CHAPTER 7

Bayesian Causal Phenotype Network Incorporating Genetic Variation and Biological Knowledge

JEE YOUNG MOON, ELIAS CHAIBUB NETO, XINWEI DENG, AND BRIAN S. YANDELL

In a segregating population, quantitative trait loci (QTL) mapping can identify QTLs with a causal effect on a phenotype. Several approaches in the literature take advantage of QTLs identified by QTL mapping, to determine causal relations among phenotypes. A common feature of these methods is that QTL mapping and phenotype network reconstruction are conducted separately. As both tasks have to benefit from each other, this chapter presents an approach which jointly infers a causal phenotype network and causal QTLs. The joint network of causal phenotype relationships and causal QTLs is modeled as a Bayesian network. Besides, a prior distribution on phenotype network structures is adjusted by biological knowledge. This integrative approach can incorporate several sources of biological knowledge such as protein–protein interactions, gene ontology annotations, transcription factor and DNA binding information. This framework allows a flexible tuning on the confidence of the various sources of knowledge through priors on biological knowledge weights. A Metropolis–Hastings scheme is described that iterates between accepting a network structure and accepting k weights corresponding to the k types of biological knowledge. The way to encode biological knowledge is described in the case of protein–protein interactions, similarity-based measures derived from a gene ontology, and transcription factor bindings to DNA. The integrative method is then applied to reconstruct a network involved in the cell cycle in yeast, relying on transcription factor binding knowledge.

Probabilistic Graphical Models for Genetics, Genomics, and Postgenomics. First Edition. Christine Sinoquet & Raphaël Mourad (Eds). Published in 2014 by Oxford University Press.

7.1 Introduction

A key interest in molecular biology is to understand how DNA, RNA, proteins, and metabolic products regulate each other. In this regard, people have considered constructing regulatory networks from microarray expression data with time-series measurements or transcriptional perturbations [14, 15]. A regulatory network can also be constructed in a segregating population where genotypes perturb the gene expression, protein, and metabolite levels. The genetic variation information can decipher genetic effects on traits and help discover causal regulatory relationships between phenotypes. In addition, knowledge of regulatory relationships is available in various biological databases, which can improve the reconstruction of causal networks. This chapter focuses on combining genetic variations in a segregating population and biological knowledge to improve the inference of causal networks.

Given the quantitative nature of a gene expression phenotype, one can perform quantitative trait loci (QTL) mapping to detect the genomic locations affecting the phenotype [29]. The genotypes at a location are often coded as *AA*, *Aa*, or *aa*, where allele *A* and *a* are distinct variant forms of a genetic locus. A quantitative phenotype/trait is any observable physical or biochemical quantitative feature of an organism such as weight, blood pressure, gene expression, or protein levels. The basic idea of QTL mapping is to detect genomic regions, or QTLs, where variation in genotype is associated with quantitative variation in phenotype. For example, tall parents tend to have tall children, whereas short parents tend to have short children. Then, it appears that there are genetic factors to be associated with the height, and the genetic factors can be identified by QTL mapping. In an experimental population, where genotypes are randomly assigned, the genetic variation at QTLs can be interpreted as causing later changes in the phenotype of interest.

In a segregating population, QTL mapping can identify QTLs with a causal effect on a phenotype. The causal effect can be direct from QTL to phenotype, or indirect via other intermediate phenotypes. We only label the direct QTLs as "causal QTLs," recognizing that they have a more proximal effect on a phenotype than indirect QTLs. We also acknowledge that there may be many other molecular factors in a pathway between the QTL and the phenotype that were not measured in a particular study. Indirect and direct QTLs can be used to help determine the direction of the edges in a causal phenotype network (i.e., a directed graph composed of phenotype nodes, whose edges represent causal relations). Several approaches in the literature take advantage of QTLs identified by QTL mapping to determine causal relations among phenotypes including: structural equation modeling [2, 34, 35] score-based methods for Bayesian networks [56, 60, 62]; causal algorithms for Bayesian networks based on independence tests [8, 53]; and causality tests on pairs of phenotypes [10, 11, 32, 38, 48]. A common feature of the above approaches is that QTL mapping and phenotype network reconstruction are conducted separately. QTL mapping without consideration of a phenotype network may find indirect QTLs. As pointed out by [9], incorrect or indirect QTLs may compromise the inference of causal relationships among phenotypes. To address this issue, several researchers [9, 20] proposed to jointly infer causal phenotype networks and causal QTLs.

Various sources of biological knowledge have been incorporated with gene expression in the reconstruction of phenotype networks, because it is difficult to determine the causal direction of gene regulation using expression data only. Transcription factor binding information was leveraged by [52], whereas [40] used protein–protein interaction knowledge to construct phenotype networks. Methods integrating multiple sorts of biological knowledge have been proposed by [25], [55], and [12].

In this chapter, we propose a Bayesian approach to jointly infer a causal phenotype network and causal QTLs with a prior distribution on phenotype network structures adjusted by biological knowledge. The joint network of causal phenotype relationships and causal QTLs is modeled as a Bayesian network[1] adopted from [9], QTLnet. Causal QTLs can be inferred by QTL mapping conditional on the phenotype network. Since the phenotype network is unknown, QTLnet traverses the space of phenotype networks and updates causal QTLs using Markov chain Monte Carlo (MCMC). We extend the framework of QTLnet by incorporating biological knowledge into the prior distribution on phenotype network structures. The incorporation of biological knowledge is expected to increase the accuracy of the model estimation, enhancing the predictive power of the network [62]. The prior probability on phenotype network structures is based on the Gibbs distribution to integrate different sources of biological information, allowing for flexible tuning of the analyst's confidence on this knowledge [55]. The consideration of reliability of biological knowledge is necessary since biological knowledge can be incomplete and inaccurate. While [62] proposed a method to incorporate genetic variation and biological knowledge to phenotype networks, their method does not consider the reliability of biological knowledge. Our proposed approach (QTLnet-prior) can integrate phenotype data, genetic variation, and several sources of biological knowledge (protein–protein interaction, gene ontology annotation, and transcription factor and DNA binding information) with the consideration of the reliability of each source of biological knowledge in the network reconstruction algorithm.

The details of our integrated framework for the joint inference of causal phenotype network and causal QTLs are organized as follows. Section 7.2 describes the QTLnet method for the joint inference of causal network and causal QTLs. Section 7.3 presents the proposed QTLnet-prior, which incorporates biological knowledge into the prior probability distribution of phenotype network structures. A simulation study is conducted in Section 7.4 to compare the proposed method with several existing approaches. In Section 7.5, the proposed method is used to reconstruct a network of 26 genes involved in the yeast cell cycle. Finally, in Section 7.6, we discuss the strengths and caveats of our approach and point out future research directions.

7.2 Joint Inference of Causal Phenotype Network and Causal QTLs

In Subsection 7.2.1, we first present a standard Bayesian network for modeling phenotype data. Next, in Subsection 7.2.2, we present an extended model, QTLnet, based on the homogeneous conditional Gaussian regression (HCGR) model, to incorporate QTL nodes into the phenotype network. Directed edges in the standard Bayesian network can be interpreted as causal relationships. By extending the phenotype network with causal QTL nodes, we can further claim causal interpretations. In Subsection 7.2.3, we present a rationale for the joint inference of the

[1] Note that Bayesian networks can be inferred in a Bayesian framework or a frequentist framework. Here, we take a Bayesian approach to infer a Bayesian network. The reason that the term "Bayesian" is used in a Bayesian network is described in [43]. The following is an excerpt from page 14 of [43]: "Bayesian networks, a term coined in Pearl (1985) to emphasize three aspects: (1) the subjective nature of the input information; (2) the reliance on Bayes's conditioning as the basis for updating information; and (3) the distinction between causal and evidential modes of reasoning, a distinction that underscores Thomas Bayes's paper of 1763.

causal phenotype network and causal QTLs, and in Subsection 7.2.4, we describe the QTL mapping conditional on the phenotype network. Finally, we give an overview of our joint approach for phenotype network and causal QTL inference in Subsection 7.2.5.

7.2.1 Standard Bayesian Network Model

A standard Bayesian network is a probabilistic graphical model whose conditional independence is represented by a directed acyclic graph (DAG). A node t in a DAG G corresponds to a random variable Y_t in the Bayesian network. A directed edge from node u to node v can supposedly represent that Y_v is causally dependent on Y_u, though an edge truly represents the conditional dependence. The local directed Markov property of Bayesian networks states that each variable is independent of its non-descendant variables conditional on its parent variables:

$$Y_t \perp\!\!\!\perp Y_{V \setminus de(t)} | Y_{pa(t)} \quad \text{for all } t \in V,$$

where V is the set of all nodes in a DAG, $de(t)$ is the set of descendants of node t, $pa(t)$ is the set of parents of node t and $Y_{pa(t)}$ is a set of variables indexed by $pa(t)$, that is, $\{Y_i : i \in pa(t)\}$. Assume the node index is ordered such that the index of descendants is always bigger than the index of their parents. Since $\{t-1, \ldots, 1\}$ is a set of non-descendants of node t and $pa(t)$ is included in the non-descendant set $\{t-1, \ldots, 1\}$, Y_t is independent of $Y_{\{t-1,\ldots,1\}}$ conditional on $Y_{pa(t)}$; that is, $P(Y_t|Y_{pa(t)})$ is equivalent to $\mathbb{P}(Y_t|Y_{t-1}, \ldots, Y_1)$. The joint distribution can be written as:

$$\begin{aligned} \mathbb{P}(Y_1, \ldots, Y_T) &= \prod_{t=1}^{T} \mathbb{P}(Y_t|Y_{t-1}, \ldots, Y_1) \\ &= \prod_{t=1}^{T} \mathbb{P}(Y_t|Y_{pa(t)}), \end{aligned} \tag{7.1}$$

where the first equality is satisfied by the chain rule in probability theory.[2]

[2] In probability theory, the chain rule permits that the joint probability of two variables X and Y can be written as

$$\mathbb{P}(X, Y) = \mathbb{P}(Y|X)\,\mathbb{P}(X) = \mathbb{P}(X|Y)\,\mathbb{P}(Y).$$

This can be extended to the joint probability of multiple variables:

$$\begin{aligned} \mathbb{P}(Y_T, \ldots, Y_1) &= \mathbb{P}(Y_T|Y_{T-1}, \ldots, Y_1)\mathbb{P}(Y_{T-1}, \ldots, Y_1) \\ &= \mathbb{P}(Y_T|Y_{T-1}, \ldots, Y_1)\mathbb{P}(Y_{T-1}|Y_{T-2}, \ldots, Y_1)\mathbb{P}(Y_{T-2}|Y_{T-3}, \ldots, Y_1) \\ &= \ldots \\ &= \textstyle\prod_{t=1}^{T} \mathbb{P}(Y_t|Y_{t-1}, \ldots, Y_1). \end{aligned}$$

7.2.2 HCGR Model

The parametric family of a Bayesian network that jointly models phenotypes and QTL genotypes corresponds to an HCGR model. Conditional on the QTL genotypes and covariates, the phenotypes are distributed according to a multivariate normal distribution, where QTLs and covariates enter the model via the mean, and the correlation structure among the phenotypes is explicitly modeled according to the DAG representing the phenotype network structure [9]. Fig. 7.1 depicts one example of a joint Bayesian network of phenotypes and QTL genotypes.

