\,\mathbb{1}_{\{1\to4\}}\,y_4\,y_1 - \lambda_4\,\beta_{42}\,\mathbb{1}_{\{2\to4\}}\,y_4\,y_2 - \lambda_4\,\beta_{43}\,\mathbb{1}_{\{3\to4\}}\,y_4\,y_3 \\
&+\lambda_4\,\beta_{41}\,\beta_{42}\mathbb{1}_{\{1\to4\,,\,2\to4\}}\,y_1\,y_2 + \lambda_4\,\beta_{41}\,\beta_{43}\mathbb{1}_{\{1\to4\,,\,3\to4\}}\,y_1\,y_3 + \lambda_4\,\beta_{42}\,\beta_{43}\mathbb{1}_{\{2\to4\,,\,3\to4\}}\,y_2\,y_3 \\
&= \left(-\lambda_1\,\beta_{12}\,\mathbb{1}_{\{2\to1\}} - \lambda_2\,\beta_{21}\,\mathbb{1}_{\{1\to2\}} + \lambda_3\,\beta_{31}\,\beta_{32}\mathbb{1}_{\{1\to3\,,\,2\to3\}} + \lambda_4\,\beta_{41}\,\beta_{42}\mathbb{1}_{\{1\to4\,,\,2\to4\}}\right)\,y_1\,y_2 \\
&+ \left(-\lambda_1\,\beta_{13}\,\mathbb{1}_{\{3\to1\}} - \lambda_3\,\beta_{31}\,\mathbb{1}_{\{1\to3\}} + \lambda_2\,\beta_{21}\,\beta_{23}\mathbb{1}_{\{1\to2\,,\,3\to2\}} + \lambda_4\,\beta_{41}\,\beta_{43}\mathbb{1}_{\{1\to4\,,\,3\to4\}}\right)\,y_1\,y_3 \\
&+ \left(-\lambda_1\,\beta_{14}\,\mathbb{1}_{\{4\to1\}} - \lambda_4\,\beta_{41}\,\mathbb{1}_{\{1\to4\}} + \lambda_3\,\beta_{31}\,\beta_{34}\mathbb{1}_{\{1\to3\,,\,4\to3\}} + \lambda_3\,\beta_{31}\,\beta_{34}\mathbb{1}_{\{1\to3\,,\,4\to3\}}\right)\,y_1\,y_4 \\
&+ \left(-\lambda_2\,\beta_{23}\,\mathbb{1}_{\{3\to2\}} - \lambda_3\,\beta_{31}\,\mathbb{1}_{\{2\to3\}} + \lambda_1\,\beta_{12}\,\beta_{13}\mathbb{1}_{\{2\to1\,,\,3\to1\}} + \lambda_4\,\beta_{42}\,\beta_{43}\mathbb{1}_{\{2\to4\,,\,3\to4\}}\right)\,y_2\,y_3 \\
&+ \left(-\lambda_2\,\beta_{24}\,\mathbb{1}_{\{4\to2\}} - \lambda_4\,\beta_{42}\,\mathbb{1}_{\{2\to4\}} + \lambda_1\,\beta_{12}\,\beta_{14}\mathbb{1}_{\{2\to1\,,\,4\to1\}} + \lambda_3\,\beta_{32}\,\beta_{34}\mathbb{1}_{\{2\to3\,,\,4\to3\}}\right)\,y_2\,y_4 \\
&+ \left(-\lambda_3\,\beta_{34}\,\mathbb{1}_{\{4\to3\}} - \lambda_4\,\beta_{43}\,\mathbb{1}_{\{3\to4\}} + \lambda_1\,\beta_{13}\,\beta_{14}\mathbb{1}_{\{3\to1\,,\,4\to1\}} + \lambda_2\,\beta_{23}\,\beta_{24}\mathbb{1}_{\{3\to2\,,\,4\to2\}}\right)\,y_3\,y_4 \\
&= \sum_{v<t}\left(-\lambda_v\,\beta_{vt}\,\mathbb{1}_{\{t\to v\}} - \lambda_t\,\beta_{tv}\,\mathbb{1}_{\{v\to t\}} + \sum_{s\neq v\,,\,s\neq t}\lambda_s\,\beta_{sv}\,\beta_{st}\,\mathbb{1}_{\{v\to s\,,\,t\to s\}}\right)\,y_v\,y_t
\end{aligned}
$$

