Gene expression

Inference of differential gene regulatory networks based on gene expression and genetic perturbation data

Xin Zhou and Xiaodong Cai*

Department of Electrical and Computer Engineering, University of Miami, FL 33146, USA

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on September 3, 2018; revised on June 9, 2019; editorial decision on June 21, 2019; accepted on June 28, 2019

Abstract

Motivation: Gene regulatory networks (GRNs) of the same organism can be different under different conditions, although the overall network structure may be similar. Understanding the difference in GRNs under different conditions is important to understand condition-specific gene regulation. When gene expression and other relevant data under two different conditions are available, they can be used by an existing network inference algorithm to estimate two GRNs separately, and then to identify the difference between the two GRNs. However, such an approach does not exploit the similarity in two GRNs, and may sacrifice inference accuracy.

Results: In this paper, we model GRNs with the structural equation model (SEM) that can integrate gene expression and genetic perturbation data, and develop an algorithm named fused sparse SEM (FSSEM), to jointly infer GRNs under two conditions, and then to identify difference of the two GRNs. Computer simulations demonstrate that the FSSEM algorithm outperforms the approaches that estimate two GRNs separately. Analysis of a dataset of lung cancer and another dataset of gastric cancer with FSSEM inferred differential GRNs in cancer versus normal tissues, whose genes with largest network degrees have been reported to be implicated in tumorigenesis. The FSSEM algorithm provides a valuable tool for joint inference of two GRNs and identification of the differential GRN under two conditions.

Availability and implementation: The R package fssemR implementing the FSSEM algorithm is available at <https://github.com/lvis4ml/fssemR.git>. It is also available on CRAN.

Contact: x.cai@miami.edu

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

A gene regulatory network (GRN) consists of a set of genes that interact with each other to govern their expression and molecular functions. For example, transcription factors (TFs) can bind to promoter regions of their target genes and regulate the expression of target genes (Harbison *et al.*, 2004). Gene-gene interactions can change under different environments, in different tissue types or disease states, and during development and speciation (Ideker and Krogan, 2012). Therefore, GRNs undergo substantial rewiring depending on specific molecular context in which they operate

(Califano, 2011). Identification of condition-specific GRNs is critical to unravel the molecular mechanism of various tissue or disease-specific biological processes (Sonawane *et al.*, 2017).

Although a number of computational methods have been developed to infer GRNs from gene expression and other relevant data, they are mainly concerned with the static structure of gene networks under one single condition. Several methods aim to infer GRNs using only gene expression data; they include the approaches constructing relevance networks based on a similarity measure, such as correlation or mutual information (Butte and Kohane, 1999; Faith

et al., 2007; Margolin *et al.*, 2006), Gaussian Graphical Model (GGM) (Friedman *et al.*, 2008), Bayesian networks (Statnikov and Aliferis, 2010) and linear regression model (Haury *et al.*, 2012). Several other methods infer GRNs by integrating genetic perturbations with gene expression data; these methods include approaches using Bayesian networks incorporating expression quantitative trait loci (eQTLs) (Zhu *et al.*, 2007), likelihood-based causal models (Neto *et al.*, 2008) and structural equation models (SEMs) (Cai *et al.*, 2013; Liu *et al.*, 2008; Logsdon and Mezey, 2010).

While it is possible to apply these methods to identify GRNs under different conditions separately, such an approach is apparently not optimal to identify the difference in GRNs, because it does not exploit the similarity in two GRNs. Several methods have been proposed to use the gene expression data of different conditions to jointly estimate GRNs under different conditions. Particularly, GRNs under multiple conditions are modeled with multiple GGMs, and these GGMs are inferred jointly from gene expression data (Danaher *et al.*, 2014). When a gene is mutated, its regulatory effect on all its target genes may be changed. Taking into account such effects, a node-based approach to joint inference of multiple GGMs was developed in Mohan *et al.* (2014). GGMs exploit the sample covariance of the gene expression levels, but they cannot integrate genetic perturbations with gene expression data. Moreover, it has been demonstrated that genetic perturbation along with gene expression data can determine directed edges in GRNs (Logsdon and Mezey, 2010), but GGMs can only identify undirected edges.

In this paper, we employ SEMs to model GRNs as described in Cai *et al.* (2013), Logsdon and Mezey (2010) and Liu *et al.* (2008). This enables us to integrate genetic perturbation data with gene expression data. Taking into account the sparsity in GRNs, we have developed a sparse-aware maximum likelihood (SML) method (Cai *et al.*, 2013) to infer a single GRN based on SEM. Here, taking into account not only the sparsity in GRNs but also the sparsity in the differences between GRNs under two different conditions, we develop an algorithm, named fused sparse SEM (FSSEM), to infer two GRNs from different conditions jointly, and then to identify difference in two GRNs. Computer simulations demonstrate the superior performance of our novel approach relative to the existing ones that infer GRNs under two conditions separately.

2 Materials and methods

2.1 GRN model

Supposed that expression levels of n genes under two different conditions are measured using, e.g. micro-array or RNA-Seq technique. Let $\mathbf{y}_i^{(k)} = [y_{i1}^{(k)}, y_{i2}^{(k)}, \dots, y_{in}^{(k)}]^T$ denote expression level of n genes in individual i under condition k , where $k = 1, 2$ and $i = 1, 2, \dots, n_k$, with n_k being the number of individuals where gene expression levels are measured under condition k . Supposed that a set of naturally occurring genetic perturbations have been observed for these genes. These perturbations can be due to, e.g. different genotypes of eQTLs or gene copy number variants (CNVs). In this paper, we will consider only eQTLs. As in Cai *et al.* (2013) and Logsdon and Mezey (2010), we assume that each gene in the GRN of interest has at least one *cis*-eQTL, so that the structure of the underlying GRN is uniquely identifiable. Let $\mathbf{x}_i^{(k)} = [x_{i1}^{(k)}, x_{i2}^{(k)}, \dots, x_{iq}^{(k)}]^T$ denote the observed genotypes of q *cis*-eQTLs in individual i under condition k , where $k = 1, 2$ and $i = 1, 2, \dots, n_k$. Since the expression level of a particular gene may be regulated by other genes and is affected by its eQTLs, we employ the following SEM to model the expression of n genes:

$$\mathbf{y}_i^{(k)} = \mathbf{B}^{(k)} \mathbf{y}_i^{(k)} + \mathbf{F}^{(k)} \mathbf{x}_i^{(k)} + \boldsymbol{\mu}^{(k)} + \boldsymbol{\epsilon}_i^{(k)}, \quad (1)$$

where $i = 1, \dots, n_k$, $k = 1, 2$, $n \times n$ matrix $\mathbf{B}^{(k)}$ defines the unknown network structure under condition k , $n \times q$ matrix $\mathbf{F}^{(k)}$ captures the effect of *cis*-eQTLs on gene expression levels under condition k , $n \times 1$ vector $\boldsymbol{\mu}^{(k)}$ accounts for the model bias in SEM, and $n \times 1$ vector $\boldsymbol{\epsilon}_i^{(k)}$ denotes the residual error, which is modeled as a Gaussian vector with zero mean and variance σ^2 and is independent for different samples. It is assumed that no self-loops are presented per gene in GRNs, which implies that the diagonal entries of $\mathbf{B}^{(k)}$ are zero. It is also assumed that q *cis*-eQTLs have been identified using an existing eQTL method, but sizes of their effects are unknown, thus, $\mathbf{F}^{(k)}$ has q nonzero entries with known locations. Our main goal is to estimate network matrices $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$, but $\mathbf{F}^{(1)}$ and $\mathbf{F}^{(2)}$ will also be estimated jointly with $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$.