The HCGR model is derived from a series of linear regression equations. For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, let Y_{ti} be the value of phenotype for individual i and trait t. Then we assume for each phenotype that Y_{ti} can be modeled as follows:

$$Y_{ti} = \mu^{\star}_{ti} + \sum_{v \in pa(t)} \beta_{tv} Y_{vi} + \epsilon_{ti}, \quad \epsilon_{ti} \sim \mathcal{N}(0, \sigma^2_t). \tag{7.2}$$

The model can be decomposed into three parts: a genetic part ($\mu^{\star}_{ti}$), a phenotypic part ($\sum_{v \in pa(t)} \beta_{tv} Y_{vi}$), and an error term ($\epsilon_{ti}$). In the phenotypic part, β_{tv} is the effect of parent phenotype v on phenotype t. The error term, ϵ_{ti}, follows a normal distribution. The genetic part, $\mu^{\star}_{ti}$, corresponds to a model of QTL genotypes and possibly covariates:

$$\mu^{\star}_{ti} = \mu_t + \sum_{k=1}^{C} \vartheta_{tk} Z_{ki} + \sum_{k=1}^{K} \gamma_{tk} \theta_{tk} X_{ki},$$

where μ_t is the overall mean for trait t, Z_{ki} represents a covariate, ϑ_{tk} represents the effect of the covariate on the phenotype, and $\sum_{k=1}^{K} \gamma_{tk} \theta_{tk} X_{ki}$ is the overall effect of QTLs. For the simplicity, we will not consider the covariates Z later on. The parameter γ_{tk} is unknown, and represents the inclusion ($\gamma_{tk} = 1$) or exclusion ($\gamma_{tk} = 0$) of the QTL located at the genomic position k, into the model. The genetic effects of QTL can be partitioned into different types of genetic effects, e.g. additive and dominance effects, and hence the genotype of the QTL is coded into the variables to estimate the different genetic effects. The vector X_{ki} represents a column vector of coded variables of the genotype at the genomic location k for individual i, and the vector θ_{tk} is a row vector of several types of genetic effects of QTL at the location k on phenotype t. The coding of a genotype may follow Cockerham's genetic model [30]. For example, in an intercross, the segregating genotypes at a locus are denoted by AA, Aa, or aa, and we can code the genotype into an additive variable by the number of A alleles in the genotype minus 1 and a dominance variable by 1/2 if it is Aa and –1/2 otherwise. In this case, the additive effect is the effect of substituting one allele a with another allele A and the dominance effect is the deviation of Aa from the mean of AA and aa. Accordingly, in an intercross, X_{ki} is a column vector of additive and dominance coding variables,

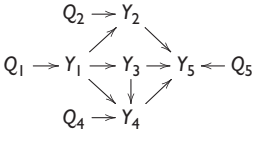

Fig 7.1 Example network with five phenotypes ($Y_1, \ldots, Y_5$) and five QTLs ($Q1, \ldots, Q5$).

and θ_{tk} is the row vector of additive and dominance effects on phenotype t. It was shown by [9] that these linear regression equations in equation (7.2) set a HCGR model for phenotypes and QTL genotypes.

7.2.3 Systems Genetics and Causal Inference

Systems genetics aims to understand the complex interrelations between genetic variations and phenotypes from large-scale genotype and phenotype data [39]. Here we explain how the systems genetics approach can infer causal networks. Causal relations from QTLs to phenotypes are justified by the unidirectional influence of the genotype on phenotype and the random allocation of genotypes to individuals. In contrast, causal relations among phenotypes are induced from conditional independence. The key idea of systems genetics is that by incorporating QTL nodes into phenotype networks, we create new sets of conditional independence relationships for distinguishing network structures that would, otherwise, belong to the same equivalence class (see Tables 7.1 and 7.2).

First, we give a more detailed description for the causal relations between QTLs and phenotypes. As stated in the central dogma of molecular biology, the hereditary DNA information is transferred to phenotypes. Thus a genotype influences phenotypes in general but not the other way around. A genotype is assumed to be randomized to other environmental factors by independent segregation of chromosomes in meiosis and random mating between gametes. These special characteristics enable us to infer causal effects of QTLs on phenotypes since, by analogy with a randomized experiment, we have that: (1) the treatment (genotype) to an experimental unit precedes the measured outcome (phenotype), and (2) random allocation of treatments to experimental units guarantees that other common causes get averaged out. Two loci on the same chromosome are highly correlated when their distance is small. But crossovers between two loci can still occur randomly in proportion to the distance. One can distinguish the true causal QTL and false nearby QTL with a large sample size. This random allocation is explicit in an experimental cross such as a backcross or an intercross.[3] While this idea can be extended to natural populations, special attention must be paid to admixture, kinship, and other forms of relatedness.

Second, the explanation of causal inference among phenotypes requires the concept of conditional independence in DAGs composed of phenotypes and QTL nodes. In the next three paragraphs, we present some definitions and results that allow us to infer phenotype-to-phenotype causal relationships.

Here are definitions. In graph theory, a *path* is defined as any unbroken, non-intersecting sequence of edges in a graph, which may go along or against the direction of arrows. We say that a path p is *d-separated* [42, 43] 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 is in Z, or (2) p contains a 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. We say that

[3] An experimental cross is generated by crossing inbred lines. An inbred line is obtained by repeated generations of inbreedings so that any genotype of the inbred line is homozygous, *AA*. Therefore, breeding within the inbred line produces genetically identical offspring to its parents. Both backcross and intercross first produce the first generation of population by mating two different inbred lines, *AA* and *BB*. The first generation is identical to each other with heterozygous genotypes, *AB*. The backcross population is produced by mating the first generation to one of its parental inbred lines such as *AA*. Then, the backcross population has genotypes either *AA* or *AB* in a ratio of 1 : 1. The intercross population is produced by mating the first generation itself so that it has genotypes *AA*, *AB*, or *BB* in a ratio of 1 : 2 : 1.

Z d-separates X from Y if and only if Z blocks every path from a node in X to a node in Y. The *skeleton* of a DAG 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.

The equivalence concept plays a key role in learning the structure of networks from the data. Here we present three important equivalence relations for graphs or statistical models of graphs. Two graphs are *Markov equivalent* if they have the same set of d-separation relations [50]. Two structures m_1 and m_2 for Y are *distribution equivalent* with respect to the distribution family F if they represent the same joint distributions for Y, that is, for every θ_1, there exists a θ_2 such that $\mathbb{P}(Y \mid \theta_1, m_1) = \mathbb{P}(Y \mid \theta_2, m_2)$ [22]. In other words, m_1 and m_2 are distribution equivalent if the parameters θ_1 and θ_2 are simple reparametrizations of each other. If m_1 and m_2 are distribution equivalent, then the invariance principle of maximum likelihood estimates guarantees $\mathbb{P}(Y \mid \hat{\theta}_1, m_1) = \mathbb{P}(Y \mid \hat{\theta}_2, m_2)$, and m_1 and m_2 cannot be distinguished using the data. In this case we say that m_1 and m_2 are *likelihood equivalent*. In a Bayesian setting, we define likelihood equivalence using the prior predictive distribution, $\int \mathbb{P}(Y \mid \theta, m_1)\, \mathbb{P}(\theta \mid m_1)\, d\theta = \int \mathbb{P}(Y \mid \theta, m_2)\, \mathbb{P}(\theta \mid m_2)\, d\theta$. If models m_1 and m_2 are distribution equivalent and we adopt a proper prior $\mathbb{P}(\theta \mid m)$, it is often reasonable to expect $\mathbb{P}(Y \mid m_1) = \mathbb{P}(Y \mid m_2)$, so that we cannot distinguish m_1 and m_2 for any data set Y [22].

Now we state four important results regarding causal inference in systems genetics: (1) Two DAGs are Markov equivalent if and only if they have the same skeletons and the same set of v-structures [54]; (2) Distribution equivalence implies Markov equivalence, but the converse is not necessarily true [50]; (3) For a Gaussian regression model, Markov equivalence implies distribution equivalence [21]; (4) For the homogeneous conditional Gaussian regression model, Markov equivalence implies distribution equivalence [9].

Therefore, for the HCGR parametric family, two DAGs are distribution and likelihood equivalent if and only if they are Markov equivalent. This implies that we can simply check if any two DAGs have the same skeleton and the same set of v-structures in order to determine if they are likelihood equivalent and hence cannot be distinguished using the data.

Getting back to the idea of causal inference among phenotypes, let G_Y be a phenotype network represented by a standard Bayesian network of phenotypes, Y. Phenotype data alone can distinguish some network structures by its likelihood but may fail to distinguish some other network structures. For example, consider the three network structures in Table 7.1. Models G_Y^1 and G_Y^3 have the same skeleton ($Y_1 - Y_2 - Y_3$) and the same set of v-structures (no v-structure) and, thus, are distribution/likelihood equivalent. Model G_Y^2, on the other hand, has the same skeleton but a different set of v-structures and, hence, is not distribution/likelihood equivalent to models G_Y^1 and G_Y^3. Therefore, phenotype data alone can identify G_Y^2 but cannot distinguish G_Y^1 and G_Y^3.

Table 7.1 Models G_Y^1 and G_Y^3 are distribution/likelihood equivalent.

DAG structures	Skeletons	v-structures
$G_Y^1 = Y_1 \rightarrow Y_2 \rightarrow Y_3$	$Y_1 - Y_2 - Y_3$	$\emptyset$
$G_Y^2 = Y_1 \rightarrow Y_2 \leftarrow Y_3$	$Y_1 - Y_2 - Y_3$	$Y_1 \rightarrow Y_2 \leftarrow Y_3$
$G_Y^3 = Y_1 \leftarrow Y_2 \rightarrow Y_3$	$Y_1 - Y_2 - Y_3$	$\emptyset$

Table 7.2 Extended models G^1 and G^3 are no longer distribution/likelihood equivalent.

Extended DAG Structures	Skeletons	v-Structures
$G^1 = Q_1 \rightarrow Y_1 \rightarrow Y_2 \rightarrow Y_3$	$Q - Y_1 - Y_2 - Y_3$	$\emptyset$
$G^3 = Q_1 \rightarrow Y_1 \leftarrow Y_2 \rightarrow Y_3$	$Q - Y_1 - Y_2 - Y_3$	$Q \rightarrow Y_1 \leftarrow Y_2$

Adding causal QTL nodes to a phenotype network allows the inference of causal relationships between phenotypes that could not be distinguishable using phenotype data alone. For example, if we add a causal QTL Q_1 to Y_1 in phenotype networks G^1_Y and G^3_Y in the above example, then the corresponding extended network structures G^1 and G^3 have different v-structures as shown in Table 7.2.

7.2.4 QTL Mapping Conditional on Phenotype Network Structure

Now we examine the inference of QTLs conditional on a phenotype network. QTL mapping can be done in a conditional or unconditional fashion. In the unconditional mapping analysis, we measure the association of a trait Y_t and QTL Q using the LOD score (logarithm of odds):

$$LOD(y_t, q) = \log_{10}\left(\frac{f(y_t \mid q)}{f(y_t)}\right),$$

where $f(y_t \mid q)$ represents the predictive density of a linear model with Q as an independent variable and $f(y_t)$ the predictive density of the baseline model. Here a predictive density is given by a maximized likelihood in a frequentist setting, or by the prior predictive density in a Bayesian setting. A high LOD score means that Y_t and Q are associated. Note that unconditional analysis can detect QTLs that directly affect the phenotype under investigation, as well as QTLs with indirect effects [9]. For example, if we consider the causal network of phenotypes and QTLs in Fig. 7.1, then the unconditional QTL mapping of Y_2 detects a direct QTL Q_2 as well as an indirect QTL Q_1 that affects Y_2 via Y_1. Fig. 7.2 shows the expected results of the unconditional analysis for each phenotype.

The conditional mapping analysis, on the other hand, incorporates other traits as covariates, and measures the association of Y_t and Q conditional on these covariates (say y_z) using the conditional LOD score:

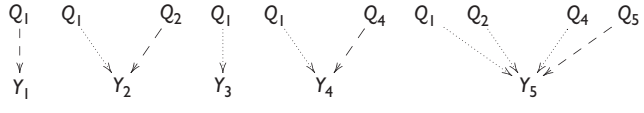

Fig 7.2 Output of the unconditional QTL mapping analysis for the phenotypes in Fig. 7.1. Dashed and pointed arrows represent direct and indirect QTL/phenotype causal relationships, respectively.

