

Identifying gene regulatory network rewiring using latent differential graphical models

Dechao Tian^{1,2}, Quanquan Gu^{3,*} and Jian Ma^{1,2,4,*}

¹Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, ²Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA,

³Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22904, USA and

⁴Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received March 06, 2016; Revised June 16, 2016; Accepted June 19, 2016

ABSTRACT

Gene regulatory networks (GRNs) are highly dynamic among different tissue types. Identifying tissue-specific gene regulation is critically important to understand gene function in a particular cellular context. Graphical models have been used to estimate GRN from gene expression data to distinguish direct interactions from indirect associations. However, most existing methods estimate GRN for a specific cell/tissue type or in a tissue-naïve way, or do not specifically focus on network rewiring between different tissues. Here, we describe a new method called Latent Differential Graphical Model (LDGM). The motivation of our method is to estimate the differential network between two tissue types directly without inferring the network for individual tissues, which has the advantage of utilizing much smaller sample size to achieve reliable differential network estimation. Our simulation results demonstrated that LDGM consistently outperforms other Gaussian graphical model based methods. We further evaluated LDGM by applying to the brain and blood gene expression data from the GTEx consortium. We also applied LDGM to identify network rewiring between cancer subtypes using the TCGA breast cancer samples. Our results suggest that LDGM is an effective method to infer differential network using high-throughput gene expression data to identify GRN dynamics among different cellular conditions.

INTRODUCTION

At the level of transcription, gene expression is controlled via transcription factor (TF) proteins that selectively bind to *cis*-regulatory elements to regulate target genes. There are less than 2000 TFs in the human genome and they work

cooperatively to regulate target genes to perform complex cellular functions in specific context (1–3). Such regulatory interactions among TFs and their target genes can be modeled as a gene regulatory network (GRN), where nodes are TFs and their target genes, and edges represent the regulatory relationships. It is acknowledged that gene expression and GRNs are highly dynamic among different tissues (4–6). In other words, some gene regulatory interactions may be very conserved and ubiquitous in different tissue types and many may only occur in certain tissues. Therefore, identifying tissue-specific gene regulation is critically important to understand gene function in a particular cellular context, providing key insights into complex biological systems (7–9). Such knowledge can also help us unravel gene-disease association in a tissue-specific manner (10). In recent years, numerous gene expression data sets across various cell/tissue types and conditions have been collected. For example, the Genotype-Tissue Expression (GTEx) consortium profiled the transcriptomes using RNA-seq over large number of different tissue types in human (5,11) with the advantage of having many biological replicates for each tissue type. Such data provide a great opportunity to more robustly infer tissue-specific GRNs.

To reconstruct GRNs from gene expression data, Gaussian graphical models have been widely used (12–16) (see Materials and Methods section for an introduction). Gaussian graphical models have the advantage of inferring *direct* dependencies between genes that correspond to edges in the estimated network, while missing edges in the estimated network indicate conditional independence. However, most existing methods estimate a GRN for a specific cell/tissue type or in a tissue-naïve way, or do not specifically focus on the network rewiring between different tissues. Therefore, methods for estimating differential networks between two tissue types remain under-explored.

One challenge of using Gaussian graphical model to estimate GRNs in the high dimensional setting (where the number of genes is much greater than the number of samples ($p > n$)) is that the sample covariance matrix is singular and the

*To whom correspondence should be addressed: Tel: +1 412 268 2776, Fax: +1 412 268 2977, Email: jianma@cs.cmu.edu
Correspondence may also be addressed to Q. Gu. Tel: +1 434 243 3428, Fax: +1 434 982 2972, Email: qg5w@virginia.edu

estimation of the GRN is impossible unless we make some assumptions on the GRN, e.g. the estimated GRN is sparse and approaches such as (14) have been developed. To estimate the differential network, one straightforward method is to estimate the network of each tissue type separately and then find the difference between the two estimated networks. However, this straightforward procedure does not take full advantage of the similarity shared between GRNs. In addition, in existing Gaussian graphical models, normal distribution is an important assumption for the gene expression values. However, the gene expression values from high-throughput method such as RNA-seq, even after being normalized, do not follow a normal distribution (17,18) (see our own analysis later). As a result, in Gaussian graphical models, the computed sample covariance matrix from gene expression data cannot precisely capture the associations among genes. The motivation of our new method in this work is to estimate the differential network between two tissue types directly without inferring the network for individual tissues, which has the clear advantage of utilizing much smaller sample size to achieve reliable differential network estimation. Importantly, our new method also does not have the normal distribution assumption for gene expression values.