2.2 Joint inference of two GRNs

Let $\mathbf{Y}^{(k)} = [\mathbf{y}_1^{(k)}, \dots, \mathbf{y}_{n_k}^{(k)}]$, $\mathbf{X}^{(k)} = [\mathbf{x}_1^{(k)}, \dots, \mathbf{x}_{n_k}^{(k)}]$ and $\mathbf{E}^{(k)} = [\boldsymbol{\epsilon}_1^{(k)}, \dots, \boldsymbol{\epsilon}_{n_k}^{(k)}]$, where $k = 1, 2$, and assume that $n_1 + n_2$ observations are independent. Then, the negative log-likelihood function of the data can be written as

$$\begin{aligned} \mathbb{L}(\mathbf{B}, \mathbf{F}, \boldsymbol{\mu}, \sigma^2) &= -\log \prod_{k=1}^2 \prod_{i=1}^{n_k} \mathbb{P}(\mathbf{y}_i^{(k)} | \mathbf{x}_i^{(k)}, \boldsymbol{\mu}^{(k)}, \mathbf{B}^{(k)}, \mathbf{F}^{(k)}) \\ &= -\sum_{k=1}^2 \frac{n_k}{2} \log |\mathbf{I} - \mathbf{B}^{(k)}|^2 + \frac{(n_1 + n_2)n}{2} \log(2\pi\sigma^2) \\ &\quad + \frac{1}{2\sigma^2} \sum_{k=1}^2 \|(\mathbf{I} - \mathbf{B}^{(k)})\mathbf{Y}^{(k)} - \mathbf{F}^{(k)}\mathbf{X}^{(k)} - \boldsymbol{\mu}^{(k)}\|_F^2, \end{aligned} \quad (2)$$

where $\mathbf{B} = [\mathbf{B}^{(1)}, \mathbf{B}^{(2)}]$, $\mathbf{F} = [\mathbf{F}^{(1)}, \mathbf{F}^{(2)}]$, $\boldsymbol{\mu} = [\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}]$, and $\|\cdot\|_F$ stands for the Frobenius norm. It is not difficult to show that minimizing (2) with respect to $\boldsymbol{\mu}$ yields $\hat{\boldsymbol{\mu}}^{(k)} = (\mathbf{I} - \mathbf{B}^{(k)})\tilde{\mathbf{Y}}^{(k)} - \mathbf{F}^{(k)}\tilde{\mathbf{X}}^{(k)}$, where $\tilde{\mathbf{Y}}^{(k)} = \mathbf{Y}^{(k)} - 1/n_k \sum_{i=1}^{n_k} \mathbf{y}_i^{(k)} \mathbf{1}$, $\tilde{\mathbf{X}}^{(k)} = \mathbf{X}^{(k)} - 1/n_k \sum_{i=1}^{n_k} \mathbf{x}_i^{(k)} \mathbf{1}$, and $\mathbf{1}$ is a vector with all its entries equal to 1.

Since a gene is regulated by a small number of other genes (Gardner *et al.*, 2003; Tegner *et al.*, 2003; Thieffry *et al.*, 1998), GRNs are sparse, meaning that most entries of $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ are zeros. Moreover, it is reasonable to expect that changes in a GRN under two different conditions are relatively small. Therefore, most entries of $\mathbf{B}^{(2)} - \mathbf{B}^{(1)}$ are zeros. Let $\hat{\sigma}^2$ be an estimate of σ^2 that will be specified later, replacing $\boldsymbol{\mu}^{(k)}$ and σ^2 in (2) with $\hat{\boldsymbol{\mu}}^{(k)}$ and $\hat{\sigma}^2$, respectively, and taking into account the sparsity in $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$, and the sparsity in $\mathbf{B}^{(2)} - \mathbf{B}^{(1)}$, we estimate $\mathbf{B}$ and $\mathbf{F}$ by minimizing the following penalized negative log-likelihood function:

$$\begin{aligned} J(\mathbf{B}, \mathbf{F}) &= -\sum_{k=1}^2 n_k \log |\mathbf{I} - \mathbf{B}^{(k)}| \\ &\quad + \frac{1}{2\hat{\sigma}^2} \sum_{k=1}^2 \|(\mathbf{I} - \mathbf{B}^{(k)})\tilde{\mathbf{Y}}^{(k)} - \mathbf{F}^{(k)}\tilde{\mathbf{X}}^{(k)}\|_F^2 \\ &\quad + \lambda \sum_{k=1}^2 \|\mathbf{B}^{(k)}\|_{1,w^{(k)}} + \rho \|\mathbf{B}^{(2)} - \mathbf{B}^{(1)}\|_{1,r}, \end{aligned} \quad (3)$$

where $\mathbf{B}_{i,i}^{(k)} = 0$, $\forall i = 1, \dots, n$, $k = 1, 2$, $\|\mathbf{B}^{(k)}\|_{1,w^{(k)}} = \sum_i \sum_j w_{ij}^{(k)} |\mathbf{B}_{ij}^{(k)}|$ is the weighted ℓ_1 -norm, $\|\mathbf{B}^{(2)} - \mathbf{B}^{(1)}\|_{1,r}$ is also a weighted ℓ_1 -norm with similar definition, λ and ρ are two nonnegative parameters. Note that the number of unknowns in $\mathbf{B}^{(k)}$ and $\mathbf{F}^{(k)}$, $k = 1, 2$ is $p = 2n(n-1) + 2q$, which is typically much larger than the number

of samples $n_1 + n_2$. Due to the ℓ_1 -norm terms in (3), our FSSEM algorithm that minimizes $J(\mathbf{B}, \mathbf{F})$ can handle the case where $p \gg n_1 + n_2$, as will be shown in computer simulations and real data analysis. Weights $w_{ij}^{(k)}$ and r_{ij} in the penalty terms are introduced to improve estimation accuracy and robustness in line with the adaptive lasso (Zou, 2006) and the adaptive generalized fused lasso (Viallon *et al.*, 2016), and they are selected as $1/|\hat{\mathbf{B}}_{ij}^{(k)}|$ and $1/|\hat{\mathbf{B}}_{ij}^{(2)} - \hat{\mathbf{B}}_{ij}^{(1)}|$, respectively, where $\hat{\mathbf{B}}^{(1)}$ and $\hat{\mathbf{B}}^{(2)}$ are preliminary estimates of $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ obtained from the following ridge regression:

$$\begin{aligned} \{\hat{\mathbf{B}}, \hat{\mathbf{F}}\} = & \underset{\{\mathbf{B}, \mathbf{F}\}}{\operatorname{argmin}} \left\{ \sum_{k=1}^2 \frac{1}{2} \|(\mathbf{I} - \mathbf{B}^{(k)})\tilde{\mathbf{Y}}^{(k)} - \mathbf{F}^{(k)}\tilde{\mathbf{X}}^{(k)}\|_F^2 + \tau \|\mathbf{B}^{(k)}\|_F^2 \right\} \\ & \text{s.t. } \mathbf{B}_{ii}^{(k)} = 0, \quad \forall i = 1, \dots, n, \quad k = 1, 2, \end{aligned} \quad (4)$$

where $\hat{\mathbf{B}} = [\hat{\mathbf{B}}^{(1)}, \hat{\mathbf{B}}^{(2)}]$, $\hat{\mathbf{F}} = [\hat{\mathbf{F}}^{(1)}, \hat{\mathbf{F}}^{(2)}]$, and the estimate of σ^2 , $\hat{\sigma}^2$ in (3) is given by