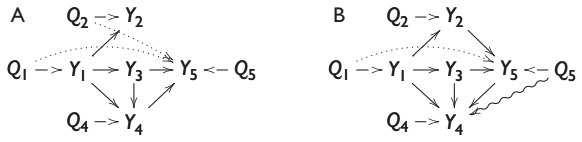

Fig 7.3 QTL mapping tailored to the network structure. Dashed, pointed, and wiggled arrows represent, respectively, direct, indirect, and incorrect QTL/phenotype causal relationships. (A) Mapping analysis of Y_5 conditional on Y_3 and Y_4 still detects Q_1 and Q_2 as QTLs for Y_5, since failing to condition on Y_2 leaves the paths $Q_1 \to Y_1 \to Y_2 \to Y_5$ and $Q_2 \to Y_2 \to Y_5$ in Fig. 7.1 open. In other words, (Y_3, Y_4) cannot d-separate (Q_1, Q_2) from Y_5 in the true causal graph. (B) Mapping analysis of Y_4 conditional on Y_1, Y_3, and Y_5 incorrectly detects Q_5 as a QTL for Y_4 because in the true network, the paths $Y_4 \to Y_5 \leftarrow Q_5$ and $Y_4 \leftarrow Y_3 \to Y_5 \leftarrow Q_5$ in Fig. 7.1 are open when we condition on Y_5.

$$\begin{aligned} LOD(y_t, q|y_z) &= \log_{10}\left(\frac{f(y_t|q, y_z)}{f(y_t)}\right) - \log_{10}\left(\frac{f(y_t|y_z)}{f(y_t)}\right) \\ &= LOD(y_t, q, y_z) - LOD(y_t, y_z). \end{aligned}$$

Now, consider QTL mapping analysis tailored to a known phenotype network structure. In this situation, we can avoid detecting indirect QTLs by simply performing mapping analysis of the phenotypes conditional on their parents. For instance, in Fig. 7.1, if we perform QTL mapping of Y_5 conditional on Y_2, Y_3 and Y_4, we do not detect Q_1, Q_2, and Q_4 because of the following independence relations: $Y_5 \perp\!\!\!\perp Q_1 \mid Y_2, Y_3, Y_4$; $Y_5 \perp\!\!\!\perp Q_2 \mid Y_2, Y_3, Y_4$; and $Y_5 \perp\!\!\!\perp Q_4 \mid Y_2, Y_3, Y_4$. We only detect Q_5 due to the following relation: $Y_5 \not\perp\!\!\!\perp Q_5 \mid Y_2, Y_3, Y_4$.

In practice, however, the structure of the phenotype network is unknown, and performing QTL mapping conditional on a misspecified phenotype network structure can result in the inference of misspecified causal QTLs, as shown in Fig. 7.3. The mapping analysis of a phenotype conditional on downstream phenotypes in the true network, induces dependences between the phenotype and QTLs affecting downstream phenotypes. This leads to the erroneous inference that the phenotype includes downstream QTLs as its QTLs. For example, the mapping analysis of Y_4 conditioning on Y_1, Y_3 and a downstream phenotype Y_5 includes downstream Q_5 as its QTLs in Fig. 7.3B. However, a model with misspecified phenotype network and QTLs will generally have a lower marginal likelihood score than the model with the correct causal order for the phenotypes and correct QTLs. Since in practice QTLnet adopts a model selection procedure to traverse the space of network structures, it tends to prefer models closer to the true data-generating process. Simulation studies presented in [9] corroborate this point.

Note that, as pointed out in [9], the conditional LOD score can be adopted as a formal measure of independence between a phenotype and QTLs. Even though we restrict our attention to HCGR models, conditional LOD profiling is a general framework for the detection of conditional independences between continuous and discrete random variables. Contrary to partial correlations, the conditional LOD score does not require the assumption of multinormality of the data in order to formally test for independence, and it can handle QTL by covariate interactions.

7.2.5 Joint Inference of Phenotype Network and Causal QTLs

Subsection 7.2.4 describes the QTL mapping conditional on a phenotype network. In practice, the phenotype network is generally unknown and we cannot directly infer the correct causal QTLs. Therefore, we need to perform a joint inference of phenotype network and causal QTLs.

Recall that Y are phenotypes, X are genetic variations as defined in Subsection 7.2.2, and G is a Bayesian network structure of phenotypes and QTLs. Let G_Y represent a phenotype network and let $G_{Q\to Y}$ represent a graph from causal QTL nodes to phenotype nodes. Note that G_Y and $G_{Q\to Y}$ are subgraphs of the extended network structure G. Conforming to the HCGR model in equation (7.2), G_Y corresponds to the collection of causal relations from $pa(t)$ to trait t and $G_{Q\to Y}$ corresponds to the collection of causal relations from non-zero γ_{tk} to traits. Denote θ_G to be the parameter sets $(\beta_{tv}, \sigma_t^2, \mu_t, \theta_{tk})$. From equation (7.2), the likelihood of a Bayesian network of phenotypes and causal QTLs can be written as a product of normal densities:

$$\begin{aligned}\mathbb{P}(Y|G, X, \theta_G) &= \mathbb{P}(Y|G_Y, G_{Q\to Y}, X, \theta_G)\\ &= \prod_{t=1}^{T}\prod_{i=1}^{n} \mathcal{N}\left(\mu_{ti}^{\star} + \sum_{y_k \in pa(y_t)} \beta_{tk} y_{ki}\,,\, \sigma_t^2\right).\end{aligned}$$

The marginal likelihood of phenotypes and causal QTLs $P(Y|G, X)$ is calculated by integrating parameters θ_G out in the Bayesian network:

$$\mathbb{P}(Y|G, X) = \int \mathbb{P}(Y|G, X, \theta_G)\, \mathbb{P}(\theta_G|G)\, d\theta_G.$$

The posterior probability of G conditional on the data is given by:

$$\mathbb{P}(G|Y, X) = \frac{\mathbb{P}(Y|G, X)\mathbb{P}(G)}{\sum_G \mathbb{P}(Y|G, X)\mathbb{P}(G)},$$

where $\mathbb{P}(G)$ represents the prior probability of the network structure G. In the next section, we devote our attention to the specification of $\mathbb{P}(G)$ using integrated biological knowledge.

Following [9], we adopt the QTLnet framework that jointly infers the phenotype network structure and causal QTLs. Most of the current literature in genetical network reconstruction has treated the problems of QTL inference and phenotype network reconstruction separately, generally performing QTL inference first, and then using QTLs to help determine the phenotype network structure [8, 62]. As indicated in Subsection 7.2.4, such a strategy can include QTLs with indirect effects into the network.

7.3 Causal Phenotype Network Incorporating Biological Knowledge

Besides the causal QTLs, biological knowledge is another useful and important information resource to enhance the construction of the phenotype network. Such knowledge can be integrated on top of the causal network to provide a more comprehensive picture of how genes are regulated. This integrated network could generate a new hypothesis on gene regulation, having an overall consistency with biological knowledge.

In this section, we propose a network inference method, QTLnet-prior, from phenotype data with genetic variations, integrating biological knowledge. The QTLnet-prior extends the framework of QTLnet referred to at the end of Subsection 7.2.5. It specifies the prior probability on phenotype network structures to integrate multiple sources of biological knowledge with flexible tuning parameters on confidence of knowledge [55]. The weighted integration of biological knowledge could produce a more predictive Bayesian network. The details of our extended framework, QTLnet-prior, are presented in Subsection 7.3.1. In Subsection 7.3.2, we sketch a Metropolis–Hastings MCMC scheme for QTLnet-prior implementation that integrates the sampling of network structures [19, 37], QTL mapping, and sampling of biological knowledge weights. In Subsection 7.3.3, we present how to encode biological knowledge into the prior distribution over phenotype network structures.

7.3.1 Model

EXTENDED MODEL

Denote by G a Bayesian network structure of phenotypes and QTLs. The graph G consists of a phenotype network (G_Y) and causal QTLs to phenotypes ($G_{Q\to Y}$). Let Y be phenotype data, X be genetic variations, and W represent weights set on various sources of biological knowledge B. The biological knowledge B is considered to be relations between phenotypes such as transcription factor binding, protein–protein interaction, and gene ontology annotation; that is, biological knowledge B can give a prior probability only for the phenotype network G_Y. The QTLnet framework presented in Section 7.2 assumes intrinsically a uniform prior over phenotype network structures. Additionally, we specify a prior distribution on the weights of biological knowledge in order to control the consistency between phenotype data and knowledge. Because the prior information can be inaccurate or incompatible with the phenotype data, it is important to quantify its uncertainty. We write the extended model as follows:

$$\begin{aligned}
\mathbb{P}(G, W|Y, X, B) &\propto \mathbb{P}(Y|G, W, X, B)\mathbb{P}(G, W|X, B) \\
&= \mathbb{P}(Y|G, X)\mathbb{P}(G, W|X, B) \\
&= \mathbb{P}(Y|G, X)\mathbb{P}(G_Y, W|X, B)\mathbb{P}(G_{Q\to Y}|X, B) \\
&= \mathbb{P}(Y|G, X)\mathbb{P}(G_Y, W|B)\mathbb{P}(G_{Q\to Y}|X) \\
&= \mathbb{P}(Y|G, X)\mathbb{P}(G_Y|B, W)\mathbb{P}(W|B)\mathbb{P}(G_{Q\to Y}|X). \qquad (7.3)
\end{aligned}$$

In the first step, the posterior probability of a network G and weights W is calculated by multiplying the marginal likelihood $\mathbb{P}(Y|G, W, X, B)$ of the traits given the network G and the prior probability $\mathbb{P}(G, W|X, B)$ of a network and weights given genetic variations and biological knowledge. The marginal likelihood $\mathbb{P}(Y|G, W, X, B)$ can be simplified to be $\mathbb{P}(Y|G, X)$ as in the second step. In the third step, the prior probability $\mathbb{P}(G, W|X, B)$ can be decomposed into $\mathbb{P}(G_Y, W|X, B)$ and $\mathbb{P}(G_{Q\to Y}|X, B)$ by assuming the independence between a phenotype network G_Y along with the weights W and causal QTLs $G_{Q\to Y}$ given genetic variations X and biological knowledge B. The fourth step is provided by the fact that $\mathbb{P}(G_Y, W|X, B)$ is equal to $\mathbb{P}(G_Y, W|B)$ because the genetic variations are not included in the structure of the phenotype network G_Y, and $\mathbb{P}(G_{Q\to Y}|X, B)$ is equal to $\mathbb{P}(G_{Q\to Y}|X)$ because B affects G_Y but not $G_{Q\to Y}$. The extended model in equation (7.3) shows that prior distributions on phenotype network structure $\mathbb{P}(G_Y|B, W)$, biological knowledge

weights $\mathbb{P}(W|B)$, and causal QTLs of traits $\mathbb{P}(G_{Q\to Y}|X)$ must be specified. We will describe how to set $\mathbb{P}(G_Y|B, W)$, $\mathbb{P}(W|B)$, and $\mathbb{P}(G_{Q\to Y}|X)$ in the following.