In this paper, we develop a new method to address the following problem of inferring differential networks using graphical models. Instead of assuming that the GRN is sparse, we only assume that the differential network between two tissue types is sparse. This assumption is reasonable and much milder because for many tissue types, the corresponding GRNs are very similar to each other. For example, on average 72.4% of regulatory interactions are shared between any pair of networks from the 41 transcriptional regulatory networks in (4). In particular, we propose to directly estimate the differential network from high-throughput gene expression data. Our new method is called Latent Differential Graphical Model (LDGM). In LDGM, we have two random vectors X and Y that represent gene expression profiles in two networks and follow two different nonparanormal distributions (19,20). More specifically, a random vector X is said to follow a nonparanormal distribution, if there exists a set of univariate monotonic functions $\{f_j\}_{j=1}^p$ such that $f(X) = [f_1(X_1), \dots, f_p(X_p)]^T \sim N(\mathbf{0}, \Sigma)$ with $\text{diag}(\Sigma) = \mathbf{I}$. It is denoted by $X \sim NPN(\mathbf{f}, \Sigma)$, where $\mathbf{f} = (f_j)_{j=1}^p$. Σ is called the latent correlation matrix for $NPN(\mathbf{f}, \Sigma)$. In other words, we assume $X \sim NPN(\mathbf{f}, \Sigma_X)$ and $Y \sim NPN(\mathbf{g}, \Sigma_Y)$, where $\mathbf{g} = (g_j)_{j=1}^p$. The corresponding latent precision matrices are denoted by $\Theta_X = \Sigma_X^{-1}$ and $\Theta_Y = \Sigma_Y^{-1}$. Different from other high-dimensional Gaussian graphical model based methods, where we need to assume that Θ_X and Θ_Y are sparse, here we *only* assume their difference $\Delta = \Theta_Y - \Theta_X$ is sparse. The key novelty of LDGM is that it directly estimates $\Delta = \Theta_Y - \Theta_X$ from sample latent correlation matrices $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ by quasi log likelihood function maximization with ℓ_1 norm penalty. Based on our simulation evaluation and real data application, we found that LDGM is a very effective model that can be applied to a wide range of differential network inference scenarios.

Our new method is conceptually different from existing approaches. As mentioned above, our model only assumes that the differential network between two tissue types is sparse, while it needs to be assumed that the GRN itself is sparse in Gaussian graphical model based approaches such as graphical lasso (Glasso) (14), joint graphical lasso (JGL) (16) and co-hub node joint graphical lasso (CN-JGL) (15) that we will directly compare performance with in the Results section. Our method is also different from recently developed approaches for constructing tissue-specific networks. In (21), the authors used 987 publicly available genome-scale expression data sets in $\sim 38\,000$ conditions to identify tissue-specific networks. However, the method relies on a compiled list of known interactions from databases such as BioGRID and Gene Ontology annotations, while the goal of our method is to estimate the tissue-specific network rewiring *only* from the gene expression data with the potential to identify novel interactions that have not been annotated. In (22), the authors developed an algorithm called GNAT to derive shared and tissue-specific gene co-expression networks utilizing hierarchy of multiple related tissues. It uses multiple graphical lasso to estimate the precision matrices of Gaussian graphical models on different tissues, and constrains the precision matrices of tissues that were nearby in the hierarchy to have similar entries. When applying the method to two tissues (which is the goal in our work also), the method in GNAT is very similar to JGL with fused lasso penalty (GNAT uses ℓ_2 penalty). Even though the methods developed in (21) and (22) consider related tissues simultaneously to enhance its ability of estimating networks, both of them need to specify reliable tissue relationships or hierarchies. A more relevant work to our method is (23), which was proposed to directly estimate the difference of the precision matrices from two multivariate normal distributions. In contrast, we consider estimating the difference of the latent precision matrices from two nonparanormal distributions. Since nonparanormal distribution is a strictly larger family of distributions including multivariate normal distribution as a special case, our proposed LDGM is more powerful in modeling without the limit of Gaussian data. Moreover, the estimator proposed by (23) is based on estimating equation and solved by linear programming, which is very time consuming in practice (note that we did not compare with (23) in this work because the code from (23) is too slow to be evaluated comprehensively). Our estimator is based on pseudo likelihood and can be solved by accelerated proximal gradient descent efficiently (24). Therefore, our LDGM method not only has unique methodology contribution but also has much broader application potential.

The rest of this paper is organized as follows. We first introduce the details of our LDGM algorithm and the principles of other graphical lasso models in the Materials and Methods section. In the Results section, we first demonstrate the performance of LDGM as compared to other methods on simulated data sets, including JGL and CN-JGL. We then evaluate LDGM by applying to the GTEx dataset to identify network rewiring between brain and blood. Finally, we apply LDGM to the TCGA breast cancer samples to study network differences between cancer subtypes.

MATERIALS AND METHODS

Brief introduction on gaussian graphical models

Before we introduce our new method, we first briefly review Gaussian graphical models. In Gaussian graphical model, a p -dimensional random vector $X = (X_1, \dots, X_p)^T$ follows a multivariate normal distribution $N(\mathbf{0}, \Sigma)$. The conditional independence structure of a pair of marginal random variables X_j and X_k is exactly encoded by the precision matrix $\Theta = \Sigma^{-1}$ (25). More specifically, X_j and X_k are independent conditioned on the other marginal random variables iff $\Theta_{jk} = 0$. Gaussian graphical models can be used for GRN estimation, where each marginal random variable X_j , $1 \leq j \leq p$, corresponds to the expression level of a gene, and the edge weight between X_j and X_k is Θ_{jk} . Therefore, the estimation of GRN can be reduced to the estimation of the precision matrix Θ in Gaussian graphical model. The remaining question is how to estimate Θ based on the covariance matrix $\hat{\Sigma}$. In the high-dimensional setting, the number of genes is much larger than the number of samples, thus the sample covariance matrix $\hat{\Sigma}$ is singular and not invertible. In order to overcome this problem, one has to make some assumptions on Θ . For example, we can assume that Θ is sparse. Under the sparsity assumption, graphical lasso (14) was proposed to estimate the sparse precision matrix Θ as follows:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \log(\det(\Theta)) - \operatorname{tr}(\hat{\Sigma}\Theta) - \lambda \|\Theta\|_1, \quad (1)$$