□

Proof. **(Result 1)**

A set of discrete ($\mathbf{q}$) and continuous ($\mathbf{y}$) variables follow a conditional Gaussian distribution (Lauritzen 1996) if the joint distribution of the vectors of discrete and continuous variables have a density f such that

$$\log f(\mathbf{q}\,,\,\mathbf{y}) \;=\; g(\mathbf{q}) \;+\; \mathbf{h}'(\mathbf{q})\,\mathbf{y} \;-\; \mathbf{y}'\,\mathbf{K}(\mathbf{q})\,\mathbf{y}/2 \tag{15}$$

where $g(\mathbf{q})$ is a scalar, $\mathbf{h}(\mathbf{q})$ is a vector and $\mathbf{K}(\mathbf{q})$ is a positive definite matrix. The above statement is equivalent to

$$P(\mathbf{q}) > 0 \quad \text{and} \quad (\mathbf{y}\mid\mathbf{q}) \;\sim\; N\left(\mathbf{K}(\mathbf{q})^{-1}\mathbf{h}(\mathbf{q})\,,\,\mathbf{K}(\mathbf{q})^{-1}\right). \tag{16}$$

In the following we will rewrite $\log p(\mathbf{y}\,,\,\mathbf{q}\mid\mathbf{m})$ (note we dropped the index i from the notation) in the formate of (15) and we will obtain the explicit forms of $g(\mathbf{q})$, $\mathbf{h}(\mathbf{q})$ and $\mathbf{K}(\mathbf{q})$. Let $\boldsymbol{\epsilon} = (\epsilon_1,\ldots,\epsilon_T)'$ be the vector of independent normal error terms and $\mathbf{y} = (y_1,\ldots,y_T)'$ the traits. Denote the Jacobian element for the transformation from $\boldsymbol{\epsilon}\to\mathbf{y}$ by $J_{\boldsymbol{\epsilon}\to\mathbf{y}}$. Then

the joint density of the phenotype traits conditional on the respective genetic architectures is given by

$$p(\mathbf{y} \mid \mathbf{q}) \;=\; \prod_{t=1}^{T} p\big(y_t \mid \mathbf{q}_t\,,\, pa(y_t)\big)$$

$$= \frac{|J_{\boldsymbol{\epsilon}\to\mathbf{y}}|}{(2\pi)^{T/2}\prod_{t=1}^{T}\sigma_t} \exp\left\{-\sum_{t=1}^{T}\frac{1}{2\sigma_t^2}\left(y_t - \mu_t^\star - \sum_{y_k\,\in\, pa(y_t)}\beta_{tk}\,y_k\right)^2\right\}$$

where

$$\left(y_t - \mu_t^\star - \sum_{y_k\,\in\, pa(y_t)}\beta_{tk}\,y_k\right)^2 \;=\; \left(\mu_t^{\star\,2} \;-\; 2\,\mu_t^\star\,y_t \;+\; 2\,\mu_t^\star \sum_{y_k\,\in\, pa(y_t)}\beta_{tk}\,y_k\right) \;+$$

$$+\left(y_t^2 \;+\; \sum_{y_k\,\in\, pa(y_t)}\beta_{tk}^2\,y_k^2 \;-\; 2\sum_{y_k\,\in\, pa(y_t)}\beta_{tk}\,y_t\,y_k \;+\; 2\sum_{k<j}\beta_{tk}\,\beta_{tj}\,y_k\,y_j\right)\;.$$