PRIOR ON PHENOTYPE NETWORK STRUCTURES $\mathbb{P}(G_Y|B, W)$

Incorporation of a priori biological knowledge into a prior on network structures can help to discriminate Bayesian networks having the same likelihood [55, 61]. If G^1 and G^2 have the same likelihood ($\mathbb{P}(Y|G^1) = \mathbb{P}(Y|G^2)$) but have different prior probabilities ($\mathbb{P}(G^1) \neq \mathbb{P}(G^2)$), the posterior probabilities would become different ($\mathbb{P}(G^1|Y) \neq \mathbb{P}(G^2|Y) \propto \mathbb{P}(Y|G^2)\mathbb{P}(G^2)$). For example, consider two graphs for nodes t and v: one is $t \to v$ and the other is $v \to t$. Their likelihoods are the same because they are Markov equivalent. If a prior indicates that one direction ($t \to v$) is more likely than the other direction ($t \leftarrow v$), then the posterior of one direction ($t \to v$) becomes higher than the other direction ($v \to t$). The biological knowledge B along with its weight W can therefore give different prior probabilities $\mathbb{P}(G_Y|B, W)$ for the phenotype network G_Y.

Various types of information can supplement the learning of a phenotype network. We can encode this supplementary information into unequal priors on network structures. A transcription factor binding location can be used to prefer the direction from a transcription factor to the target gene [4]. Pathway information can also guide to infer directions among phenotypes [55]. For example, consider a network with three nodes t, v, and u, where a path from t to v is known. Then, we can at least distinguish these two relations: $t \to v \leftarrow u$ and $t \to v \to u$. Regulation inference [41, 44, 57] from knockout data and protein–protein interaction [26] can be used as a prior for network structure. We will describe how to encode this information in Subsection 7.3.3. Since QTLnet is a Bayesian approach, we can flexibly incorporate various sources of biological knowledge by constructing meaningful priors for the network structures.

Now, it remains to set the prior distribution on phenotype network structure G_Y with respect to biological knowledge B. Since a Bayesian network distribution can be factored by its parent–child relations $\prod_t \mathbb{P}(Y_t|Y_{pa(t)})$, it is natural to assume the prior on DAG structures to be factored by its parent–child relations. Adapting the prior formulation over network structures in [55], we will show below that the prior satisfies the parent–child factorization.

Let us define the *energy* of a phenotype network G_Y relative to the biological knowledge B to be

$$\mathcal{E}(G_Y) = \sum_{i,j=1}^{T} |B(i,j) - G_Y(i,j)|, \tag{7.4}$$

where B is an encoding meant to describe biological knowledge ranging from 0 to 1 and G_Y is represented by the adjacency matrix of a network structure. The adjacency matrix is a 0-1 matrix which assigns $G_Y(i,j)$ to be 1 if there is a directed edge from node i to j, and to be 0 otherwise. The energy $\mathcal{E}(G_Y)$ acts as a distance measure between biological knowledge and a network structure G_Y. For a fixed biological knowledge matrix B, network structures will have small energy if they agree with the biological knowledge, and will have large energy if they disagree with the knowledge.

The energy can be decomposed into the sum of local pseudo-energies defined by parent–child relations for each trait:

$$\mathcal{E}(G_Y) = \sum_{j=1}^{T} \left(\sum_{i \in pa(j)} (1 - B(i,j)) + \sum_{i \notin pa(j)} B(i,j) \right)$$
$$= \sum_{j=1}^{T} \left(\frac{|B|}{T} + \sum_{i \in pa(j)} (1 - 2B(i,j)) \right) = \sum_{j=1}^{T} \left(\frac{|B|}{T} + \mathcal{E}_{j,pa(j)}(G_Y) \right),$$

where $|B| = \sum_{i,j=1}^{T} B(i,j)$ and $\mathcal{E}_{j,pa(j)}(G_Y) = \sum_{i \in pa(j)} (1 - 2B(i,j))$, which is the local pseudo-energy defined by phenotype j and its parents. Therefore, the prior distribution on network structures can be constructed in terms of energy, and it is shown to be the Gibbs distribution factorized by parent–child relations:

$$\mathbb{P}(G_Y|B,W) = \frac{\exp(-W\mathcal{E}(G_Y))}{Z(W)} \tag{7.5}$$
$$= \frac{\prod_{j=1}^{T} \exp(-W\mathcal{E}_{j,pa(j)}(G_Y))}{Z'(W)}, \quad G_Y \in \text{DAG}$$

where $Z(W)$ is a normalizing constant given by $\sum_{G_Y \in \text{DAG}} \exp(-W\mathcal{E}(G_Y))$ and $Z'(W)$ is another normalizing constant given by $Z(W)/\exp(-W|B|)$. For a fixed W, network structures with small energy will have higher prior probabilities than network structures with large energy. The weight W of biological knowledge B is introduced to tune the confidence of biological information which sometimes can be inaccurate or incompatible with expression data. As W goes toward 0, the influence of a priori knowledge becomes negligible, and the prior distribution of network structure is assumed to be almost uniform. On the contrary, as W goes to the infinity, the prior on network structure peaks at the biological knowledge.

Multiple sources of biological knowledge can be integrated into a prior on network structures with different weights:

$$\mathbb{P}(G_Y|B,W) = \frac{\exp(-\sum_k W_k \mathcal{E}_k(G_Y))}{Z(W)}, \quad G_Y \in \text{DAG}$$

where B_k is an encoding matrix of biological knowledge from source k, B is the vector of biological knowledge matrices $(B_1, \ldots, B_k)$, W_k is the weight of B_k, W is the weight vector $(W_1, \ldots, W_k)$, and $Z(W)$ is the summation of the numerator over all DAGs.

PRIOR ON BIOLOGICAL KNOWLEDGE WEIGHTS $\mathbb{P}(W|B)$

The weight parameter is introduced to control the influence of biological knowledge on the phenotype network. A higher value of the weight would increase the influence of the biological knowledge on the posterior distribution of networks. Specifically, a large W puts significant prior probability on the phenotype network structures which consistently agree with biological knowledge B. Conversely, a small W puts fairly equal prior probabilities on all possible networks. If biological knowledge B is similar to the true network from which the expression data are generated, then the posterior probability will peak at high W. On the contrary, if biological knowledge deviates substantially from the true network, the posterior will peak at small W. This happens because a smaller W leads to a smaller ratio of prior probabilities of the deviated network and the

true network. Consequently, the posterior of the true network can be larger than the posterior of the deviated network by the virtue of likelihood ratio overcoming the prior ratio at a small W.

For each biological knowledge B_k, we specify the prior probability distribution of the weight W_k to be an exponential distribution such that

$$\mathbb{P}(W_k|B_k) = \phi \exp(-\phi W_k), \tag{7.6}$$

with the rate parameter ϕ. Such an exponential prior for W_k has several advantages. First, it does not impose an upper bound on W_k. Second, it would not allow the weight to go to infinity too easily, since an infinite weight always results in a network closer to the biological knowledge regardless of expression data. Third, when biological knowledge is inaccurate or incompatible with expression data, the exponential distribution can control the contribution of negative biological knowledge more easily than a uniform distribution. The rate parameter ϕ is set to be 1 in our simulation because this rate balances the prior and likelihood well in the empirical study.

PRIOR ON CAUSAL QTLS $\mathbb{P}(G_{Q\to Y}|X)$

Without any specific information about the causal QTLs, we set the prior of causal QTLs to be a uniform distribution. Several alternative specifications can be found in Bayesian QTL mapping such as in [59] and in [58].

7.3.2 Sketch of MCMC

A main challenge in the reconstruction of networks is that the graph space grows superexponentially with the number of nodes. Exhaustive search over all network structures is impractical even for small networks. Hence, heuristic approaches are needed to efficiently traverse the graph space. We adopt a Metropolis-Hastings MCMC scheme that integrates the sampling of network structures [24, 37], QTL mapping, and the sampling of biological knowledge weights W. The MCMC scheme iterates between accepting a network structure G and accepting k weights $W_1, \cdots, W_k$ corresponding to k types of biological knowledge.

1. Sample a new phenotype network structure G_Y^{new} from a network structure proposal distribution $R(G_Y^{new}|G_Y^{old})$.
2. Given the phenotype network structure G_Y^{new}, sample a new set of causal QTLs $G_{Q\to Y}$ from a QTL proposal distribution $R(G_{Q\to Y}^{new}|G_{Q\to Y}^{old})$.
3. Accept the new extended network structure G^{new} composed of G_Y^{new} and $G_{Q\to Y}^{new}$ given the biological knowledge weights W with a probability

$$A_G = \min\left\{1, \frac{\mathbb{P}(Y|G^{new}, X)\mathbb{P}(G_Y^{new}|B, W)\mathbb{P}(G_{Q\to Y}^{new}|X)}{\mathbb{P}(Y|G^{old}, X)\mathbb{P}(G_Y^{old}|B, W)\mathbb{P}(G_{Q\to Y}^{old}|X)} \times \frac{R(G_Y^{old}|G_Y^{new})R(G_{Q\to Y}^{old}|G_{Q\to Y}^{new})}{R(G_Y^{new}|G_Y^{old})R(G_{Q\to Y}^{new}|G_{Q\to Y}^{old})}\right\}.$$

4 For each biological knowledge k,

(a) Sample a new weight W_k^{new} for biological knowledge B_k from a weight proposal distribution $R(W_k^{new}|W_k^{old})$.

(b) Accept the new biological weight W_k^{new} given the phenotype network G_Y with a probability

$$A_{W_k} = \min\left\{1, \frac{\mathbb{P}(G_Y|W_k^{new}, W_{-k}^{old}, B)}{\mathbb{P}(G_Y|W^{old}, B)} \frac{\mathbb{P}(W_k^{new}|B)}{\mathbb{P}(W_k^{old}|B)} \frac{R(W_k^{old}|W_k^{new})}{R(W_k^{new}|W_k^{old})}\right\}.$$

5 Iterate steps 1–4 until the chain converges.

In step 1, a new phenotype network structure is proposed by a mixture of single edge operations (single edge addition, single edge deletion, single edge reversal) and edge reversal moves with orphaning [19]. The edge reversal move with orphaning consists of selecting an edge $i \to j$, removing the parents of each node on the selected edge, sampling new parents of node i (including node j), and sampling new parents of node j, as long as it does not make a cycle. It has been shown that edge reversal moves can significantly improve the convergence of an MCMC sampler [19]. The proposal distribution puts the same probability, summing to 1, to the graphs that can be reached by a corresponding edge move.

In step 2, causal QTLs can be sampled conditional on the phenotypes' parents. There are several ways to sample causal QTLs. One way is a Bayesian QTL mapping proposed in [59] for each phenotype. The prior distribution for the indicators of QTLs is $\prod w_k^{\gamma_k}(1-w_k)^{\gamma_k}$ where $w_k = \mathbb{P}(\gamma_k = 1)$ is the prior inclusion probability for the k^{th} QTL. We can use this independent prior for the prior distribution and the proposal distribution for a causal QTL. Another way is the classical interval mapping of QTL for each phenotype conditional on its phenotypic parents. The classical interval mapping regresses a phenotype on a single QTL and picks every QTL over the significance threshold computed by permutations. Thus, this approach is deterministic as it chooses the same set of QTLs given the same set of parent phenotypes. It is a fast algorithm approximating the Bayesian mapping of QTL though it might fail to satisfy the irreducibility of the Markov Chain. We use the interval mapping for practical reasons.

In step 3, the computation of the ratio of marginal likelihoods, or Bayes factor, $\mathbb{P}(Y|G^{new}, X)/\mathbb{P}(Y|G^{old}, X)$, can be approximated by the difference of BIC scores [31] when the sample size is large:

$$\frac{\mathbb{P}(Y|G^{new}, X)}{\mathbb{P}(Y|G^{old}, X)} \approx \exp(-\frac{1}{2}(BIC_{G^{new}} - BIC_{G^{old}})).$$

The BIC score is defined to be $-2\log L + k\log n$, where L is the maximized value of the likelihood for the estimated model, k is the number of free parameters estimated, and n is the sample size.