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^2 \|(\mathbf{I} - \hat{\mathbf{B}}^{(k)})\tilde{\mathbf{Y}}^{(k)} - \hat{\mathbf{F}}^{(k)}\tilde{\mathbf{X}}^{(k)}\|_F^2}{(n_1 + n_2)n}. \quad (5)$$

Of note, the likelihood function (2) is obtained under the assumption that $\epsilon_i^{(1)}$ and $\epsilon_i^{(2)}$ are independent, or equivalently, that $\mathbf{y}_i^{(1)}$ and $\mathbf{y}_i^{(2)}$ are independent. In some applications, $\mathbf{y}_i^{(1)}$ and $\mathbf{y}_i^{(2)}$ may be correlated, when, for example, $\mathbf{y}_i^{(1)}$ is the gene expression of a tumor, and $\mathbf{y}_i^{(2)}$ is the gene expression level of the normal tissue of the same individual. It turns out that our FSSEM algorithm, that minimizes the penalized negative likelihood function $J(\mathbf{B}, \mathbf{F})$ in (3), is quite robust with respect to (w.r.t.) the correlation between $\epsilon_i^{(1)}$ and $\epsilon_i^{(2)}$. Our computer simulations will show that performance of the FSSEM algorithm does not change significantly, when the correlation coefficient between $\epsilon_i^{(1)}$ and $\epsilon_i^{(2)}$ varies from 0 to 1.

2.3 Ridge regression

In the first stage, we solve the ridge regression problem (4) to find initial values of $\mathbf{B}$, $\mathbf{F}$, weights $w_{ij}^{(k)}$, $k = 1, 2$, and r for the FSSEM algorithm to minimize (3). Let $\mathbf{B}_i^{(k)}$, $\mathbf{F}_i^{(k)}$ and $\mathbf{Y}_i^{(k)}$ be the i -th row of $\mathbf{B}^{(k)}$, $\mathbf{F}^{(k)}$ and $\mathbf{Y}^{(k)}$, respectively. Define $\mathbf{B}_{i,-i}^{(k)}$ as the $1 \times (n-1)$ vector obtained by removing the i -th entry from $\mathbf{B}_i^{(k)}$. Let $S_q(i)$ be the set of indices of non-zero entries in the $\mathbf{F}_i^{(k)}$, $\mathbf{F}_{i,S_q(i)}^{(k)}$ be the vector that contains the nonzero entries of $\mathbf{F}_i^{(k)}$, $\tilde{\mathbf{X}}_{S_q(i)}$ be the matrix formed by taking rows of $\tilde{\mathbf{X}}$ whose indices are in $S_q(i)$, and $\tilde{\mathbf{Y}}_{-i}$ be the matrix formed by removing the i th row of $\tilde{\mathbf{Y}}$.

Then, the ridge regression problem (4) can be decomposed into n separate problems

$$\underset{\mathbf{B}_{i,-i}, \mathbf{F}_{i,S_q(i)}}{\operatorname{argmin}} \left\{ \sum_{k=1}^2 \frac{1}{2} \|\tilde{\mathbf{Y}}_i^{(k)} - \mathbf{B}_{i,-i}^{(k)}\tilde{\mathbf{Y}}_{-i}^{(k)} - \mathbf{F}_{i,S_q(i)}^{(k)}\tilde{\mathbf{X}}_{S_q(i)}^{(k)}\|_F^2 + \tau \|\mathbf{B}_{i,-i}^{(k)}\|_F^2 \right\}, i = 1, \dots, n. \quad (6)$$

Minimizing the objective function in (6) w.r.t. $\mathbf{F}_{i,S_q(i)}^{(k)}$ yields the following closed-form solution

$$\hat{\mathbf{F}}_{i,S_q(i)}^{(k)} = (\tilde{\mathbf{Y}}_i^{(k)} - \hat{\mathbf{B}}_{i,-i}^{(k)}\tilde{\mathbf{Y}}_{-i}^{(k)})\tilde{\mathbf{X}}_{S_q(i)}^{(k)}(\tilde{\mathbf{X}}_{S_q(i)}^{(k)}\tilde{\mathbf{X}}_{S_q(i)}^{(k)T})^{-1}. \quad (7)$$

Substituting $\hat{\mathbf{F}}_{i,S_q(i)}^{(k)}$ into (6) and minimizing w.r.t. $\hat{\mathbf{B}}_{i,-i}^{(k)}$ gives $\hat{\mathbf{B}}_{i,-i}^{(k)} = \tilde{\mathbf{Y}}_i^{(k)}\mathbf{P}_i^{(k)}\tilde{\mathbf{Y}}_{-i}^{(k)T}(\tilde{\mathbf{Y}}_{-i}^{(k)}\mathbf{P}_i^{(k)}\tilde{\mathbf{Y}}_{-i}^{(k)T} + \tau\mathbf{I})^{-1}$, which in

turn results in $\hat{\mathbf{F}}_{i,S_q(i)}^{(k)} = \tilde{\mathbf{Y}}_i^{(k)}\mathbf{P}_i^{(k)}\tilde{\mathbf{X}}_{S_q(i)}^{(k)T}(\tilde{\mathbf{X}}_{S_q(i)}^{(k)}\tilde{\mathbf{X}}_{S_q(i)}^{(k)T})^{-1}$, where $\mathbf{P}_i^{(k)} = \mathbf{I} - \mathbf{P}_i^{(k)}\tilde{\mathbf{Y}}_{-i}^{(k)T}(\tilde{\mathbf{Y}}_{-i}^{(k)}\mathbf{P}_i^{(k)}\tilde{\mathbf{Y}}_{-i}^{(k)T} + \tau\mathbf{I})^{-1}\tilde{\mathbf{Y}}_{-i}^{(k)}$ and $\mathbf{P}_i^{(k)} = \mathbf{I} - \tilde{\mathbf{X}}_{S_q(i)}^{(k)T}(\tilde{\mathbf{X}}_{S_q(i)}^{(k)}\tilde{\mathbf{X}}_{S_q(i)}^{(k)T})^{-1}\tilde{\mathbf{X}}_{S_q(i)}^{(k)}$. After $\hat{\mathbf{B}}^{(k)}$ and $\hat{\mathbf{F}}^{(k)}$ are estimated, the estimate of $\hat{\sigma}^2$ is given in (5). The hyper-parameter τ in ridge regression (4) or (6) is selected by 5-fold cross-validation.

2.4 FSSEM algorithm

In this section, we will develop the FSSEM algorithm to minimize the objective function $J(\mathbf{B}, \mathbf{F})$ in (3) with the initial values of $\mathbf{B}^{(k)}$ and $\mathbf{F}^{(k)}$ given in (7). The objective function is non-convex due to the log-determinant term, and non-smooth due to the ℓ_1 norm terms. Recently, the proximal alternating linearized minimization (PALM) method (Bolte *et al.*, 2014) was developed to solve a broad classes of non-convex and non-smooth minimization problems. We next apply the PALM approach to develop the FSSEM algorithm.