where $\hat{\Sigma}$ is the sample covariance matrix, λ is a non-negative regularization parameter, Θ_{jk} is the (j, k) -th element in Θ and $\|\Theta\|_1 = \sum_{j,k} |\Theta_{jk}|$ is the element-wise ℓ_1 norm of the matrix Θ . The graphical lasso estimator guarantees that the solution $\hat{\Theta}$ is positive definite. Generally, increasing λ will decrease the number of non-zero elements in $\hat{\Theta}$ that makes $\hat{\Theta}$ sparse. In terms of the network, there is only a small number of edges in the corresponding network, i.e. the estimated GRN is sparse. Based on the estimated sparse precision matrix $\hat{\Theta}$, we can immediately obtain an estimated GRN as follows: if $\hat{\Theta}_{jk} = 0$, there is no edge between the j -th gene and the k -th gene; if $\hat{\Theta}_{jk} \neq 0$, there is an edge between the j -th gene and the k -th gene, and the edge weight is $\hat{\Theta}_{jk}$.

Latent differential graphical models (LDGM)

In this paper, we only consider the problem in the context of high-dimensional gene expression data from two different tissue types. In Figure 1, we illustrate the workflow of LDGM and also the key differences between LDGM and other Gaussian graphical model based methods. Let X and Y be the expression of the p genes in two tissue types. We assume that the gene expression data are sampled from two different nonparanormal distributions (19,20), i.e. $X \sim NPN(\mathbf{f}, \Sigma_X)$ and $Y \sim NPN(\mathbf{g}, \Sigma_Y)$. In other words, we consider two nonparanormal graphical models together. As we explained before, GRNs for the two tissue types can be characterized by the latent precision matrix $\Theta_X = \Sigma_X^{-1}$ and $\Theta_Y = \Sigma_Y^{-1}$, respectively. Some gene regulations are rewired while the other regulatory relationships remain unchanged. These rewired interactions form the differential GRN be-

tween the two tissue types. In particular, the differential network can be defined as $\Delta = \Theta_Y - \Theta_X$ (Figure 1A). There is an estimated edge between the j -th gene and the k -th gene in the differential network iff the corresponding element in Δ , i.e. $\hat{\Delta}_{jk}$, is non-zero. Given the high-dimensional gene expression samples $\mathbf{x}_1, \dots, \mathbf{x}_{n_X}$ of X , and samples $\mathbf{y}_1, \dots, \mathbf{y}_{n_Y}$ of Y (Figure 1B), our goal is to estimate the differential network Δ . A straightforward procedure is to estimate Δ by $\hat{\Delta} = \hat{\Theta}_Y - \hat{\Theta}_X$ (Figure 1G), where $\hat{\Theta}_X$ and $\hat{\Theta}_Y$ are estimators of Θ_X and Θ_Y respectively, using (20) introduced before (Figure 1E and F). However, in order to obtain $\hat{\Theta}_X$ and $\hat{\Theta}_Y$, we have to assume that Θ_X and Θ_Y are sparse and we need sufficient sample size for both tissue types. The goal of our new method is to significantly relax this assumption. We observe that the differential network between two tissue types is typically more sparse and we only need about half the sample size (as compared to estimating the GRN for individual tissues separately) if we focus on estimating the differential network directly. Based on this key rationale, we propose a novel graphical model named LDGM, for differential network inference. In LDGM, instead of assuming that Θ_X and Θ_Y are sparse, we only assume Δ is sparse (Figure 1D).

LDGM directly models the differential network between two tissue types based on high-throughput gene expression data. To achieve this, we directly estimate $\Delta = \Theta_Y - \Theta_X$ based on quasi log likelihood maximization with ℓ_1 norm penalization. Recall that $\Delta = \Theta_Y - \Theta_X$, $\Theta_X = \Sigma_X^{-1}$ and $\Theta_Y = \Sigma_Y^{-1}$, we have:

$$\Sigma_X \Delta \Sigma_Y - (\Sigma_X - \Sigma_Y) = 0. \quad (2)$$

Therefore, a reasonable procedure to estimate Δ is to solve the following estimating equation:

$$\hat{\Sigma}_X \Delta \hat{\Sigma}_Y - (\hat{\Sigma}_X - \hat{\Sigma}_Y) = 0, \quad (3)$$