Let $\lambda_t = \prod_{j\neq t}\sigma_j^2$. Then

$$\log p(\mathbf{y}\,,\,\mathbf{q} \mid \mathbf{m}) \;=\; \log p(\mathbf{q} \mid \mathbf{m}) \;+\; \log\left\{\frac{|J_{\boldsymbol{\epsilon}\to\mathbf{y}}|}{(2\pi)^{T/2}\prod_{t=1}^{T}\sigma_t}\right\} \;+\; (I) \;+\; (II)$$

where

$$(I) \;=\; -\,\frac{1}{2\prod_{t=1}^{T}\sigma_t^2}\sum_{t=1}^{T}\lambda_t\left(\mu_t^{\star\,2} \;-\; 2\,\mu_t^\star\,y_t \;+\; 2\,\mu_t^\star \sum_{y_k\,\in\, pa(y_t)}\beta_{tk}\,y_k\right)$$

$$= -\,\frac{1}{2\prod_{t=1}^{T}\sigma_t^2}\left[\sum_{t=1}^{T}\lambda_t\,\mu_t^{\star\,2} \;-\; 2\sum_{t=1}^{T}\lambda_t\,\mu_t^\star\left(y_t \;-\; \sum_{y_k\,\in\, pa(y_t)}\beta_{tk}\,y_k\right)\right]$$

$$= -\,\frac{1}{2\prod_{t=1}^{T}\sigma_t^2}\left[\sum_{t=1}^{T}\lambda_t\,\mu_t^{\star\,2} \;-\; 2\sum_{t=1}^{T}y_t\left(\lambda_t\,\mu_t^\star \;-\; \sum_{s\neq t}\lambda_s\,\beta_{st}\,\mu_s^\star\,\mathbb{1}_{\{t\to s\}}\right)\right]$$

$$= -\,\frac{1}{2}\sum_{t=1}^{T}\frac{\mu_t^{\star\,2}}{\sigma_t^2} \;+\; \sum_{t=1}^{T}\left(\frac{\mu_t^\star}{\sigma_t^2} - \sum_{s\neq t}\frac{\beta_{st}\,\mu_s^\star}{\sigma_s^2}\,\mathbb{1}_{\{t\to s\}}\right)y_t$$

$$= -\,\frac{1}{2}\sum_{t=1}^{T}\frac{\mu_t^{\star\,2}}{\sigma_t^2} \;+\; \boldsymbol{\gamma}'\,\mathbf{y}$$

where the third equality follows from Result 5 (a) and $\boldsymbol{\gamma}' = (\frac{\mu_1^\star}{\sigma_1^2} - \sum_{s\neq 1}\frac{\beta_{s1}\,\mu_s^\star}{\sigma_s^2}\,\mathbb{1}_{\{1\to s\}}, \ldots, \frac{\mu_T^\star}{\sigma_T^2} -$

$\sum_{s\neq T} \frac{\beta_{sT}\,\mu_s^\star}{\sigma_s^2}\,\mathbb{1}_{\{T\to s\}}$).