Without loss of generality, we define the proximal operator associated with a proper and lower semi-continuous function $h(\mathbf{x}) : \mathbb{R}^d \rightarrow (-\infty, +\infty]$ as $\operatorname{prox}_{\alpha}^h(\mathbf{v}) = \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^d} \{\alpha/2 \|\mathbf{u} - \mathbf{v}\|^2 + h(\mathbf{u})\}$, where $\alpha > 0$ and $\mathbf{v} \in \mathbb{R}^d$ are given. We also define the fused lasso signal approximator (Friedman *et al.*, 2007; Hoeffling, 2010) on $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$ as the following proximal operator:

$$\operatorname{prox}_{\alpha}^{f(\mathbf{x})}(\mathbf{z}) = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^{2d}} \left\{ \frac{\alpha}{2} \sum_{k=1}^2 \|\mathbf{x}_k - \mathbf{z}_k\|^2 + f(\mathbf{x}) \right\}, \quad (8)$$

where $f(\mathbf{x}) = \lambda \sum_{k=1}^2 \|\mathbf{x}_k\|_1 + \rho \|\mathbf{x}_2 - \mathbf{x}_1\|_1$, $\mathbf{z}_1$ and $\mathbf{z}_2$ are two $d \times 1$ vectors and $\mathbf{z} = [\mathbf{z}_1^T, \mathbf{z}_2^T]^T$. Denote solution of (8) as $(\mathbf{x}_1(\lambda), \mathbf{x}_2(\lambda))$, and let $\mathbf{x}_{1j}(\lambda)$ and $\mathbf{x}_{2j}(\lambda)$ be the j th element of $\mathbf{x}_1(\lambda)$ and $\mathbf{x}_2(\lambda)$, respectively. Then we have

$$(\mathbf{x}_{1j}(0), \mathbf{x}_{2j}(0)) = \begin{cases} (\mathbf{z}_{1j} - \rho/\alpha, \mathbf{z}_{2j} + \rho/\alpha), & \mathbf{z}_{1j} - \mathbf{z}_{2j} > 2\rho/\alpha \\ (\mathbf{z}_{1j} + \rho/\alpha, \mathbf{z}_{2j} - \rho/\alpha), & \mathbf{z}_{1j} - \mathbf{z}_{2j} < -2\rho/\alpha \\ \left(\frac{\mathbf{z}_{1j} + \mathbf{z}_{2j}}{2}, \frac{\mathbf{z}_{1j} + \mathbf{z}_{2j}}{2} \right), & |\mathbf{z}_{1j} - \mathbf{z}_{2j}| \leq 2\rho/\alpha \end{cases}, \quad (9)$$

where $\mathbf{z}_{1j}$ and $\mathbf{z}_{2j}$ are the j th element of $\mathbf{z}_1$ and $\mathbf{z}_2$, respectively.

Define soft-thresholding function $S(\beta, \lambda)$ as

$$S(\beta, \lambda) = \begin{cases} \beta - \lambda & \text{if } \beta > \lambda \\ \beta + \lambda & \text{if } \beta < -\lambda \\ 0 & \text{if } |\beta| \leq \lambda \end{cases}. \quad (10)$$

If $\boldsymbol{\beta} = [\beta_1, \dots, \beta_d]$ is a $d \times 1$ vector, the $S(\boldsymbol{\beta}, \lambda)$ is a $d \times 1$ vector whose j th element is $S(\beta_j, \lambda)$. The solution of (8) at $\lambda > 0$ is given in terms of the soft-thresholding operator as follows (Friedman *et al.*, 2007):

$$\operatorname{prox}_{\alpha}^{f(\mathbf{x})}(\mathbf{z}) = (S(\mathbf{x}_1(0), \lambda/\alpha), S(\mathbf{x}_2(0), \lambda/\alpha)). \quad (11)$$

Minimizing (3) w.r.t. $\mathbf{F}^{(k)}$ yields $\hat{\mathbf{F}}_{i,S_q(i)}^{(k)}$ in (7). Substituting $\hat{\mathbf{F}}_{i,S_q(i)}^{(k)}$ in (7) into (3) gives

$$J(\mathbf{B}) = H(\mathbf{B}) + \sum_{i=1}^{N_g} f_i(\mathbf{B}_{i,-i}), \quad (12)$$

where

$$H(\mathbf{B}) = -\sum_{k=1}^2 \frac{n_k}{2} \log |\mathbf{I} - \mathbf{B}^{(k)}|^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{N_g} \sum_{k=1}^2 \|\tilde{\mathbf{Y}}_i^{(k)} \mathbf{P}_i^{(k)} - \mathbf{B}_{i,-i}^{(k)} \tilde{\mathbf{Y}}_{-i}^{(k)} \mathbf{P}_i^{(k)}\|_2^2, \quad (13)$$

and

$$f_i(\mathbf{B}_{i,-i}) = \lambda (\|\mathbf{B}_{i,-i}^{(1)}\|_{1,w^{(1)}} + \|\mathbf{B}_{i,-i}^{(2)}\|_{1,w^{(2)}}) + \rho \|\mathbf{B}_{i,-i}^{(1)} - \mathbf{B}_{i,-i}^{(2)}\|_{1,r}. \quad (14)$$

Using the inertial version of the PALM approach (Pock and Sabach, 2016), the FSSEM algorithm efficiently minimizes the non-convex non-smooth function $J(\mathbf{B})$ with the block coordinate descent (BCD) method in an iterative fashion. More specifically, in each cycle of the iteration, $J(\mathbf{B})$ is minimized successively w.r.t. $[\mathbf{B}_{i,-i}^{(1)}, \mathbf{B}_{i,-i}^{(2)}]$, while $[\mathbf{B}_{i,-j}^{(1)}, \mathbf{B}_{i,-j}^{(2)}]$, $j = 1, \dots, n, j \neq i$ are fixed.

Let us consider updating the i th block of variables $\mathbf{B}_{i,-i} = [\mathbf{B}_{i,-i}^{(1)}, \mathbf{B}_{i,-i}^{(2)}]$ in the $(t+1)$ th cycle. Let $\mathbf{B}[t] = [\mathbf{B}^{(1)}[t], \mathbf{B}^{(2)}[t]]$ be the estimate of $\mathbf{B}$ in the t th cycle. Define $\tilde{\mathbf{B}}_{i,-j} = \mathbf{B}_{i,-j}[t_1] + \alpha_t(\mathbf{B}_{i,-j}[t_1] - \mathbf{B}_{i,-j}[t_1 - 1])$, where $t_1 = t + 1, \forall j < i, t_1 = t, \forall j \geq i$, and α_t is a constant in the interval $[0, 1]$. We obtain $\mathbf{B}_{i,-i}$ from the FLSA proximal operator (11) as follows:

$$\mathbf{B}_{i,-i} = \text{prox}_{\gamma_i}^{f_i(\cdot)}(\tilde{\mathbf{B}}_{i,-i} - \frac{1}{\gamma_i} \nabla_{\mathbf{B}_{i,-i}} H(\tilde{\mathbf{B}})), \quad (15)$$

where $1/\gamma_i$ is the step-size for the i -th block that will be given later, and $\nabla_{\mathbf{B}_{i,-i}} H(\tilde{\mathbf{B}})$ is the partial derivative of $H(\mathbf{B})$ w.r.t. $\mathbf{B}_{i,-i}$ at $\tilde{\mathbf{B}}$.