In step 4, a new weight W_k^{new} can be sampled from a moving uniform distribution $U(W_k^{old} - 1, W_k^{old} + 1)$, and if the sampled W_k^{new} is less than 0, we take a negative of the new weight. This proposal distribution makes the ratio of proposal distributions, $R(W_k^{old}|W_k^{new})/R(W_k^{new}|W_k^{old})$, to be 1. In addition, we need to compute

$$\frac{\mathbb{P}(G_Y|W_k^{new}, W_{-k}^{old}, B)}{\mathbb{P}(G_Y|W^{old}, B)} = \frac{\frac{\exp(-W_k^{new}\mathcal{E}_k(G_Y) - \sum_{k'\neq k} W_{k'}^{old}\mathcal{E}_{k'}(G_Y))}{Z(W_k^{new}, W_{-k}^{old})}}{\frac{\exp(-\sum_k W_k^{old}\mathcal{E}_k(G_Y))}{Z(W^{old})}},$$

where $Z(W) = \sum_{G_Y \in \text{DAG}} \exp(-\sum_k W_k \mathcal{E}_k(G_Y))$ is a normalizing constant. Note that it is not feasible to compute the exact $Z(W)$ due to the exclusion of cyclic networks. We approximate the normalizing constant by the summation over directed graphs with restriction on the number of parents, e.g., 3 as adopted by [55].

After running an MCMC chain, we need to efficiently summarize the chain for the inference of a network structure. The choice by the highest posterior network structure might not produce a convincing model because the graph space grows rapidly with the number of phenotype nodes and the most probable network structure might still have a very low probability. Therefore, instead of selecting the network structure with the highest posterior probability, we perform Bayesian model averaging [8] over the causal links between phenotypes, to infer an averaged network. Explicitly, let Δ_{uv} represent a causal link from u to v, that is, $\Delta_{uv} = \{Y_u \rightarrow Y_v\}$. Then

$$\begin{aligned}\mathbb{P}(\Delta_{uv} \mid Y, X) &= \sum_G \mathbb{P}(\Delta_{uv} \mid G, Y, X)\, \mathbb{P}(G \mid Y, X) \\ &= \sum_G \mathbb{P}\{\Delta_{uv} \in G\}\, \mathbb{P}(G \mid Y, X).\end{aligned}$$

The averaged network is represented by the causal links with maximum posterior probability or with posterior probability above a predetermined threshold, e.g., 0.5.

7.3.3 Summary of Encoding of Biological Knowledge

In equation (7.5), we have constructed a prior distribution on a network structure G_Y in terms of energy $\mathcal{E}(G_Y)$ relative to biological knowledge B. Now we describe how to encode a biological knowledge matrix B from several sources of biological information. Recall that B is an encoding meant to describe biological knowledge ranging from 0 to 1, and energy $\mathcal{E}(G_Y)$ is defined to be a distance measure between B and G_Y in equation (7.4). When there is no available biological knowledge, we would put every element in B as 1/2. Then all DAGs have the same energy, and therefore the probability of a network structure conditional on W is $1/K$, with K as the number of all DAGs. In contrast, when biological knowledge is available, we will look at several ways of encoding biological knowledge into B, such as transcription factor and DNA binding [4], protein–protein interaction [28], and gene ontology annotations [36].

TRANSCRIPTION FACTOR AND DNA BINDING

Chromatin immunoprecipitation with microarray experiments is used to investigate the interaction of proteins and DNA in vivo. This technology has been employed to generate putative lists of transcription factor/target gene interactions [33]. In [4], an approach was suggested to convert a p-value P_{ij}, quantifying the evidence that a transcription factor i binds to a putative target

gene j, into a posterior probability for the presence and orientation of an edge in a Bayesian network. Following [4], we assume that the p-value P_{ij} follows a truncated exponential distribution with mean λ when the transcription factor i binds to a target gene j ($G_Y(i,j) = 1$) and a uniform distribution when the transcription factor does not bind to a target gene ($G_Y(i,j) = 0$).

$$\mathbb{P}_\lambda(P_{ij} = p|G_Y(i,j) = 1) = \frac{\lambda e^{-\lambda p}}{1 - e^{-\lambda}},$$
$$\mathbb{P}_\lambda(P_{ij} = p|G_Y(i,j) = 0) = 1.$$

The probability of the directed edge before observing any biological data is assumed to be $\mathbb{P}(G_Y(i,j) = 1) = 1/2$ so that without any biological data, the probability of presence of edge only depends on the expression data. By Bayes' rule, the probability of presence of an edge after observing a p-value is:

$$\mathbb{P}_\lambda(G_Y(i,j) = 1|P_{ij} = p) = \frac{\lambda e^{-\lambda p}}{\lambda e^{-\lambda p} + (1 - e^{-\lambda})}.$$

Here λ is assumed to be uniformly distributed over the interval $[\lambda_L, \lambda_H]$ and the integration over λ is performed to obtain the probability of the presence of an edge:

$$\mathbb{P}(G_Y(i,j) = 1|P_{ij} = p) = \frac{1}{\lambda_H - \lambda_L} \int_{\lambda_L}^{\lambda_H} \frac{\lambda e^{-\lambda p}}{\lambda e^{-\lambda p} + (1 - e^{-\lambda})}\, d\lambda.$$

This can be solved numerically, for instance, by choosing λ in the range $[0, 10000]$. We should thus obtain the following estimate: $B(i,j) = \mathbb{P}(G_Y(i,j) = 1|P_{ij} = p)$.

PROTEIN—PROTEIN INTERACTION

Since protein–protein interaction is non-directional, we put the same probability on both directions. If we do not consider the diverse reliabilities of protein–protein interaction from several experiments, we set $B(i,j)$ and $B(j,i)$ to be $\delta > 1/2$ when we find any interaction on any experiment. If there are gold standards for positive and negative protein–protein interactions, and experiments have diverse reliabilities, then we can use the Bayes classifier proposed by [28] to combine heterogeneous data. Positive gold standards are well-known true protein–protein interactions, whereas negative gold standards are interactions which cannot happen, such as those between a pair of proteins in different subcellular compartments. An interaction experimental data set is a collection of observations over all pairs of proteins by binaries concerning whether the interaction is present or absent for each pair. Suppose there are L interaction experimental data sets with different false positive rates. We can calculate the posterior odds of an interaction from the binary observations $f_1, \ldots, f_L$ using the likelihood ratio LR:

$$O_{posterior} = \frac{\mathbb{P}(pos|f_1, \ldots, f_L)}{\mathbb{P}(neg|f_1, \ldots, f_L)} = O_{prior} \times LR$$
$$= \frac{\mathbb{P}(pos)}{\mathbb{P}(neg)} \times \frac{\mathbb{P}(f_1, \ldots, f_L|pos)}{\mathbb{P}(f_1, \ldots, f_L|neg)}.$$

In the positive gold standard interactions, we can find a set of interactions which have the observed values $f_1, \ldots, f_L$. The likelihood under the positive gold standard can be defined to be the proportion of the set with the values $f_1, \ldots, f_L$ in the positive gold standard. Similarly we define the likelihood $\mathbb{P}(f_1, \ldots, f_L|neg)$ under the negative gold standard. Then we can take the ratio of the two likelihoods to calculate the likelihood ratio LR. The prior odds O_{prior} can be defined by an expert. The encoding of B can be obtained by transforming the posterior odds into a posterior positive rate:

$$B(i,j) = B(j,i) = \frac{O_{posterior}}{1 + O_{posterior}}.$$

When the posterior odds is equal to 1, $B(i,j)$ and $B(j,i)$ are equal to 1/2. As the posterior odds increases, the values of $B(i,j)$ and $B(j,i)$ also increase.

GENE ONTOLOGY

The Gene Ontology (GO) [1] is a well-controlled vocabulary of terms describing the molecular functions, biological processes, and cellular components of a gene. A GO is structured as a DAG in which each node represents a GO term. The GO terms annotate a large fraction of genes. The distance between two genes can be defined in terms of their GO annotations. One well-defined distance is Lord's similarity [36]. This measure takes into account the hierarchy of GO and GO term occurrences in the myriad of genes. If two genes share a more specific GO term positioned in the lower part of the GO hierarchy, they are more likely to be similar. However, even if the shared GO terms lie in the same level of the hierarchy, the frequencies of the GO terms in the whole genes are different, which affects the similarity. Consider that two GO terms, c_1 and c_2, lie in the same level of the hierarchy. Suppose there are 100 genes annotated with term c_1 and there are 1000 genes annotated with term c_2. Then the chance of two genes sharing the term c_2 is higher than the chance of sharing the term c_1. This then implies that the term c_1 is more informative. The information content $IC(c)$ for a GO term c is defined to be the negative logarithm of the number of times the term or any of its descendant terms occurs in the myriad of genes, divided by the total number of occurrences of GO terms. The root of the hierarchy will have zero information content, whereas a leaf of the hierarchy will have high information content. Once the information content $IC(c)$ for each node in the GO is set up, we can define GO term similarity and gene similarity. The similarity between two GO terms is defined to be the maximum information content among the shared parents of the two terms, which is

$$\text{sim}(c_1, c_2) = \max_{c \in (pa(c_1) \cap pa(c_2))} IC(c).$$

Then, since a gene is annotated with a set of GO terms, the similarity between two genes g_1 and g_2 can be defined as the average of similarities calculated for all pairs of GO terms between two genes; that is,

$$\text{sim}(g_1, g_2) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \text{sim}(c_{1,i}, c_{2,j})}{nm}.$$

This Lord's measure can be used as an encoding of B if it is rescaled to be in the interval $[0, 1]$.

Table 7.3 Four methods for causal phenotype network inference, which differ in their use of genetic variation information and biological knowledge.

Method	Use of Genetic Variation Information	Use of Biological Knowledge
QTLnet-prior	YES	YES
QTLnet	YES	NO
WH-prior	NO	YES
Expression	NO	NO

7.4 Simulations

We performed a simulation study for comparing the proposed method (QTLnet-prior) with three other methods: QTLnet [9], WH-prior [55], and Expression. Table 7.3 provides a summary of these four methods in terms of using the genetic variation information and biological knowledge. QTLnet was implemented using R/QTLnet, QTLnet-prior was implemented with prior setting on R/QTLnet, and WH-prior was programmed as in [55], with a modification of approximating the marginal likelihood with the BIC score instead of using the BGe score [17].[4] Expression was programmed by modifying R/QTLnet to exclude QTL mapping.

We simulated expression data and a priori knowledge matrix according to the network structure in Fig. 7.1 and produced 100 simulated data sets. To generate expression data based on the network in Fig. 7.1, the genetic information was simulated first. The genetic map described five chromosomes of 100 cM with ten equally spaced markers in each chromosome, and the markers were simulated for 500 mice in an F2 population using R/qtl [7]. We assumed QTL Q_t was located in the middle of chromosome t. Then, each expression data set of the F2 population was generated with different genetic effects and partial regression coefficients between phenotypes. Genetic additive effects were sampled from a uniform distribution $U[0, 0.5]$, and dominance effects were sampled from $U[0, 0.25]$. The partial regression coefficients β_{ut} were sampled from $U[-0.5, 0.5]$. The residual phenotypic variance was 1. The biological knowledge matrix B was generated for several cases. The value $B(t, u)$ was generated from one of two $[0, 1]$-truncated normal distributions $\mathcal{N}_{\pm}(0.5 \pm \delta, 0.1)$ [16]. The distribution was truncated at 0 and 1 to guarantee that the value $B(t, u)$ ranged from 0 to 1. When no biological knowledge is available, the natural choice for $B(t, u)$ is 1/2. Consequently, the evidence for the presence (absence) of edge $t \to u$ is necessarily specified through a value of $B(t, u)$ greater (lower) than 1/2. In the simulation, the $B(t, u)$ value of true edge was generated from $\mathcal{N}_+$ and the $B(t, u)$ value of false edge was generated from $\mathcal{N}_-$.