$$
\begin{aligned}
(II) \;&=\; -\frac{1}{2\prod_{t=1}^T \sigma_t^2}\sum_{t=1}^T \lambda_t \left(y_t^2 \;+\; \sum_{y_k\,\in\, pa(y_t)} \beta_{tk}^2\, y_k^2 \;-\; 2 \sum_{y_k\,\in\, pa(y_t)} \beta_{tk}\, y_t\, y_k \;+\; 2\sum_{k<j} \beta_{tk}\,\beta_{tj}\, y_k\, y_j \right) \\
&=\; -\frac{1}{2\prod_{t=1}^T \sigma_t^2}\left[\sum_{t=1}^T \lambda_t \left(y_t^2 \;+\; \sum_{y_k\,\in\, pa(y_t)} \beta_{tk}^2\, y_k^2 \right) + \right. \\
&\qquad\qquad \left. +\, 2\sum_{t=1}^T \lambda_t \left(-\sum_{y_k\,\in\, pa(y_t)} \beta_{tk}\, y_t\, y_k \;+\; \sum_{k<j} \beta_{tk}\,\beta_{tj}\, y_k\, y_j \right) \right] \\
&=\; -\frac{1}{2\prod_{t=1}^T \sigma_t^2}\left[\sum_{t=1}^T \left(\lambda_t + \sum_{s\neq t} \lambda_s\,\beta_{st}^2\,\mathbb{1}_{\{t\to s\}} \right) y_t^2 + \right. \\
&\quad \left. +\, 2\sum_{v<t} \left(-\lambda_v\,\beta_{vt}\,\mathbb{1}_{\{t\to v\}} - \lambda_t\,\beta_{tv}\,\mathbb{1}_{\{v\to t\}} + \sum_{s\neq v\,,\, s\neq t} \lambda_s\,\beta_{sv}\,\beta_{st}\,\mathbb{1}_{\{v\to s\,,\, t\to s\}} \right) y_v\, y_t \right] \\
&= -\frac{1}{2}\left[\sum_{t=1}^T \left(\frac{1}{\sigma_t^2} + \sum_{s\neq t} \frac{\beta_{st}^2}{\sigma_s^2}\,\mathbb{1}_{\{t\to s\}} \right) y_t^2 + \right. \\
&\quad \left. +\, 2\sum_{v<t} \left(-\frac{\beta_{vt}}{\sigma_v^2}\,\mathbb{1}_{\{t\to v\}} - \frac{\beta_{tv}}{\sigma_t^2}\,\mathbb{1}_{\{v\to t\}} + \sum_s \frac{\beta_{sv}\,\beta_{st}}{\sigma_v^2}\,\mathbb{1}_{\{v\to s\,,\, t\to s\}} \right) y_v\, y_t \right] \\
&=\; -\frac{1}{2}\,\mathbf{y}'\,\mathbf{\Omega}\,\mathbf{y}
\end{aligned}
$$

where the third equality follows from Result 5 (b) and (c) and $\mathbf{\Omega}$ is a concentration matrix with entries

$$
\omega_{tv} = \begin{cases} \dfrac{1}{\sigma_t^2} + \sum_s \dfrac{\beta_{st}^2}{\sigma_s^2}\mathbb{1}_{\{t\to s\}}, & \text{for} \quad t = v \\[2ex] -\dfrac{\beta_{vt}}{\sigma_v^2}\mathbb{1}_{\{t\to v\}} - \dfrac{\beta_{tv}}{\sigma_t^2}\mathbb{1}_{\{v\to t\}} + \sum_s \dfrac{\beta_{sv}\,\beta_{st}}{\sigma_v^2}\mathbb{1}_{\{v\to s\,,\, t\to s\}}, & \text{for} \quad t \neq v \end{cases}
$$

Therefore,

$$\log p(\mathbf{q}\,,\,\mathbf{y}\mid\mathbf{m}) \;=\; g(\mathbf{q}) \;+\; \mathbf{h}'(\mathbf{q})\,\mathbf{y} \;-\; \mathbf{y}'\,\mathbf{K}(\mathbf{q})\,\mathbf{y}/2$$

with $\mathbf{K}(\mathbf{q}) = \mathbf{\Omega}$, $\mathbf{h}(\mathbf{q}) = \boldsymbol{\gamma}$ and $g(\mathbf{q}) = \log p(\mathbf{q}\mid\mathbf{m}) - \frac{1}{2}\,(T\log 2\pi - \log det(\mathbf{\Omega}) + \sum_{t=1}^T \mu_t^{\star\,2}/\sigma_t^2)$. Note that since $\mathbf{\Omega}$ is not a function of $\mathbf{q}$ our model is in the homogeneous conditional Gaussian parametric family (Lauritzen 1996). Finally, it follows from equation (16) that

$$\mathbf{y}\mid\mathbf{q} \;\sim\; N_T\left(\mathbf{\Omega}^{-1}\,\boldsymbol{\gamma}\,,\,\mathbf{\Omega}^{-1}\right)\;.$$