Since $\mathbf{B}_{i,-i} = [\mathbf{B}_{i,-i}^{(1)}, \mathbf{B}_{i,-i}^{(2)}]$, we have $\nabla_{\mathbf{B}_{i,-i}} H(\mathbf{B}) = [\nabla_{\mathbf{B}_{i,-i}^{(1)}} H(\mathbf{B}), \nabla_{\mathbf{B}_{i,-i}^{(2)}} H(\mathbf{B})]$. The determinant of $\mathbf{I} - \mathbf{B}^{(k)}$ can be expressed as $\mathbf{c}_{ii}^{(k)} - \mathbf{B}_{i,-i}^{(k)} \mathbf{c}_i^{(k)}$, where $\mathbf{c}_{ii}^{(k)}$ is the (i, i) co-factor of $\mathbf{I} - \mathbf{B}^{(k)}$, and the j th entry of the $(n-1) \times 1$ column vector $\mathbf{c}_i^{(k)}$ is the co-factor of $\mathbf{I} - \mathbf{B}^{(k)}$ corresponding to the j th entry of $\mathbf{B}_{i,-i}^{(k)}$. Defining $\mathbf{B}_{-i}^{(k)} = \{\mathbf{B}_{j,-i}^{(k)}, j = 1, \dots, n, j \neq i\}$, we can write $\nabla_{\mathbf{B}_{i,-i}^{(k)}} H(\mathbf{B}), k = 1, 2$, with $\mathbf{B}_{-i}^{(k)}$ fixed, as follows:

$$\nabla_{\mathbf{B}_{i,-i}^{(k)}} H(\mathbf{B}) = \frac{n_k \mathbf{c}_i^{(k)T}}{\mathbf{c}_{ii}^{(k)} - \mathbf{B}_{i,-i}^{(k)} \mathbf{c}_i^{(k)}} + \frac{1}{\sigma^2} (\mathbf{B}_{i,-i}^{(k)} \tilde{\mathbf{Y}}_{-i}^{(k)} \mathbf{P}_i^{(k)} \tilde{\mathbf{Y}}_{-i}^{(k)T} - \tilde{\mathbf{Y}}_i^{(k)} \mathbf{P}_i^{(k)} \tilde{\mathbf{Y}}_{-i}^{(k)T}), \quad (16)$$

where $\mathbf{c}_i^{(k)}$ is calculated from $\mathbf{B}_{-i}^{(k)}$. In [Supplementary Text S1](#), we prove that given $\mathbf{B}_{-i}^{(k)}, \nabla_{\mathbf{B}_{i,-i}^{(k)}} H(\mathbf{B}), k = 1, 2$ are Lipschitz continuous. Specifically, we can write $\nabla_{\mathbf{B}_{i,-i}^{(k)}} H(\mathbf{B})$ as $\nabla_{\mathbf{B}_{i,-i}^{(k)}} H(\mathbf{B}_{i,-i}^{(k)}, \mathbf{B}_{-i}^{(k)})$, which satisfies

$$\|\nabla_{\mathbf{B}_{i,-i}^{(k)}} H(\mathbf{x}, \mathbf{B}_{-i}^{(k)}) - \nabla_{\mathbf{B}_{i,-i}^{(k)}} H(\mathbf{y}, \mathbf{B}_{-i}^{(k)})\| \leq L_i(\mathbf{B}_{-i}^{(k)}) \|\mathbf{x} - \mathbf{y}\|, \quad (17)$$

where the Lipschitz constant $L_i(\mathbf{B}_{-i}^{(k)})$ is derived in the [Supplementary Text S1](#), and is given by

$$L_i(\mathbf{B}_{-i}^{(k)}) = n_k \|\mathbf{c}_i^{(k)}\|_2^2 / \min_{\mathbf{B}_{i,-i}^{(k)}} (\det(\mathbf{I} - \mathbf{B}^{(k)}))^2 + \lambda_{\max}(\tilde{\mathbf{Y}}_{-i}^{(k)} \mathbf{P}_i^{(k)} \tilde{\mathbf{Y}}_{-i}^{(k)T}) / \sigma^2. \quad (18)$$

Algorithm 1 Fused Sparse SEM (FSSEM)

```

Select  $\tau^*$  in (4) via cross-validation
Solve (4) with  $\tau^*$  to obtain  $(\tilde{\mathbf{B}}, \tilde{\mathbf{F}})$ , and compute  $\hat{\sigma}^2$  from (5).
Set  $w_{ij}^{(k)} = 1/|\tilde{\mathbf{B}}_{ij}^{(k)}|$ ,  $r_{ij} = 1/|\tilde{\mathbf{B}}_{ij}^{(2)} - \tilde{\mathbf{B}}_{ij}^{(1)}|$ .
Initialize  $\mathbf{B}[0] = \tilde{\mathbf{B}}$ .
for  $t$  in  $1, 2, \dots$  do
  Select  $\alpha_t \in [0, 1]$ 
  for  $i$  in  $1, \dots, n$  do
    Compute  $L_i(\mathbf{B}_{-i})$  from (18), set  $\gamma_i = L_i(\mathbf{B}_{-i})$ 
    Update  $\mathbf{B}_{i,-i}^{(k)}, k = 1, 2$ , with (15)
    Set  $\tilde{\mathbf{B}}_{i,-i} = \mathbf{B}_{i,-i}[t] + \alpha_t(\mathbf{B}_{i,-i}[t] - \mathbf{B}_{i,-i}[t-1])$ 
  end for
  Update  $\mathbf{F}_i^{(k)}$  with (7) and  $\hat{\sigma}^2$  with (5)
  if convergence then
    Break
  end if
end for
Return  $\{\hat{\mathbf{B}}^{(k)}, \hat{\mathbf{F}}^{(k)}, k = 1, 2\}$ 

```

Here $\lambda_{\max}(\tilde{\mathbf{Y}}_{-i}^{(k)} \mathbf{P}_i^{(k)} \tilde{\mathbf{Y}}_{-i}^{(k)T})$ is the maximum eigenvalue of $\tilde{\mathbf{Y}}_{-i}^{(k)} \mathbf{P}_i^{(k)} \tilde{\mathbf{Y}}_{-i}^{(k)T}$, and the value of $\min_{\mathbf{B}_{i,-i}^{(k)}} \det(\mathbf{I} - \mathbf{B}^{(k)})^2$ can be computed by solving the optimization problem as shown in (S12) in [Supplementary Text S1](#). Let $L_i(\mathbf{B}_{-i}) = \max\{L_i(\mathbf{B}_{-i}^{(k)}), k = 1, 2\}$. Then, the step size in (15) is chosen to be $1/\gamma_i = 1/L_i(\mathbf{B}_{-i})$.

The FSSEM algorithm is summarized in Algorithm 1. The convergence criterion is defined as

$$\left\{ \sum_{k=1}^2 \|\mathbf{B}^{(k)}[t+1] - \mathbf{B}^{(k)}[t]\|_F^2 / \sum_{k=1}^2 \|\mathbf{B}^{(k)}[t]\|_F^2 + \sum_{k=1}^2 \|\mathbf{F}^{(k)}[t+1] - \mathbf{F}^{(k)}[t]\|_F^2 / \sum_{k=1}^2 \|\mathbf{F}^{(k)}[t]\|_F^2 \right\} < \epsilon_v \quad (19)$$

$$|J(\mathbf{B}[t+1]) - J(\mathbf{B}[t])| / |J(\mathbf{B}[t])| < \epsilon_o,$$

where $\epsilon_v > 0$ and $\epsilon_o > 0$ are pre-specified small constants. Since the objective function is not convex, it is not guaranteed that the FSSEM algorithm converges to the global minimization. However, we prove in [Supplementary Text S1](#) that the FSSEM algorithm always converges to a stationary point of the objective function. Note that if we drop the fused lasso term $\rho \|\mathbf{B}^{(1)} - \mathbf{B}^{(2)}\|_{1,r}$ in (3), then minimizing $J(\mathbf{B}, \mathbf{F})$ is equivalent to estimating two network matrices $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ separately. The BCD approach used in FSSEM can also be employed to solve this problem, because the proximal operator in (15) can be easily solved in terms of the soft-thresholding function $S(\beta, \lambda)$ defined in (10). This BCD approach is much more efficient than the SML algorithm in [Cai et al. \(2013\)](#), which employs the element-wise coordinate ascent approach. Parameters λ and ρ in (3) can be determined with cross-validation (CV) or Bayesian information criterion (BIC). In [Supplementary Text S1](#), we derive the expression for the maximum values of λ and ρ and describe the CV process.