where we replace the population latent correlation matrices Σ_X and Σ_Y in Equation (2) with the sample latent correlation matrices $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$, respectively. Note that Equation (3) is a Z-estimator (26). The Z-estimator can be translated into an M-estimator (26) by noticing that $\hat{\Sigma}_X \Delta \hat{\Sigma}_Y - (\hat{\Sigma}_X - \hat{\Sigma}_Y)$ can be seen as the score function of the following negative quasi log likelihood function:

$$\ell(\Delta) = \frac{1}{2} \operatorname{tr}(\Delta \hat{\Sigma}_X \Delta \hat{\Sigma}_Y) - \operatorname{tr}(\Delta(\hat{\Sigma}_X - \hat{\Sigma}_Y)), \quad (4)$$

where $\operatorname{tr}(\cdot)$ denotes trace operator of a matrix. Since Equation (4) is the negative quasi log likelihood, we can estimate Δ by using maximum likelihood principle. Moreover, since we assume Δ is sparse, we can use ℓ_1 norm penalty in addition to the maximum likelihood estimator. This leads to the following ℓ_1 norm penalized M-estimator:

$$\hat{\Delta}^1 = \underset{\Delta}{\operatorname{argmin}} \frac{1}{2} \operatorname{tr}(\Delta \hat{\Sigma}_X \Delta \hat{\Sigma}_Y) - \operatorname{tr}(\Delta(\hat{\Sigma}_X - \hat{\Sigma}_Y)) + \lambda \|\Delta\|_1, \quad (5)$$

where $\lambda > 0$ is a regularization parameter and $\|\Delta\|_1 = \sum_{j,k} |\Delta_{jk}|$ is the element-wise ℓ_1 norm of Δ that encourages $\hat{\Delta}^1$ to be sparse. Since $\hat{\Delta}^1$ is not guaranteed to be symmetric, we symmetrize $\hat{\Delta}^1$ by the following procedure (27):

$$\hat{\Delta}_{jk} = \hat{\Delta}_{jk}^1 \mathbb{1}(|\hat{\Delta}_{jk}^1| \geq |\hat{\Delta}_{kj}^1|) + \hat{\Delta}_{kj}^1 \mathbb{1}(|\hat{\Delta}_{jk}^1| < |\hat{\Delta}_{kj}^1|). \quad (6)$$