□

Proof of Result 2

Result 2. *Consider a class of Markov equivalent DAGs $\mathcal{G}$. Let Y_1 and Y_2 be any two adjacent nodes in the graphs in $\mathcal{G}$. Assume that for each such pair there exists at least one variable, Q, directly affecting Y_1 but not Y_2. Let $\mathcal{G}_E$ represent the class of extended graphs. Then the graphs in $\mathcal{G}_E$ are not Markov equivalent.*

Proof. Consider first subgraphs composed by Y_1, Y_2 and Q. Since Q is a cause of Y_1 but not of Y_2, the three variables can only be represented by two causal subgraphs: $Q \to Y_1 \leftarrow Y_2$ and $Q \to Y_1 \to Y_2$. Since both subgraphs have the same skeleton, $Q - Y_1 - Y_2$, but different sets of v-structures ($Q \to Y_1 \leftarrow Y_2$ for the first and no v-structure for the second) it follows from Theorem 1 that they are not Markov equivalent. Since by assumption this is true for any pair of adjacent nodes, Y_1 and Y_2, in any graph in $\mathcal{G}$, it follows that all graphs in $\mathcal{G}_E$ will have different sets of v-structures, and the result follows.

□

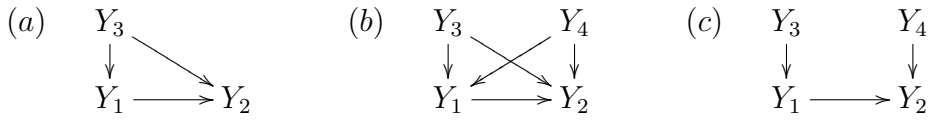
Proof of Result 3

Result 3. *For the HCGR parametric family, two DAGs are distribution equivalent if and only if they are Markov equivalent.*

Distribution equivalence imply Markov equivalence, but the converse is not generally true. For the linear regression with Gaussian errors parametric family, Markov equivalence does imply distribution equivalence (Heckerman and Geiger 1995). The proof for the HCGR model is a slightly modification of the proof for the Gaussian regression model. We therefore present the proof for later first. It is based on the following two results.

Theorem 2. *Two DAGs are Markov equivalent if and only if there exists a set of covered arrow reversals that transform one structure into the other (Chickering, 1995).*

An arrow reversal is a transformation from one DAG to the other where a single arrow between two nodes is reversed. An arrow between two nodes is said to be covered if the two nodes have the same parents when the arrow is removed. For example, the arrow from Y_1 to Y_2 is covered in graphs (a) and (b) but not in (c).



Theorem 3. *Consider the Gaussian linear regression model with local likelihoods*

$$y_t \mid \mathbf{y}_{pa(y_t)} \sim N\left(\mu_t + \sum_{k \in pa(y_t)} \beta_{tk}(y_k - \mu_k)\,,\, \sigma_t^2\right).$$

Given that node $j \in pa(y_t)$ and no other directed path exists from j to t,

$$E(y_t \mid \mathbf{y}_K) = \mu_t + \sum_{k \in K} (\beta_{tk} + \beta_{jk}\,\beta_{tj})(y_k - \mu_k), \quad \textit{and}$$

$$Var(y_t \mid \mathbf{y}_K) = \sigma_t^2 + \beta_{tj}^2\,\sigma_j^2,$$

where $K = (pa(y_t) \cup pa(y_j)) \setminus \{y_j\}$. *If neither* σ_j^2 *or* σ_t^2 *are zero, then*

$$E(y_j \mid y_t, \mathbf{y}_K) = \mu_j + \sum_{k \in K} (\beta_{jk} - (\beta_{tk} + \beta_{jk}\,\beta_{tj})\,\beta_{jt}^*)(x_k - \mu_k) + \beta_{jt}^*(x_t - \mu_t),$$

$$Var(y_j \mid y_t, \mathbf{y}_K) = \frac{\sigma_j^2\,\sigma_t^2}{\sigma_t^2 + \beta_{tj}^2\,\sigma_j^2}, \qquad \beta_{jt}^* = \frac{\beta_{tj}\,\sigma_j^2}{\sigma_t^2 + \beta_{tj}^2\,\sigma_j^2}.$$

with asterisk denoting the coefficients revised during the arrow reversal (Shachter and Kenley 1989, page 533).