Fig. 1. The PD and FDR of FSSEM, SML and QDG algorithms for the DAG with $n = 300$ genes and $n_e = 3$ eQTLs per gene. The number of samples $n_s = n_1 = n_2$ varies from 80 to 500 and noise variance $\sigma^2 = 0.25$. PD and FDR were obtained from 30 network replicates

3 Results

3.1 Computer simulations

In this section, we conduct simulation studies to compare the performance of the FSSEM algorithm with that of the SML (Cai *et al.*, 2013) and the QDG (Neto *et al.*, 2008) algorithms. FSSEM estimates network matrices $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ jointly, while SML and QDG estimate $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ separately. Other method such as the AL-based (Logsdon and Mezey, 2010) algorithm is also available to estimate $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ separately. However, as shown in Cai *et al.* (2013), SML outperforms the AL-based algorithm. Therefore, we select only SML and QDG for performance comparison.

Following the setup of Cai *et al.* (2013), both directed acyclic networks (DAG) and directed cyclic networks (DCG) are simulated in our experiments. Specifically, the adjacency matrix $\mathbf{A}^{(1)}$ of a DAG or DCG of 30 or 300 gene nodes with expected number of edges per gene $d = 1$ or $d = 0.1$ is generated for the GRN under condition 1. Another adjacency matrix $\mathbf{A}^{(2)}$ was generated by randomly changing 10% entries of $\mathbf{A}^{(1)}$, and the probabilities of changes of entries from 0 to 1 and from 1 to 0 are equal. A network matrix $\mathbf{B}^{(1)}$ was generated from $\mathbf{A}^{(1)}$ as follows. For any entry $A_{ij}^{(1)} = 1$, $B_{ij}^{(1)}$ is generated from a random variable uniformly distributed over interval $[0.5, 1]$ or $[-1, -0.5]$; for all $A_{ij}^{(1)} = 0$, we set $B_{ij}^{(1)} = 0$. The second network matrix $\mathbf{B}^{(2)}$ was generated from $\mathbf{A}^{(2)}$ and $\mathbf{B}^{(1)}$ as follows. For all $A_{ij}^{(2)} = 0$, we set $B_{ij}^{(2)} = 0$; for all $A_{ij}^{(2)} = A_{ij}^{(1)}$, we set $B_{ij}^{(2)} = B_{ij}^{(1)}$; and for all $A_{ij}^{(2)} = 0$ but $A_{ij}^{(1)} = 1$, we generate $B_{ij}^{(2)}$ from a random variable uniformly distributed over interval $[0.5, 1]$ or $[-1, -0.5]$. The genotypes of eQTLs were simulated from an F2 cross. Values 1 and 3 were assigned to two homozygous genotypes, respectively, and value 2 to the heterozygous genotype. Then, $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ were generated from ternary random variables taking on values $\{1, 2, 3\}$ with corresponding probabilities $\{0.25, 0.5, 0.25\}$. The number of eQTLs per gene n_e was chosen to be 3, and effect sizes of all eQTLs were set to 1 in $\mathbf{F}^{(1)}$ and $\mathbf{F}^{(2)}$. Error terms $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ were independently sampled from Gaussian random variables with zero mean and variance σ^2 ;

$\mu^{(1)}$ and $\mu^{(2)}$ were set to zero vectors; and the sample sizes $n_s = n_1 = n_2$ vary from 80 to 500. Finally, $\mathbf{Y}^{(k)}$ was calculated as $\mathbf{Y}^{(k)} = (\mathbf{I} - \mathbf{B}^{(k)})^{-1}(\mathbf{F}^{(k)}\mathbf{X}^{(k)} + \mathbf{E}^{(k)})$, where $k = 1, 2$.

For each configuration of the two GRNs, 30 replicates of the GRN were simulated. For each replicate, QDG, SML and FSSEM were run to infer network matrices $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$. QDG was implemented with $\alpha = 0.01$, and the hyper parameters of SML and FSSEM algorithms were selected by 5-fold cross-validation and BIC, respectively. After $\mathbf{B}^{(k)}$, $k = 1, 2$, are estimated, an edge from gene j to gene i under condition k is declared, if $B_{ij}^{(k)} \neq 0$. The power of detection (PD) and the false discovery rate (FDR) for detecting network edges were calculated from $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ estimated from the data of each of 30 network replicates. Detailed definitions of PD and FDR are given in Supplementary Text S1. The differential network was defined as $\Delta\mathbf{B} = \mathbf{B}^{(2)} - \mathbf{B}^{(1)}$, and PD and FDR for the differential network were calculated accordingly.

The results for DAGs with $n = 300$, $n_e = 3$ and $\sigma^2 = 0.25$ are depicted in Figure 1, and results of DAGs under other settings are given in Supplementary Figures S1–S5. First, let us look at the PD and FDR of $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ in the left panel of Figure 1. FSSEM offers slightly better PD than SML, and much better PD than QDG, when the sample size is ≤ 200 , and slightly better PD than QDG, when the sample size is > 200 . It offers much lower FDR than QDG and slightly lower FDR than SML. Next, let us look at the PD and FDR of $\Delta\mathbf{B} = \mathbf{B}^{(2)} - \mathbf{B}^{(1)}$ in the right panel of Figure 1. FSSEM exhibits slightly worse PD than SML, and much better PD than QDG when the sample size is ≤ 200 , and similar or slightly better PD than SML and QDG, when the sample size is > 200 . Moreover, FSSEM offers much smaller FDR than both SML and QDG across all sample sizes. The same trend was also observed in Supplementary Figures S4 and S5 for $n = 300$ with different noise levels for the PD and the FDR of $\mathbf{B}^{(1)}$, $\mathbf{B}^{(2)}$ and $\Delta\mathbf{B}$. Comparing the results in Figure 1 and Supplementary Figures S4 and S5 for DAGs with $n = 300$ with those in Supplementary Figures S1–S3 for DAGs with $n = 30$, we observed that both PD and FDR of FSSEM and SML, as well as the PD of QDG, are similar, and that the FDR of QDG improves when $n = 30$, but it is still higher than the FDR of FSSEM. Overall, our FSSEM offered lower FDR particularly in identifying changed network edges than SML and QDG, while exhibited similar or higher PD. For the GRNs of $n = 300$ genes, there are $2(n^2 - n) = 179400$ unknown entries in $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ to be estimated, and $2n_e = 1800$ unknown entries in $\mathbf{F}^{(1)}$ and $\mathbf{F}^{(2)}$. Therefore, the number of observations is much smaller than the size of features, when the sample size $n_s \leq 500$. Interestingly, the performance of FSSEM and SML algorithms did not change much when the sample size n_s varied from 80 to 500, but the performance of QDG algorithm improves when the sample size increases.