[4] BGe stands for *B*ayesian metric for *G*aussian networks having score *e*quivalence. The BGe score was developed as a scoring metric for a Bayesian network of continuous variables under the assumption that the data are sampled from a multivariate Gaussian distribution. The BGe score is first derived for a complete Bayesian network where every pair of distinct nodes is connected by a direct edge. It assumes a prior on parameter to be a normal Wishart distribution so that one can obtain a closed-form marginal distribution. Under the assumption of parameter independence and modularity, the BGe score for an arbitrary Bayesian network is derived to be $\mathbb{P}(Y|G) = \prod_{t=1}^{T} \frac{\mathbb{P}(Y_t, Y_{pa(t)}|G_c)}{\mathbb{P}(Y_{pa(t)}|G_c)}$, where G_c is any complete DAG such that each node has the same parents as in G. It is known that Markov equivalent DAGs have the same BGe score. See [17] for details.

The parameter δ controls the accuracy of prior knowledge. We denote the generated biological knowledge to be positive knowledge, non-informative knowledge, or negative knowledge based on the sign of δ: +, 0, or –, respectively. We examined 11 cases of different accuracies of prior knowledge: $\delta \in \{\pm 0.1, \pm 0.08, \pm 0.06, \pm 0.04, \pm 0.02, 0\}$. In the extreme case when δ is equal to 0.5, the prior knowledge almost correctly reflects the network structure, whereas when δ is equal to –0.5, the prior knowledge incorrectly reflects the network structure almost in the opposite way. When δ is equal to 0, the information is generated with no distinction between true and false edges. For each simulated data set, we ran an MCMC for 30 300 iterations, discarded the first 300 iterations, sampled every ten iterations, and generated 3000 samples.

We assessed these four methods by using Receiver Operating Characteristic (ROC) curves of the proportion of recovered and spurious edges. Bigger areas under the ROC curve generally indicate better performance, as the area represents the probability that the classifier ranks true edges higher than false edges [13]. The ROC curves are obtained from the set of proportions of recovered edges and spurious edges for various posterior probability thresholds ranging from 0 to 1.

First, we evaluated the effect of incorporating genetic variation information. The effect of QTL mapping can be tested by comparing QTLnet-prior and WH-prior. QTLnet-prior is more effective in recovering the network structure than WH-prior in Fig. 7.4A, and we can conclude that QTL mapping increases the effectiveness. Better causal QTLs can be inferred by conditioning on the phenotype network. Similarly, a better phenotype network can be inferred by conditioning on causal QTLs. The gain is more apparent when the biological knowledge is negative.

Second, we evaluated the effect of incorporating biological knowledge. In Fig. 7.4A, when δ is positive, QTLnet-prior performs better than QTLnet and WH-prior performs better than

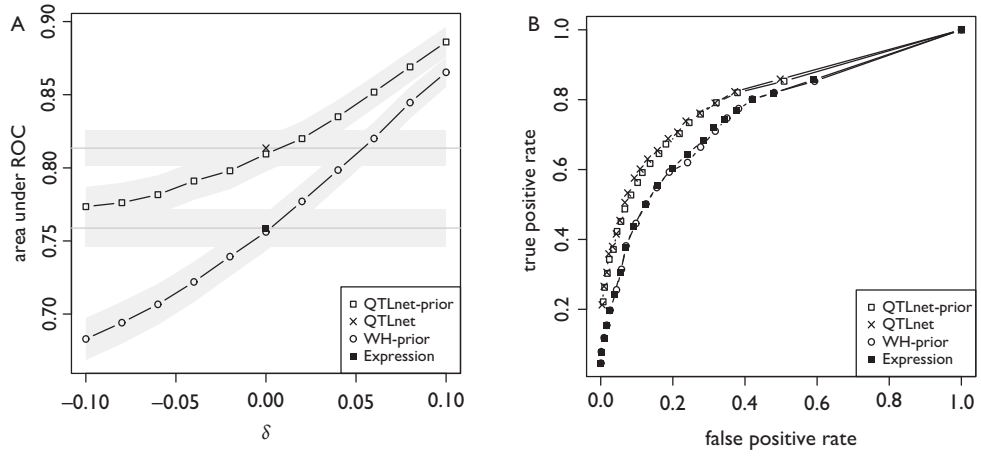

Fig 7.4 A comparison of the four methods of causal phenotype network inference by the area under the Receiver Operating Characteristic (ROC) curves with respect to the accuracy of biological knowledge. (A) The areas under ROC curves of QTLnet-prior, QTLnet, WH-prior, and Expression. The areas under ROC curves of QTLnet-prior and WH-prior are plotted against the accuracy of biological knowledge, δ. Since QTLnet and Expression do not incorporate biological knowledge, they are plotted in a single point each (×, ■). The shaded area indicates the standard error of the area under ROC curve. (B) The ROC curves of QTLnet-prior and WH-prior are drawn when non-informative biological knowledge ($\delta = 0$) is incorporated. They are compared with the ROC curves of QTLnet and Expression which do not incorporate biological information.

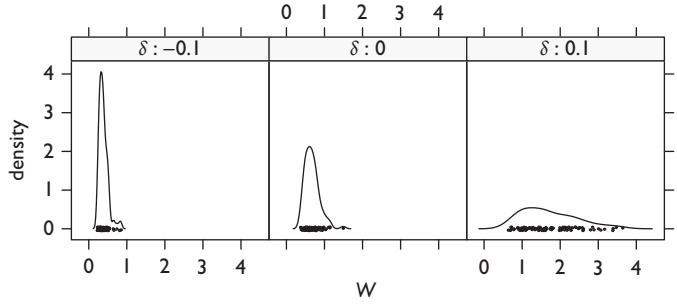

Fig 7.5 The distribution of median weight W of posterior sample by QTLnet-prior inference. Each panel shows the median W distribution when biological knowledge is defective ($\delta = -0.1$), non-informative ($\delta = 0$), and informative ($\delta = 0.1$).

Expression; whereas when δ is negative, QTLnet-prior performs worse than QTLnet and WH-prior performs worse than Expression. With a positive δ, as the accuracy of knowledge increases, QTLnet-prior and WH-prior benefit by prior knowledge incorporation. However, a negative δ, indicating that the knowledge disagrees with the true network structure, causes QTLnet-prior and WH-prior to be impaired by prior knowledge incorporation. The decreased performances in QTLnet-prior and WH-prior draw attention as to whether W can effectively control the influence of negative knowledge. Fig. 7.5 shows that the median of W in the posterior sample is close to 0 with negative knowledge. It implies that the weight W can effectively control the use of negative knowledge to some extent but not completely, based on the decreased recovery observed in comparison to the case of non-informative knowledge evidenced in Fig. 7.4A. In comparison with QTLnet and Expression, the reduced performance of QTLnet-prior and WH-prior can be explained by the remaining uncontrolled effect of prior probability incorporating negative knowledge. When non-informative knowledge is incorporated, there is no significant difference in the area under the ROC curve between QTLnet and QTLnet-prior (p-value = 0.82) and between Expression and WH-prior (p-value = 0.89), as shown in Fig. 7.4A and 7.4B.

7.5 Analysis of Yeast Cell-Cycle Genes

We used QTLnet-prior to reconstruct a network of 26 genes involved in the cell cycle in yeast (*Saccharomyces cerevisiae*), previously chosen by [4] for cell-cycle network analysis with time-dependent expression data and transcription factor binding information. Some genes express periodically according to cell-cycle phase (genome duplication phase, gap phase 2, cell division phase, and gap phase 1), and there are transcription factors that regulate these periodical genes [3]. The gene expression data and genetic variation information for this analysis were obtained from a backcross population of 112 segregates between BY4716 and RM11-1a [6]. In [6], the authors extracted gene expression data by constructing a backcross, isolating the RNA, and hybridizing the resulting complementary DNA (cDNA; see Chapter 1, Section 1.3) to microarrays. They also genotyped the population at 2957 genetic markers for genetic variations. In addition to gene expression data and genetic variations, we incorporated transcription factor

binding information as biological knowledge for QTLnet-prior analysis. The p-value for evidence of transcription factor binding from chromatin immunoprecipitation with microarray experiments is available for 106 transcription factors from [33]. For the 26 genes in our analysis, 11 of them are transcription factors and the rest are known targets of one or more transcription factors. We transformed the p-values into the biological knowledge matrix B, as described in Subsection 7.3.3.

The construction of the causal network focused on the 26 phenotypes of 112 yeast segregates, incorporating genetic variation information at 2957 markers and biological knowledge of transcription factor binding. We ran an MCMC for 760 000 iterations, discarded the first 200 000 iterations, sampled every 100 iterations, and finally got 5 600 samples used for estimation. The computation took around 14 days of CPU time on a 2.66GHz Intel(R) Core(TM)2 Quad running Red Hat 4.1.2-50. To examine the mixing and convergence of the MCMC chain, we first computed the autocorrelation of Bayesian information criterion (BIC) scores and autocorrelation of W, respectively. As shown in Fig. 7.10 in Appendix 7.A, both autocorrelation values get close to 0. It indicates that the MCMC chain may not suffer from a slow mixing rate. Furthermore, we calculated Geweke's convergence diagnostic [18] to check the convergence of the Markov chain. The Geweke's diagnostic is asymptotically $\mathcal{N}(0, 1)$ when it is equal for the two means of the first 10% and the last 50% of the Markov chain. The Geweke's diagnostic for the BIC score is 0.34 and is –0.25 for W, suggesting the convergence of the chain. Fig. 7.6 shows the causal phenotype network reconstructed by QTLnet-prior. The full network of phenotypes and causal QTLs can be found in Fig. 7.9 in Appendix 7.A.

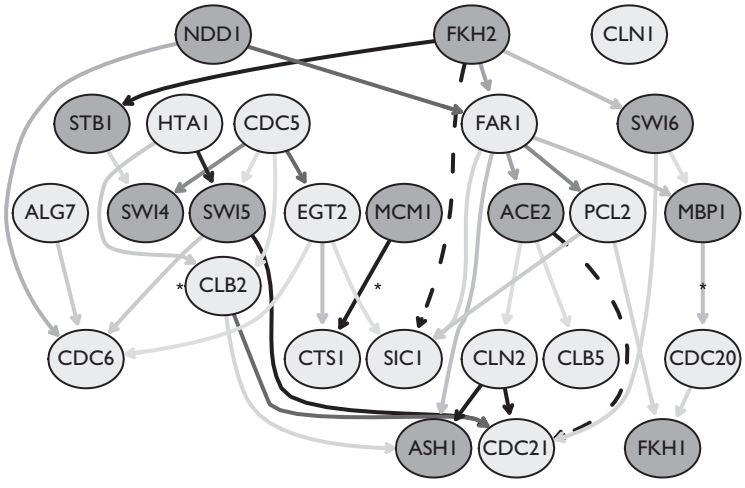

Fig 7.6 Yeast cell-cycle phenotype network by QTLnet-prior, integrating transcription factor binding information. A solid line represents an inferred edge with a posterior probability over 0.5, and the darkness of the edge is in proportion to the posterior probability. Dark nodes are transcription factors. The edge consistent with transcription factor binding information is marked with an asterisk. The transcription factor binding relation recovered by an indirect path in the inferred network is represented by a dashed line.

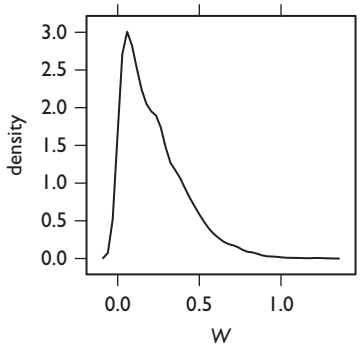

Fig 7.7 The posterior distribution of weight W of transcription factor information in reconstructing a yeast cell-cycle network by QTLnet-prior.