The above Theorem shows that for the Gaussian linear regression parametric family a single covered arrow reversal is equivalent to a re-parametrization of the model. Consider for example graph (a). The joint distribution factors as $f(y_3)f(y_1 \mid y_3)f(y_2 \mid y_1, y_3)$ where

$$(y_3) \sim N\big(\mu_3, \sigma_3^2\big),$$
$$(y_1 \mid y_3) \sim N\big(\mu_1 + \beta_{13}(y_3 - \mu_3), \sigma_1^2\big),$$
$$(y_2 \mid y_1, y_3) \sim N\big(\mu_2 + \beta_{21}(y_1 - \mu_1) + \beta_{23}(y_3 - \mu_3), \sigma_2^2\big).$$

Now let's consider the arrow reversal $Y_1 \leftarrow Y_2$. From Theorem 3 we have $K = y_3$ and the components of re-parameterized model $f(y_3)f(y_2 \mid y_3)f(y_1 \mid y_2, y_3)$ normally distributed with expectations and variances

$$E(y_3) = \mu_3, \qquad Var(y_3) = \sigma_3^2,$$
$$E(y_2 \mid y_3) = \mu_2 + (\beta_{23} + \beta_{13}\beta_{21})(y_3 - \mu_3), \qquad Var(y_2 \mid y_3) = \sigma_2^2 + \beta_{21}^2\sigma_1^2,$$
$$E(y_1 \mid y_2, y_3) = \mu_1 + \left(\beta_{13} - (\beta_{23} + \beta_{13}\beta_{21})\frac{\beta_{21}\sigma_1^2}{\sigma_2^2 + \beta_{21}^2\sigma_1^2}\right)(y_3 - \mu_3) + \frac{\beta_{21}\sigma_1^2}{\sigma_2^2 + \beta_{21}^2\sigma_1^2}(y_2 - \mu_2),$$
$$Var(y_1 \mid y_2, y_3) = \frac{\sigma_1^2\sigma_2^2}{\sigma_2^2 + \beta_{21}^2\sigma_1^2}.$$

We now present the proof, given by Heckerman and Geiger (1995), for the linear regression model with Gaussian errors.

Proof. We want to show that for the Gaussian linear regression parametric family two models are distribution equivalent if and only if they are Markov equivalent. The direction, distribution equivalence implies Markov equivalence is a simple consequence that if two models are distribution equivalent than one is re-parametrization of the other, and the set of conditional independence/dependence relations entailed by them must be the same.

To get the direction, Markov equivalence implies distribution equivalence note that by Theorem 2, if two structures are Markov equivalent, then there exists a set of covered arrow reversals that transform one structure into the other. Theorem 3, by its turn, implies that two structures differing by a set of covered arrow reversals are distribution equivalent.

□

We now present the proof for the HCGR model.

Proof. **(Result 3)** Since we only allow arrow reversals between phenotypes (but not between QTLs and phenotypes), and the QTLs enter the HCGR model through the mean, the result follows by replacing μ_t by $\mu_t^\star = \mu_t + \mathbf{X}_t \boldsymbol{\theta}_t$ in Theorem 3. (Note we are dropping the i subscript from the notation in $\mu_{ti}^\star$ and $\mathbf{X}_{ti}$).

□

Proof of Result 4

Result 4. *For the HCGR parametric family, two DAGs are distribution equivalent if and only if they have the same skeletons and same sets of v-structures.*

Proof. Follows directly from Theorem 1 and Result 3.

□