Simulation results for DCGs with $n = 300$, $n_e = 3$ and $\sigma^2 = 0.25$ are depicted in Figure 2, and results of DCGs under other settings are shown in Supplementary Figures S6–S10. The performance of the three algorithms for DCGs is similar to that in the cases of DAGs. To evaluate the performance of these algorithms in inferring relatively large GRNs, we increased n to 600. For both DAGs and DCGs, the performance of FSSEM is similar to the cases when $n = 300$, as shown in Supplementary Figure S11. However QDG and SML were too time consuming and failed to obtain any results for $n = 600$.

For the convenience of comparison, the simulation results of DAG and DCG with $n_1 = n_2 = 500$, $n_e = 3$ and $\sigma^2 = 0.25$ are summarized in Table 1, which clearly shows that FSSEM outperforms both SML and QDG. Particularly, FSSEM offers much lower FDR than SML and QDG in inferring the differential GRN. In all

simulations shown in Figures 1 and 2 and Supplementary Figures S1–S11, and in Table 1, we assumed that data samples are independent. We also tested the performance of the FSSEM algorithm for the case of paired data samples, where the paired samples $y_i^{(1)}$ and $y_i^{(2)}$ are correlated. To this end, we jointly generated the (i, j) th elements of $E^{(1)}$ and $E^{(2)}$ from two Gaussian random variables with zero mean, variance σ^2 and correlation coefficient ρ . The simulation results in Supplementary Figures S12 and S13 for DAG and DCG networks of 30 genes show that the performance of FSSEM for estimating the differential GRN (ΔB) remains almost the same, when ρ varies from 0 to 1. The FDR for estimating B degrades slightly, when ρ varies from 0 to 1. This shows that FSSEM is robust w.r.t the correlation between paired data samples.

3.2 Real data analysis

In Lu *et al.* (2011), gene expression levels in 42 tumors and their adjacent normal tissues of non-smoking female patients with lung adenocarcinoma were measured with 54 675 probe sets from Affymetrix Human Genome U133 Plus 2.0 arrays. The genotypes of single nucleotide polymorphisms (SNPs) in the same set of tissues were obtained using 906 551 SNP probes from Affymetrix Genome-

Wide Human SNP 6.0 arrays. We applied FSSEM to this dataset to infer GRNs in lung cancer and normal tissues.

Both gene expression and SNP data in the gene expression omnibus (GEO) database (GSE33356) were downloaded. The R package *affy* (Gautier *et al.*, 2004) was employed to transform raw microarray data to normalized gene expression levels. Specifically, the raw gene expression data in the custom CDF format (Dai *et al.*, 2005) were normalized using the robust multi-array average (RMA) method (Irizarry *et al.*, 2003a). In total, gene expression levels of 18 807 genes with their Entrez IDs were obtained from 54 675 probe sets. The genotypes of the 906 551 SNP probes in the 84 tissue samples were transformed to values {0, 1, 2} using the following mapping: AA $\rightarrow$ 0, AB $\rightarrow$ 1 and BB $\rightarrow$ 2. The missing genotypes of SNP probes were imputed by Beagle (Browning and Browning, 2007) and SNPs with a minor allele frequency (MAF) of 5% or less were removed (Altshuler *et al.*, 2005). Finally, R package MatrixEQTL (Shabalin, 2012) was adopted to identify *cis*-eQTLs of genes. In total, 1260 genes were found to have at least one *cis*-eQTLs within 1M base pairs (bps) from the open reading frame (ORF) of the gene at an FDR < 0.05 and MAF > 5%.

We applied the FSSEM algorithm to the expression levels and genotypes of the eQTLs of these 1,260 genes to infer the GRNs in lung tumor and normal tissues. An edge from gene i to gene j was detected if $B_{ji}^{(k)} \neq 0, k = 1, 2$, where $B^{(1)}$ and $B^{(2)}$ specify the GRNs in normal and tumor tissues, respectively. Then, we identified the differential GRN based on $\Delta B = B^{(2)} - B^{(1)}$. Since small changes of coefficients B_{ji} may not have much biological effect, we regarded the regulatory effect from gene i to gene j to be different using the following two criteria rather than the simple criterion $B_{ji}^{(2)} \neq B_{ji}^{(1)}$. The first criterion is $|B_{ji}^{(2)} - B_{ji}^{(1)}| > \min\{|B_{ji}^{(1)}|, |B_{ji}^{(2)}|\}$, which ensures that there is at least one-fold change relative to $\min\{|B_{ji}^{(1)}|, |B_{ji}^{(2)}|\}$. However, when one of $B_{ji}^{(k)}, k = 1, 2$ is zero or near zero, this criterion still failed to filter out very small changes. To avoid this issue, we added another criterion. Specifically, we obtained all nonzero entries of $B^{(k)}, k = 1, 2$, and compute the 20 percentile value of all nonzero $|B_{ji}^{(k)}|, k = 1, 2$ as η . Then, we defined the second criterion as $\max\{|B_{ji}^{(k)}|, k = 1, 2\} > \eta$.

Finally, our network analysis with FSSEM identified 325 genes that are associated with at least one edge in the differential network. The differential network has 765 edges and is depicted in Figure 3. To assess whether the 325 genes are related to the cancer status, we performed gene set enrichment analysis (GSEA) with the 4762 C2 gene sets in the molecular signatures database (MSigDB) (Subramanian *et al.*, 2005). Using Fisher's exact test, we found that 25 C2 gene sets are enriched in the set of 325 genes at a

Fig. 2. The PD and FDR of FSSEM, SML and QDG algorithms for the DCG with $n = 300$ genes and $n_e = 3$ eQTLs per gene. The number of samples $n_s = n_1 = n_2$ varies from 80 to 500 and noise variance $\sigma^2 = 0.25$. PD and FDR were obtained from 30 network replicates

Table 1. The PD and FDR of FSSEM, SML and QDG algorithms

Network	n	FSSEM				SML				QDG			
		PD_B	FDR_B	$PD_{\Delta B}$	$FDR_{\Delta B}$	PD_B	FDR_B	$PD_{\Delta B}$	$FDR_{\Delta B}$	PD_B	FDR_B	$PD_{\Delta B}$	$FDR_{\Delta B}$
DAG	30	1.000	0.012	1.000	0.029	0.948	0.077	1.000	0.751	0.974	0.037	0.969	0.245
	300	1.000	0.008	1.000	0.040	0.940	0.159	1.000	0.751	0.996	0.885	0.996	0.980
DCG	30	0.962	0.021	0.981	0.040	0.904	0.077	0.977	0.735	0.911	0.052	0.936	0.301
	300	0.997	0.010	0.991	0.031	0.937	0.170	0.991	0.762	0.993	0.882	0.986	0.980

Note: Expected number of eQTLs per gene is $n_e = 3$, number of samples is $n_1 = n_2 = 500$, noise variance $\sigma^2 = 0.25$. PD and FDR were obtained from 30 network replicates.

Fig. 3. The differential GRN of 325 genes with 765 edges in lung tumors versus normal tissues. The size of a node is proportional to its degree, the number of edges that the node connects to. The top ten genes with the highest degrees are labeled

p -value < 0.05 . These 25 gene sets are listed in [Supplementary Table S1](#). After searching over the C2 gene sets with key words ‘lung cancer’, ‘lung tumor’, ‘lung carcinoma’ and ‘LUCA’ in their description text, we identified 140 gene sets that are related to lung cancer. Comparing the 140 lung cancer-related gene sets with the 25 enriched gene sets, we found that 5 of the 25 enriched gene sets are related to lung cancer, which is significant (Fisher’s exact test, p -value $< 7.2 \times 10^{-4}$). Furthermore, we ranked the 325 genes according to the number of edges in the differential GRN that they are involved in. The top ten genes are C4BPA, ANPEP, LTF, SELE, HLA-DQA1, CLC, PNMAL1, TPSAB1, ERAP2 and PPP1R14C; all ten top genes have been reported in the literature to be implicated in lung cancer, as discussed in [Supplementary Text S1](#).