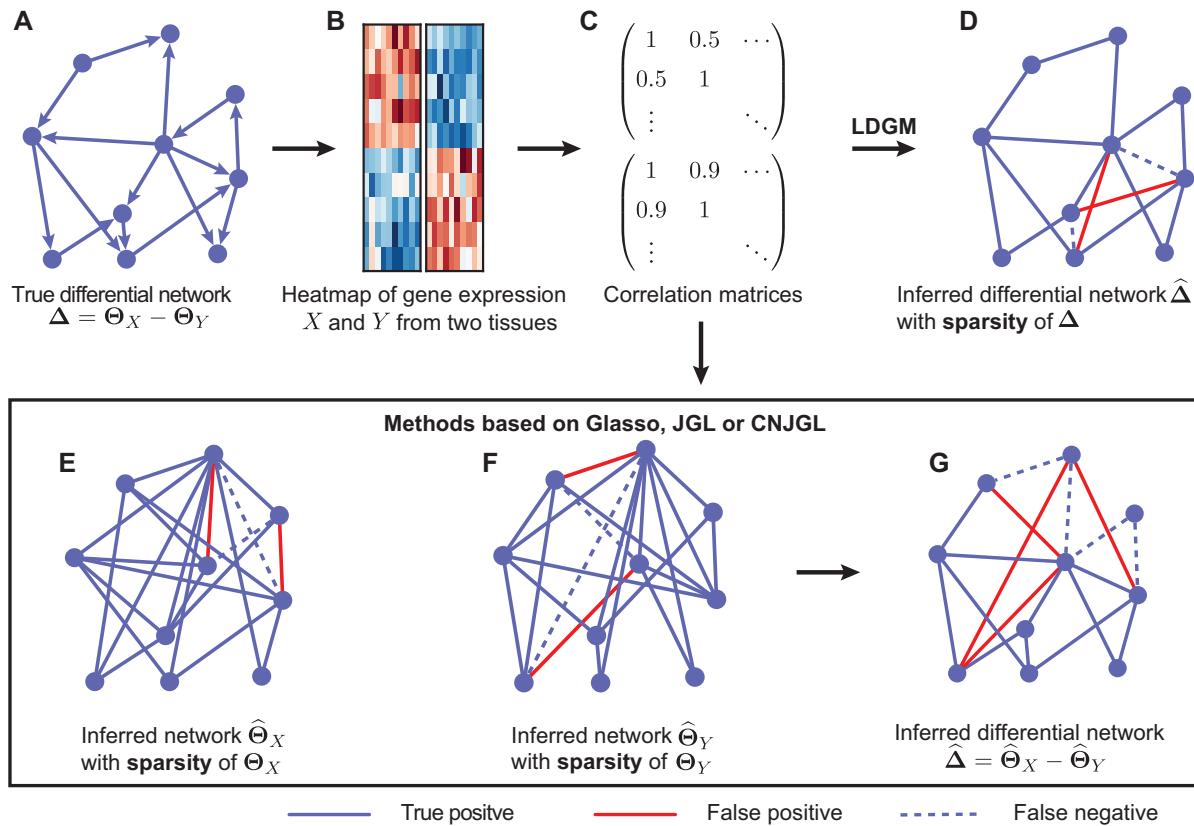


Figure 1. Illustration of Latent Differential Graphical Model (LDGM) as compared to other graphical model based methods. (A) A toy example of a differential network between two tissues. (B) Gene expression levels of genes X and Y involved in the differential network in two tissues. (C) Sample correlation matrices of gene expression levels in the two tissues. (D) LDGM directly infers the differential network from the two correlation matrices. In contrast, other graphical model based methods (Glasso, JGL or CNJGL) first infer individual gene regulatory networks of tissues (E) X and (F) Y from correlation matrices separately, then infer the differential network by the difference between the (G) two reconstructed networks. Red solid lines represent false positive interactions while blue dashed lines represent false negative interactions in the reconstructed networks.

In other words, we take either $\hat{\Delta}_{jk}^1$ or $\hat{\Delta}_{kj}^1$ depending on whose magnitude is larger. Note that we can also add a constraint such as $\Delta = \Delta^T$ in Equation (6) to make $\hat{\Delta}$ symmetric. However, this additional constraint would make the resulting optimization problem more complex and inefficient to solve. The remaining question is how to estimate the latent correlation matrices Σ_X and Σ_Y , respectively. Due to the existence of the marginal monotonic transformations $\{f_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^p$, the estimation of Σ_X and Σ_Y depends on the estimation of $\{f_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^p$. In order to address this challenge, following the idea in (20), instead of estimating $\{f_j\}_{j=1}^p$, $\{g_j\}_{j=1}^p$, Σ_X and Σ_Y simultaneously, we avoid the estimation of $\{f_j\}_{j=1}^p$, $\{g_j\}_{j=1}^p$ by exploiting the relation between the latent correlation matrix and the Kendall tau correlation matrix. In detail, it is shown in (28) that the Kendall tau statistics between X_j and X_k , i.e. denoted by τ_{jk} , and the Pearson correlation coefficient between X_j and X_k , i.e. Σ_{jk} , satisfy the following:

$$\Sigma_{jk} = \sin\left(\frac{\pi}{2} \cdot \tau_{jk}\right).$$

To this end, for $X \sim NPN(\mathbf{f}, \Sigma_X)$, we use the following estimator for the latent correlation matrix Σ_X :

$$[\hat{\Sigma}_X]_{jk} = \begin{cases} \sin\left(\frac{\pi}{2} \hat{\tau}_{jk}\right), & j \neq k, \\ 1, & j = k. \end{cases} \quad (7)$$

where $\hat{\tau}_{jk}$ is the estimator for the Kendall tau statistic

$$\hat{\tau}_{jk} = \frac{2}{n(n-1)} \sum_{1 \leq i < i' \leq n} \text{sign}[(X_{ij} - X_{i'j})(X_{ik} - X_{i'k})].$$

Similarly, we can estimate Σ_Y in the same way. By plugging the above estimators $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ into Equation (5), we can estimate Δ directly.

In order to solve Equation (5) efficiently, by some linear algebra identities, we have $\text{tr}(\mathbf{A}^T \mathbf{B}) = \text{vec}(\mathbf{A})^T \text{vec}(\mathbf{B})$ and $\text{tr}(\mathbf{A}^T \mathbf{B} \mathbf{C} \mathbf{D}^T) = \text{vec}(\mathbf{A})^T (\mathbf{D} \otimes \mathbf{B}) \text{vec}(\mathbf{C})$ for any matrices \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{D} with appropriate size. Note that $\mathbf{D} \otimes \mathbf{B}$ is the Kronecker product (29) of matrices \mathbf{B} and \mathbf{D} . Using these identities, we can rewrite the negative quasi log likelihood in Equation (4) as:

$$\ell(\beta) = \frac{1}{2} \beta^T \mathbf{Q} \beta - \mathbf{b}^T \beta, \quad (8)$$

where $\beta = \text{vec}(\Delta)$, $\mathbf{Q} = \widehat{\Sigma}_X \otimes \widehat{\Sigma}_Y$, $\mathbf{b} = \text{vec}(\widehat{\Sigma}_X - \widehat{\Sigma}_Y)$. Therefore, the estimator in Equation (5) can be rewritten as:

$$\widehat{\beta} = \underset{\beta}{\text{argmin}} \frac{1}{2} \beta^T \mathbf{Q} \beta - \mathbf{b}^T \beta + \lambda \|\beta\|_1, \quad (9)$$

where λ is a non-negative regularization parameter. Increasing λ will make $\widehat{\beta}$ sparse. It is worth noting that Equation (9) can be solved by accelerated proximal gradient descent efficiently (24). Given $\widehat{\beta}$, we can obtain $\widehat{\Delta}^1$ immediately by converting $\widehat{\beta}$ back into a matrix. The estimated differential network can be sparse when we properly choose λ . Note that when we estimate the differential network directly by Equation (9), we only assume the unknown differential network Δ is sparse and we do not assume individual networks Θ_X and Θ_Y are sparse. In addition, LDGM only has one regularization parameter to tune. In contrast, if we apply Gaussian graphical models or non-paranormal graphical models to estimate Θ_X and Θ_Y separately, we need to tune two regularization parameters, which is less convenient in practice.