In the transcription factor biological knowledge matrix B, we defined a pair (i.e., an edge from node i to node j) to be *significant* if its $B(i, j)$ value is over 0.5. There are 44 significant transcription factor pairs in B. For the constructed network with 36 inferred edges in Fig. 7.6, we found three significant direct pairs (MBP1 → CDC20, SWI5 → CDC6, and MCM1 → CTS1) and two significant indirect pairs (ACE2 → CDC21 and FKH2 → SIC1). Interestingly, we did not find any reverse relations, that is, causal relations from target genes to transcription factors, in the inferred network. The remaining 39 causal relations in B were not inferred in the phenotype network.

Fig. 7.7 shows the posterior density distribution of the weight W of transcription factor information, which has a mode of approximately 0 with right skewness. To further examine the contribution of transcription factor information on phenotype network reconstruction, we applied QTLnet to construct the phenotype network without using transcription factor information. For the two networks inferred by QTLnet-prior and QTLnet, the posterior probability of every possible directed edge is very similar to each other, as shown in Fig. 7.8. Although the transcription factor knowledge did not improve the reconstruction of the cell-cycle network, it did not have a negative impact on reconstruction, either. The weight parameter was actually effective in protecting the network reconstruction against the inconsistent transcription factor information. The inconsistency between transcription factor information and expression data

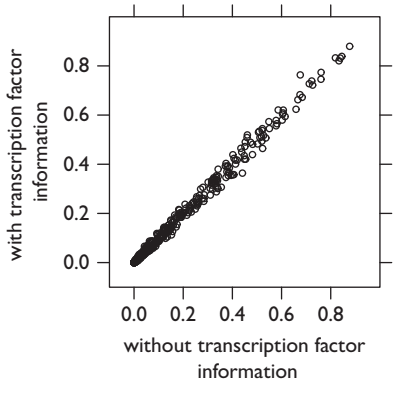

Fig 7.8 Comparison of the posterior probability of every possible directed edge between the network inferred by QTLnet-prior and the network inferred by QTLnet.

may be due to any of the following reasons: (1) inconsistency between physical regulation of transcription binding and transcriptional regulation level of expression changes; (2) necessary post-translational modification of transcription factor or construction of complexes with other proteins for regulation of target genes; (3) cell cycle phase or tissue-dependent transcription factor binding or false binding information; and (4) inability to capture the cyclicity of the cell-cycle network from static expression data relative to a single time point.

7.6 Conclusion

We have developed a phenotype network inference method (QTLnet-prior) to incorporate genetic variation information and biological knowledge. Genotypes are known to control phenotypes but not the other way and thereby can help to distinguish phenotype network structures. Biological knowledge can improve the clustering and directional inference between phenotypes. The simulation study shows that the proposed method can improve the reconstruction of the gene network by integrating genetic variation information and biological knowledge as long as knowledge agrees with data. When biological knowledge does not agree with data, the weight of knowledge controls the contribution of prior probability of biological knowledge on the likelihood of data, reducing (to some extent) the negative impact of the defective knowledge. We applied QTLnet-prior to estimate a yeast cell-cycle network of 26 genes with causal QTLs by integrating transcription factor binding information and compared its performance to QTLnet. The distribution of weight suggests that the transcription factor binding information was inconsistent with the expression data. Nonetheless, comparison with QTLnet's output showed a fairly similar result, suggesting the weight parameter of knowledge was effective in controlling the negative impact of inconsistent knowledge in this case.

When we interpret the inferred networks, we need to be cautious. Even though, in theory, the incorporation of causal QTLs allows us to distinguish network structures that would otherwise be likelihood equivalent, in practice some of the detected expression-to-expression causal relationships might be invalid. The possible explanation is that the inferred expression network represents a projection of real causal relationships that might take place outside the transcriptional regulation level. For instance, the true causal regulations could be due to transcription factor binding, direct protein–protein interaction, phosphorylation, methylation, etc. and might not be well reflected at the gene expression level. The incorporation of diffused biological knowledge, mined from different levels of biological regulation, could potentially improve the reconstruction of gene-expression regulatory networks. In any case, the inference of these networks can still play an important role in generating hypothetically possible causal relations.

There are several factors that could change the inference by QTLnet-prior. One is the prior distribution specification. We have used the Gibbs distribution as a prior distribution for network structures in equation (7.5) in terms of an absolute distance measure in equation (7.4) to incorporate biological knowledge. The exponential distribution is used for the weight of biological knowledge in equation (7.6) with the rate parameter (see Subsection 7.3.1). However, we could consider different choices of network structure distributions, measures to incorporate information, weight distributions, and hyperparameters. Another factor is the sample size of

expression data. As the sample size increases, the contribution of biological knowledge will be generally reduced. This shows the limited contribution of biological knowledge on the reconstruction of networks, even though biological knowledge B can also be obtained from a number of experiments, as discussed in [55]. The third factor is the global control of biological knowledge on network reconstruction. Illustrated by the yeast cell-cycle network, every transcription factor/target regulation was controlled by the same weight parameter. It may have resulted in no contribution of any biological knowledge, even though five transcription factor/target regulations were inferred to be consistent with expression data. This suggests incorporation of biological knowledge by local control parameters when reconstructing a network. Finally, the encoding of biological knowledge plays an important role. We have proposed to use the encoding for a transcription factor and its targets from [4], protein–protein interactions from [28], and gene ontology annotations from [36]. These encodings are mainly about direct relationships in separate biological regulation levels. As discussed in the previous paragraph, this diffused biological knowledge can improve the Bayesian network reconstruction.

There are shortcomings to the QTLnet-prior framework inherited from QTLnet. One of the assumptions of QTLnet is an absence of latent variables. Latent variables can make it impossible to find the marginalized model in the class of DAGs as shown in [46] and can induce erroneous relations. Suppose there are three nodes y_1, y_2, and y_3, and y_1 and y_2 have a common parent c_1 while y_2 and y_3 have a common parent c_2. If the common parents c_1 and c_2 are not observed, we obtain the following independence relations: $y_1 \perp\!\!\!\perp y_3$ and $y_1 \not\perp\!\!\!\perp y_3 \mid y_2$. Then, we mistakenly infer that y_1 and y_3 are parents of y_2. To address this problem, one can consider the more general class of ancestral graphs, which takes care of latent variables. Ancestral graphs open up the possibility of latent variables, although they do not explicitly include the latent variables in the network structures [46].

A persistent challenge in Bayesian network analysis is how to cope with large networks, since the DAG space size grows superexponentially with the number of nodes. Approaches based on Markov blankets with and without restrictions on the number of parent nodes have been proposed [45, 47, 49]. In [27], the authors approximated the Bayesian network problem to a linear programming problem. In [51], the authors developed a parallel algorithm that infers subnetworks restricted on a Markov blanket and merges the subnetworks. Likewise, in phylogeny estimation, the supertree reconstruction from small trees has been studied [5]. We think the rigorous development of super Bayesian network methodology to integrate small subnetworks is a promising direction for the inference of a large network, since the inference of small subnetworks is computationally inexpensive and multiple subnetworks can be parallelized for computation. In this era of vast biological data and knowledge in various aspects, integrating them reasonably on a large scale can be an interesting topic for future research.

ACKNOWLEDGMENTS

The authors wish to thank NIH/NIDDK 58037 (JYM, BSY), and 66369 (JYM, BSY), NIH/NIGMS 74244 (JYM, BSY) and 69430 (JYM, BSY), NCI ICBP U54-CA149237 (ECN), and NIH R01MH090948 (ECN), for supporting this work. Alan D. Attie and Mark Keller motivated the work through our collaboration on causal models for diabetes and obesity.

APPENDIX 7.A INFERRED YEAST CELL CYCLE NETWORK WITH CAUSAL QTLS INTEGRATING TRANSCRIPTION FACTOR INFORMATION BY QTLNET-PRIOR

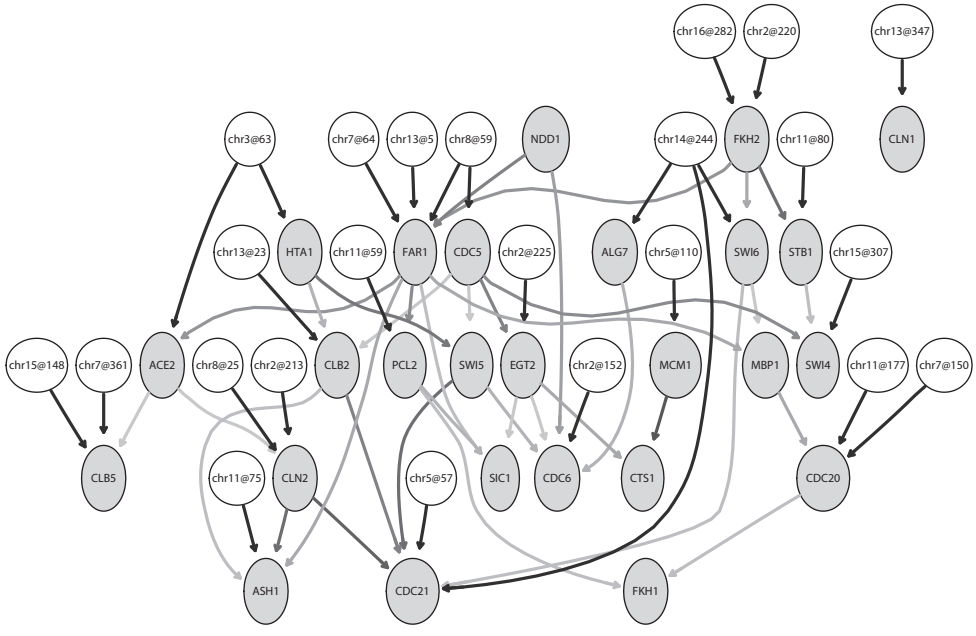

Fig 7.9 Yeast cell-cycle network integrating transcription factor binding information inferred by QTLnet-prior. The edge darkness is in proportion to the posterior probability.

APPENDIX 7.B CONVERGENCE DIAGNOSTICS OF YEAST CELL CYCLE NETWORK

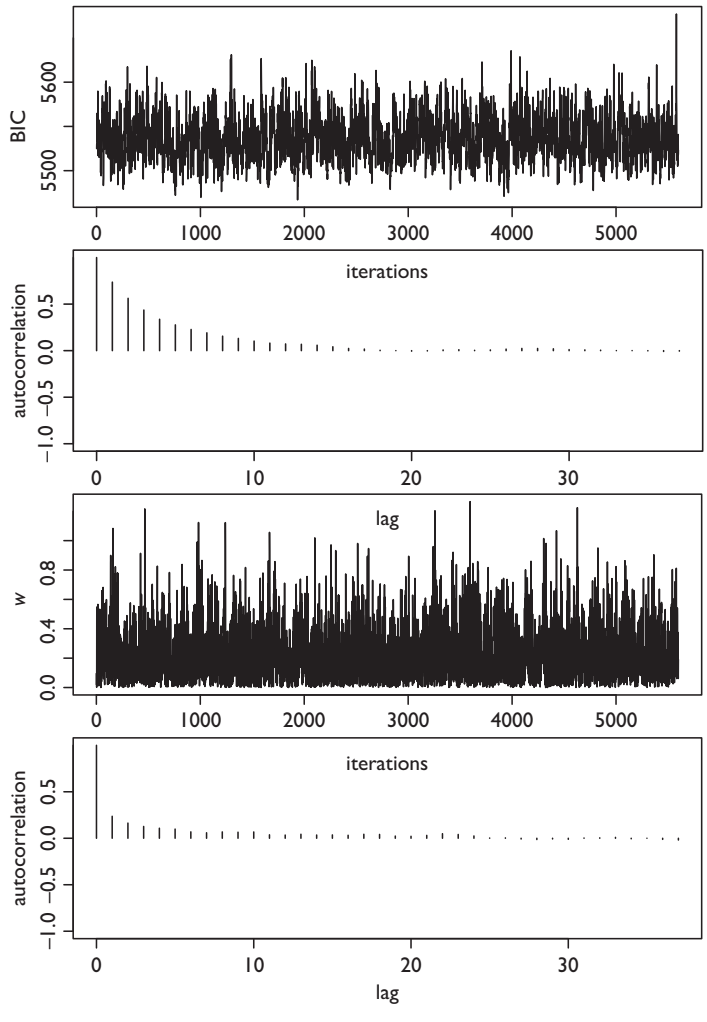

Fig 7.10 The top two figures are the trace plot and the autocorrelation plot of BIC scores for sampled causal networks. The bottom two figures are the trace and the autocorrelation plots of the sampled weights (W) on transcription factor binding information.