In [Holbrook et al. \(2011\)](#), gene expression levels and SNPs of 49 gastric tumors and normal tissues were measured with Illumina mRNA expression arrays and Affymetrix SNP arrays, respectively. We downloaded the dataset that contains Illumina sequencing data and SNP data from the GEO database (GSE29999). A total of 558 genes were found to have at least one *cis*-eQTLs within 1 M base pairs from the ORF of the gene at an FDR < 0.01 and MAF $> 5\%$. Genotypes of the eQTLs and expression levels of these 558 genes were analyzed with FSSEM, which yielded a differential GRN of 88 genes in gastric cancer versus normal tissues. Two MSigDB C2 gene sets were found to be enriched in the set of 88 genes, as described in [Supplementary Table S2](#). One of the enriched gene sets is related to gastric cancer. The 88 genes were ranked according to the number of edges that each genes associated with, and the top ten genes are CLC, NLRP2, SNCAIP, IL33, WARS, PGGHG, CXCL13, CIDEA, TMEM45B and RASEF. Six of the top ten genes are related to gastric cancer, and other 2 of top 10 genes are related to other cancers; see [Supplementary Text S1](#) for more detailed description. Therefore, analysis of both lung and gastric datasets with FSSEM algorithm demonstrates that the genes in the differential GRN identified by FSSEM are relevant to tumorigenesis.

4 Discussion

In this paper, we developed a very efficient algorithm, named FSSEM, for joint inference of two similar GRNs by integrating genetic perturbations with gene expression data under two different conditions with the SEM. Computer simulations showed that our FSSEM offered much better accuracy in identifying changed gene-

gene interactions than both SML and QDG algorithms, which infer two GRN separately. Particularly, the FDR of gene interactions in the differential GRN estimated by FSSEM was significantly lower than that resulted from SML and QDG. This result is expected because FSSEM exploits the similarity in the two GRNs and penalizes the changes of gene interactions in the inference process. We also analyzed real datasets of lung and gastric cancers. The FSSEM algorithm identified a set of genes involved in the differential GRNs in cancers, and these genes have been reported to be relevant in tumorigenesis.

The number of unknowns in network matrices $\mathbf{B}^{(k)}$, $k = 1, 2$ is $2n(n-1)$ and the number of unknowns in $\mathbf{F}^{(k)}$, $k = 1, 2$ is $2q$. When n is large, the number of unknown is huge, which incurs huge computation. In the analysis of the lung cancer data, we have $n = 1,260$, and we were able to complete the analysis in several days on a desktop with Intel i7-5820K CPU. If n is larger, analysis using FSSEM may be very time consuming. In the future work, we will consider parallelizing the FSSEM to increase its speed. In this paper, we consider joint inference of two similar GRNs. In some situations, we may need to infer more than two GRNs, e.g. in different developmental stages. Our future work also aims to develop a new algorithm for joint inference of more than two GRNs.

Funding

This work was supported by the National Science Foundation [Grant No. CCF-1319981], and the National Institute of General Medical Sciences [Grant No. 5R01GM104975].

Conflict of Interest: none declared.

References

- Altshuler, D. et al. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299.
- Bohte, J. et al. (2014) Proximal alternating linearized minimization or nonconvex and nonsmooth problems. *Math. Program.*, **146**, 459–494.
- Browning, S.R. and Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Human Genet.*, **81**, 1084–1097.
- Butte, A.J. and Kohane, I.S. (1999). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. In: *Biocomputing 2000*, pp. 418–429. World Scientific.
- Cai, X. et al. (2013) Inference of gene regulatory networks with sparse structural equation models exploiting genetic perturbations. *PLoS Comput. Biol.*, **9**, e1003068.
- Califano, A. (2011) Rewiring makes the difference. *Mol. Syst. Biol.*, **7**, 463.
- Dai, M. et al. (2005) Evolving gene/transcript definitions significantly alter the interpretation of genechip data. *Nucleic Acids Res.*, **33**, e175.
- Danaher, P. et al. (2014) The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Series B Stat. Method.*, **76**, 373–397.
- Faith, J.J. et al. (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, **5**, e8.
- Friedman, J. et al. (2007) Pathwise coordinate optimization. *Ann. Appl. Stat.*, **1**, 302–332.
- Friedman, J. et al. (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**, 432–441.
- Gardner, T.S. et al. (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, **301**, 102–105.
- Gautier, L. et al. (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.
- Harbison, C.T. et al. (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99.

- Haury,A.-C. *et al.* (2012) Tigress: trustful inference of gene regulation using stability selection. *BMC Syst. Biol.*, **6**, 145.
- Hoefling,H. (2010) A path algorithm for the fused lasso signal approximator. *J. Comput. Graphical Stat.*, **19**, 984–1006.
- Holbrook,J.D. *et al.* (2011) Deep sequencing of gastric carcinoma reveals somatic mutations relevant to personalized medicine. *J. Transl. Med.*, **9**, 119.
- Ideker,T. and Krogan,N.J. (2012) Differential network biology. *Mol. Syst. Biol.*, **8**, 1–565.
- Irizarry,R.A. *et al.* (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
- Liu,B. *et al.* (2008) Gene network inference via structural equation modeling in genetical genomics experiments. *Genetics*, **178**, 1763–1776.
- Logsdon,B.A. and Mezey,J. (2010) Gene expression network reconstruction by convex feature selection when incorporating genetic perturbations. *PLoS Comput. Biol.*, **6**, e1001014.
- Lu,T.-P. *et al.* (2011) Integrated analyses of copy number variations and gene expression in lung adenocarcinoma. *PLoS One*, **6**, e24829.
- Margolin,A.A. *et al.* (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, **7**, S7.
- Mohan,K. *et al.* (2014) Node-based learning of multiple Gaussian graphical models. *J. Mach. Learn. Res.*, **15**, 445–488.
- Neto,E.C. *et al.* (2008) Inferring causal phenotype networks from segregating populations. *Genetics*, **179**, 1089–1100.
- Pock,T. and Sabach,S. (2016) Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems. *SIAM J. Imag. Sci.*, **9**, 1756–1787.
- Shabalin,A.A. (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, **28**, 1353–1358.
- Sonawane,A.R. *et al.* (2017) Understanding tissue-specific gene regulation. *Cell Rep.*, **21**, 1077–1088.
- Statnikov,A. and Aliferis,C.F. (2010) Analysis and computational dissection of molecular signature multiplicity. *PLoS Comput. Biol.*, **6**, e1000790.
- Subramanian,A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.
- Tegner,J. *et al.* (2003) Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc. Natl. Acad. Sci. USA*, **100**, 5944–5949.
- Thieffry,D. *et al.* (1998) From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays*, **20**, 433–440.
- Viallon,V. *et al.* (2016) On the robustness of the generalized fused lasso to prior specifications. *Stat. Comput.*, **26**, 285–301.
- Zhu,J. *et al.* (2007) Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput. Biol.*, **3**, e69.
- Zou,H. (2006) The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.*, **101**, 1418–1429.