Other related works

There are existing methods that estimate the precision matrices of two Gaussian graphical models simultaneously. These methods can be directly applied to gene expression data from two tissue types to estimate the corresponding GRNs. To facilitate the method comparison in the Results section, here we briefly introduce two methods developed by others very recently: JGL with fused lasso penalty function (16) and CNJGL (15). Both methods are able to estimate the precision matrices Θ_X and Θ_Y of two Gaussian graphical models simultaneously. The key idea of these methods is to borrow information across different cell types when estimating each network that would lead to more accurate estimation than simply estimating each network individually.

JGL encourages network estimates to share similar edges. It estimates the precision matrices of two Gaussian graphical models simultaneously based on penalized joint log likelihood maximization as follows:

$$\max_{\Theta_X, \Theta_Y} L(\Theta_X) + L(\Theta_Y) - \lambda_1 (\|\Theta_X\|_1 + \|\Theta_Y\|_1) - \lambda_2 \sum_{j \neq k} |\Theta_{jk}^X - \Theta_{jk}^Y|, \quad (10)$$

where $L(\Theta_X)$ is the log likelihood for the Gaussian graphical model on X , i.e. $L(\Theta_X) = \log(\det(\Theta_X)) - \text{tr}(\widehat{\Sigma}_X \Theta_X)$, and similarly, $L(\Theta_Y)$ is the log likelihood for the Gaussian graphical model on Y . λ_1 and λ_2 are non-negative regularization parameters. Note that the fused lasso penalty $\lambda_1 (\|\Theta_X\|_1 + \|\Theta_Y\|_1) + \lambda_2 \sum_{j \neq k} |\Theta_{jk}^X - \Theta_{jk}^Y|$ has been incorporated to encourage the two networks to share similar edges.

CNJGL estimates the precision matrices of two Gaussian graphical models in a similar way to JGL. The only difference is that instead of using fused lasso penalty, CNJGL uses a penalty that encourages the two precision matrices to have a common set of hub nodes in the networks. In particular, CNJGL estimates the precision matrices based on

penalized joint log likelihood maximization as follows:

$$\max_{\Theta_X, \Theta_Y} L(\Theta_X) + L(\Theta_Y) - \lambda_1 (\|\Theta_X\|_1 + \|\Theta_Y\|_1) - \lambda_2 \Omega_q [\Theta_X - \text{diag}(\Theta_X), \Theta_Y - \text{diag}(\Theta_Y)], \quad (11)$$

where $L(\Theta_X)$ and $L(\Theta_Y)$ are the log likelihood for the Gaussian graphical models on X and Y , respectively, and Ω_q is the row-column overlap norm proposed in (15) that encourages network estimates to have a common set of hub nodes.

Note that in this study, when we compare the performance from different methods, we always ran JGL and CNJGL with different λ_2 and reported the best result without explicitly mentioning the corresponding λ_2 . For JGL, $\lambda_2 = 1e-04, 1e-03, \dots, 10$. For CNJGL, $\lambda_2 = c \times n$, where $c = 1e-04, 1e-03, \dots, 10$, following (15). The approaches for selecting λ for LDGM and Glasso and λ_1 for JGL and CNJGL are described in Supplementary Text.

Additionally, (23) proposed to directly estimate the difference of two precision matrices from two multivariate normal distributions that is given by the following estimator:

$$\min_{\Delta} \|\Delta\|_1 \quad (12)$$

subject to $\|\widehat{\Sigma}_X \Delta \widehat{\Sigma}_Y - \widehat{\Sigma}_X + \widehat{\Sigma}_Y\|_{\infty} \leq \lambda$,

where $\widehat{\Sigma}_X$ and $\widehat{\Sigma}_Y$ are sample covariance matrices for the two multivariate normal distributions, $\lambda > 0$ is a tuning parameter.

Methodological comparisons

Both JGL and CNJGL suffer from the problem that they estimate Δ indirectly. In other words, JGL and CNJGL both first estimate Θ_X and Θ_Y separately, and then estimate Δ by $\widehat{\Delta} = \widehat{\Theta}_Y - \widehat{\Theta}_X$ (Figure 1E–G). This requires JGL and CNJGL to access twice the number of observations (i.e. sample size) than LDGM, because estimating Θ_X and estimating Θ_Y individually is as difficult as estimating Δ . In addition, the assumptions in JGL and CNJGL on network topology similarity between two tissues may not hold in reality for many application settings. For example, the hub nodes in the GRNs in two different tissues may not be the same, which is what CNJGL assumes, because key TF proteins in two tissues could be quite different. The above aspects suggest clear conceptual advantages of LDGM over JGL and CNJGL. On the other hand, the idea of our method is similar to (23). Nevertheless, the method in (23) is limited to multivariate normal distributions, while our method is applicable to the non-paranormal family of distributions that admits multivariate normal distribution as a special case. Furthermore, our estimator in Equation (9) can be solved by accelerated proximal gradient descent efficiently, while the estimator in Equation (12) is a constrained optimization problem, which is computationally very time consuming.