REFERENCES

[1] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, and Consortium Gene Ontology. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, 2000.

[2] J. E. Aten, T. F. Fuller, A. J. Lusis, and S. Horvath. Using genetic markers to orient the edges in quantitative trait networks: The NEO software. *BMC Systems Biology*, 2:34, 2008.

[3] J. Bähler. Cell-cycle control of gene expression in budding and fission yeast. *Annual Review of Genetics*, 39(1):69–94, 2005.

[4] A. Bernard and A. J. Hartemink. Informative structure priors: joint learning of dynamic regulatory networks from multiple types of data. *Pacific Symposium on Biocomputing 2005*, pages 459–470, 2005.

[5] O. R. P. Bininda-Emonds, J. L. Gittleman, and M. A. Steel. The (Super)tree of life: procedures, problems, and prospects. *Annual Review of Ecology and Systematics*, 33:265–289, 2002.

[6] R. B. Brem and L. Kruglyak. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 102(5):1572–1577, 2005.

[7] K. W. Broman, H. Wu, S. Sen, and G.A. Churchill. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7):889–890, 2003.

[8] E. Chaibub Neto, C. T. Ferrara, A. D. Attie, and B. S. Yandell. Inferring causal phenotype networks from segregating populations. *Genetics*, 179(2):1089–1100, 2008.

[9] E. Chaibub Neto, M. P. Keller, A. D. Attie, and B. S. Yandell. Causal graphical models in systems genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes. *Annals of Applied Statistics*, 4(1):320–339, 2010.

[10] E. Chaibub Neto, M. P. Keller, A. T. Broman, A. D. Attie, and B. S. Yandell. Causal model selection tests in systems genetics. Technical Report 1157, Department of Statistics, University of Wisconsin-Madison, 2010.

[11] L. S. Chen, F. Emmert-Streib, and J. D. Storey. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biology*, 8(10):R219, 2007.

[12] S. Christley, Q. Nie, and X. Xie. Incorporating existing network information into gene network inference. *PLOS ONE*, 4(8):e6799, 2009.

[13] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[14] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3–4):601–620, 2000.

[15] T. S. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301(5629):102–105, 2003.

[16] F. Geier, J. Timmer, and C. Fleck. Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge. *BMC Systems Biology*, 1:11, 2007.

[17] D. Geiger and D. Heckerman. Learning Gaussian networks. Technical Report MSR-TR-94-10, Microsoft Research, 1994.

[18] J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. Smith, editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, 1992.

[19] M. Grzegorczyk and D. Husmeier. Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. *Machine Learning*, 71:265–305, 2008.

[20] R. S. Hageman, M. S. Leduc, R. Korstanje, B. Paigen, and G. A. Churchill. A Bayesian framework for inference of the genotype-phenotype map for segregating populations. *Genetics*, 187:1163–1170, 2011.

[21] D. Heckerman and D. Geiger. Likelihoods and parameter priors for Bayesian networks. Technical Report MSR-TR-95-94, Microsoft Research, 1996.
[22] D. Heckerman, C. Meek, and G. Cooper. A Bayesian approach to causal discovery. In D. Holmes and L. Jain, editors, *Innovations in Machine Learning*, volume 194 of *Studies in Fuzziness and Soft Computing*, pages 1–28. Springer Berlin/Heidelberg, 2006.
[23] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: a tutorial. *Statistical Science*, 14:382–417, 1999.
[24] D. Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19(17):2271–2282, 2003.
[25] S. Imoto, T. Higuchi, T. Goto, K. Tashiro, S. Kuhara, and S. Miyano. Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks. *Journal of Bioinformatics and Computational Biology*, 2(1):77–98, 2004.
[26] S. Imoto, S. Kim, T. Goto, S. Miyano, S. Aburatani, K. Tashiro, and S. Kuhara. Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *Journal of Bioinformatics and Computational Biology*, 1(2):231–52, 2003.
[27] T. Jaakkola, D. Sontag, A. Globerson, and M. Meila. Learning Bayesian network structure using LP relaxations. *Journal of Machine Learning Research*, 9:358–365, 2010.
[28] R. Jansen, H. Yu, D. Greenbaum, Y. Kluger, N. J. Krogan, S. Chung, A. Emili, M. Snyder, J. F. Greenblatt, and M. Gerstein. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, 302(5644):449–453, 2003.
[29] R. C. Jansen and J.-P. Nap. Genetical genomics: the added value from segregation. *Trends in Genetics*, 17(7):388–391, 2001.
[30] C.-H. Kao and Z.-B. Zeng. Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics*, 160:1243–1261, 2002.
[31] R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.
[32] D. Kulp and M. Jagalur. Causal inference of regulator-target pairs by gene mapping of expression phenotypes. *BMC Genomics*, 7(1):125, 2006.
[33] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298(5594):799–804, 2002.
[34] R. Li, S.-W. Tsaih, K. Shockley, I. M. Stylianou, J. Wergedal, B. Paigen, and G. A. Churchill. Structural model analysis of multiple quantitative traits. *PLOS Genetics*, 2(7):e114, 2006.
[35] B. Liu, A. de la Fuente, and I. Hoeschele. Gene network inference via structural equation modeling in genetical genomics experiments. *Genetics*, 178(3):1763–1776, 2008.
[36] P. W. Lord, R. D. Stevens, A. Brass, and C. A. Goble. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics*, 19(10):1275–1283, 2003.
[37] D. Madigan and J. York. Bayesian graphical models for discrete data. *International Statistical Review*, 63:215–232, 1995.
[38] J. Millstein, B. Zhang, J. Zhu, and E. Schadt. Disentangling molecular relationships with a causal inference test. *BMC Genetics*, 10(1):23, 2009.
[39] J. H. Nadeau and A. M. Dudley. Systems genetics. *Science*, 331(6020):1015–1016, 2011.
[40] N. Nariai, S. Kim, S. Imoto, and S. Miyano. Using protein-protein interactions for refining gene networks estimated from microarray data by Bayesian networks. *Pacific Symposium on Biocomputing*, 336–47, 2004.
[41] O. Ourfali, T. Shlomi, T. Ideker, E. Ruppin, and R. Sharan. SPINE: a framework for signaling-regulatory pathway inference from cause-effect experiments. *Bioinformatics*, 23(13):i359–i366, 2007.

[42] J. Pearl. *Probabilistic Reasoning in Intelligent Systems:Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.
[43] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.
[44] T. Peleg, N. Yosef, E. Ruppin, and R. Sharan. Network-free inference of knockout effects in yeast. *PLOS Computational Biology*, 6(1):e1000635, 2010.
[45] E. Perrier, S. Imoto, and S. Miyano. Finding optimal Bayesian network given a super-structure. *Journal of Machine Learning Research*, 9:2251–2286, 2008.
[46] T. Richardson and P. Spirtes. Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.
[47] C. Riggelsen. MCMC learning of Bayesian network models by Markov blanket decomposition. In J. Gama, R. Camacho, P. Brazdil, A. Jorge, and L. Torgo, editors, *Machine Learning: ECML 2005*, volume 3720 of *Lecture Notes in Computer Science*, pages 329–340. Springer Berlin/Heidelberg, 2005.
[48] E. E. Schadt, J. Lamb, X. Yang, J. Zhu, S. Edwards, D. Guhathakurta, S. K. Sieberts, S. Monks, M. Reitman, C. S. Zhang, P. Y. Lum, A. Leonardson, R. Thieringer, J. M. Metzger, L. M. Yang, J. Castle, H. Y. Zhu, S. F. Kash, T. A. Drake, A. Sachs, and A. J. Lusis. An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics*, 37(7):710–717, 2005.
[49] M. Schmidt, A. Niculescu-Mizil, and K. Murphy. Learning graphical model structure using L1-regularization paths. In *Proceedings of the 22nd National Conference on Artificial intelligence—Volume 2*, pages 1278–1283. AAAI Press, 2007.
[50] P. Spirtes, C. Glymour, and R. Scheines. *Causation, prediction, and search*. The MIT Press, 2nd edition, 2000.
[51] Y. Tamada, S. Imoto, and S. Miyano. Parallel algorithm for learning optimal Bayesian network structure. *Journal of Machine Learning Research*, 12:2437–2459, 2011.
[52] Y. Tamada, S. Kim, H. Bannai, S. Imoto, K. Tashiro, S. Kuhara, and S. Miyano. Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection. *Bioinformatics*, 19(suppl 2):ii227–ii236, 2003.
[53] B. D. Valente, G. J. M. Rosa, G. de los Campos, D. Gianola, and M. A. Silva. Searching for recursive causal structures in multivariate quantitative genetics mixed models. *Genetics*, 185:633–644, 2010.
[54] T. Verma and J. Pearl. Equivalence and synthesis of causal models. In G. Shafer and J. Pearl, editors, *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence (UAI90)*, pages 220–227. Morgan Kaufmann Publishers, 1990.
[55] A. V. Werhli and D. Husmeier. Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of prior knowledge. *Statistical Applications in Genetics and Molecular Biology*, 6:15, 2007.
[56] C. J. Winrow, D. L. Williams, A. Kasarskis, J. Millstein, A. D. Laposky, H. S. Yang, K. Mrazek, L. Zhou, J. R. Owens, D. Radzicki, F. Preuss, E. E. Schadt, K. Shimomura, M. H. Vitaterna, C. Zhang, K. S. Koblan, J. J. Renger, and F. W. Turek. Uncovering the genetic landscape for multiple sleep-wake traits. *PLOS ONE*, 4(4):e5161, 2009.
[57] C. H. Yeang, T. Ideker, and T. Jaakkola. Physical network models. *Journal of Computational Biology*, 11(2–3):243–262, 2004.
[58] N. J. Yi, D. Shriner, S. Banerjee, T. Mehta, D. Pomp, and B. S. Yandell. An efficient Bayesian model selection approach for interacting quantitative trait loci models with many effects. *Genetics*, 176:1865–1877, 2007.
[59] N. J. Yi, B. S. Yandell, G. A. Churchill, D. B. Allison, E. J. Eisen, and D. Pomp. Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics*, 170(3):1333–1344, 2005.
[60] J. Zhu, P. Y. Lum, J. Lamb, D. Guhathakurta, S. W. Edwards, R. Thieringer, J. P. Berger, M. S. Wu, J. Thompson, A. B. Sachs, and E. E. Schadt. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenetic Genome Research*, 105(2–4):363–374, 2004.

[61] J. Zhu, M. C. Wiener, C. Zhang, A. Fridman, E. Minch, P. Y. Lum, J. R. Sachs, and E. E. Schadt. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLOS Computational Biology*, 3(4):e69, 2007.

[62] J. Zhu, B. Zhang, E. N. Smith, B. Drees, R. B. Brem, L. Kruglyak, R. E. Bumgarner, and E. E. Schadt. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics*, 40(7):854–861, 2008.