RESULTS

Performance evaluation using simulation

We assessed the performance of LDGM by comparing to Glasso, JGL and CNJGL on simulated data sets. We used *hug* function with ‘method=glasso’ from the R package

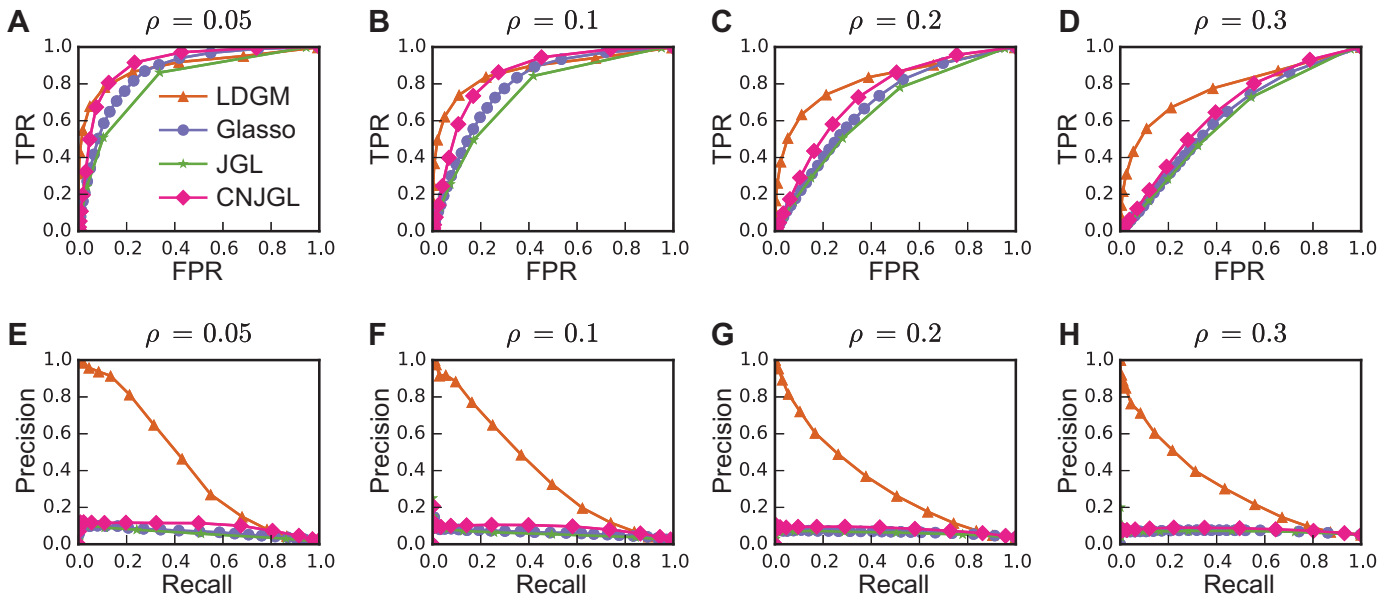


Figure 2. Performance of different methods on simulated data with varied individual network density ρ . (A–D) are ROC curves. (E–H) are precision-recall curves. LDGM consistently achieves better performance than other models. Here the proportion of network-specific edges $\rho_1 = 0.1$, the number of nodes $p = 100$ and the sample size $n = 300$. Each curve is the average over 30 runs.

huge (30) for Glasso (14). For JGL, we used the JGL function with ‘penalty=fused’ from the R package *JGL*. We used the CNJGL source code from (15). Note that we did not include the method from (23) in the comparison because their code is too time consuming, making the comprehensive evaluation infeasible.

We first briefly describe our method that generated the synthetic data. In the simulated networks, we set number of nodes $p = 50, 100$. Network density ρ is defined as the number of edges divided by $p \times (p - 1)/2$. $\rho = 0$ if no edge is in the graph, $\rho = 1$ if every pair of nodes are connected by an edge. We chose ρ as individual network sparsity parameter and set $\rho = 0.05, 0.1, 0.2, 0.3$ in our evaluation. Let ρ_1 be the proportion of edges only found in network G_X when it is compared to network G_Y or only found in G_Y when compared to G_X . When ρ is fixed, the higher ρ_1 is, the more different the networks G_X and G_Y are; so the differential network is denser. We used a combination of ρ and ρ_1 as differential network sparsity parameters. We set $\rho_1 = 0.025, 0.05, 0.1$. We set sample size $n = 100, 200, 300$ for $p = 50$ and $n = 200, 300, 400$ for $p = 100$. In order to assess the performance of LDGM and the other graphical models under various individual network sparsity levels, ρ starts from 0.05 to approximately match the average density (0.045) of GRNs reported in (4). Then ρ gradually increases to 0.3 to cover more general network sparsity levels. Sample size n is selected to guarantee that the true differential network structure can be recovered by graphical models under different combinations of p, ρ, ρ_1 . To make a fair comparison with the other methods which have a Gaussian assumption, we simulated gene expression samples from multivariate normal distributions. Then we computed sample Pearson correlation matrices as the input for all the methods.

For every combination of p, ρ, ρ_1 and n , we repeated the following steps 30 times:

1. We first generated an undirected scale-free (SF) network G_X (since SF networks and biological networks share many topological properties (31)). We generated another network G_Y in the following way. First we made a copy of G_X , denoted as G_Y . Then we randomly rewired a proportion (ρ_1) of edges in G_Y . At each rewiring step, two edges $A-B, C-D$ were randomly chosen from G_Y and substituted with $A-D, C-B$ if they do not already exist in G_Y .
2. We constructed the precision matrices Θ_X and Θ_Y separately from adjacency matrices of G_X and G_Y following the method used in (30). We constructed Δ by $\Theta_Y - \Theta_X$.
3. We applied each of the graphical models to compute $\hat{\Delta}$ from $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$, which were sample correlation matrices computed from n independent samples drawn from $N(\mathbf{0}, \Theta_X^{-1})$ and $N(\mathbf{0}, \Theta_Y^{-1})$, respectively.
4. We computed true positive rate (TPR) and recall by $TP/(TP + FN)$, false positive rate (FPR) by $FP/(TN + FP)$, precision by $TP/(TP + FP)$. Here, TP, FP, TN and FN stand for true positives, false positives, true negatives and false negatives, respectively:

$$TP = \sum_{j < k} \mathbb{1}\{\hat{\Delta}_{jk} \neq 0, \Delta_{jk} \neq 0\},$$

$$FP = \sum_{j < k} \mathbb{1}\{\hat{\Delta}_{jk} \neq 0, \Delta_{jk} = 0\},$$

$$TN = \sum_{j < k} \mathbb{1}\{\hat{\Delta}_{jk} = 0, \Delta_{jk} = 0\},$$

$$FN = \sum_{j < k} \mathbb{1}\{\hat{\Delta}_{jk} = 0, \Delta_{jk} \neq 0\}.$$

To assess the performance of differential network estimation, we first compared the ROC curves (Figure 2A–D)

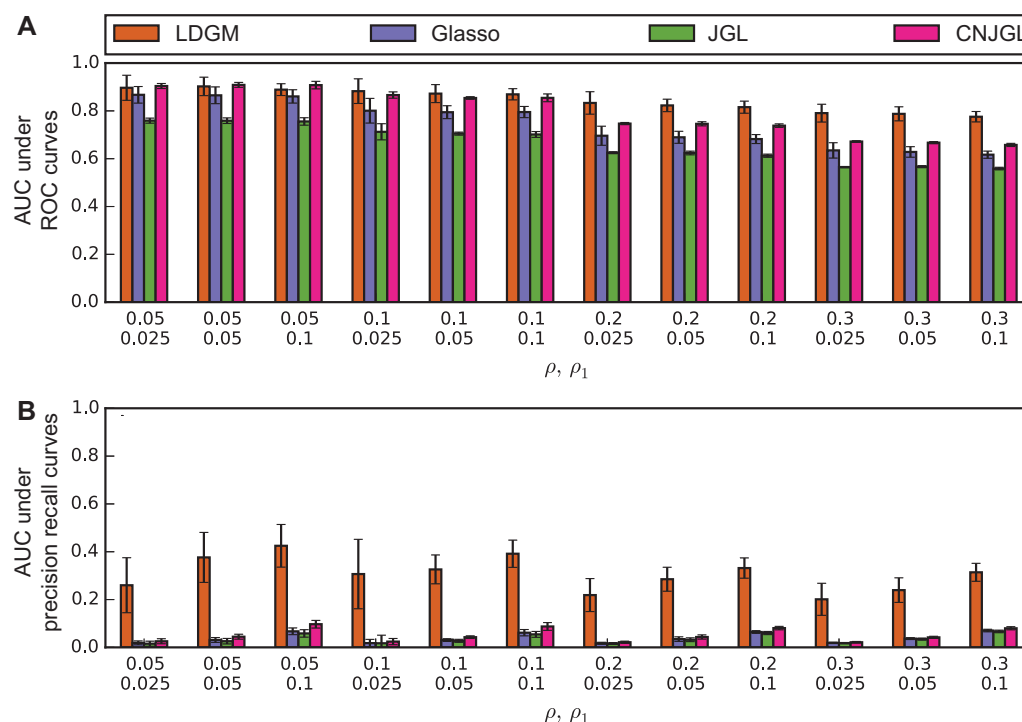


Figure 3. Performance of different methods on simulated data with different ρ and ρ_1 when $p = 100$, $n = 300$. (A) AUC under ROC curves. Advantage of LDGM on ROC becomes more visible when differential networks are more dense with an increased network density ρ . (B) AUC under precision-recall curves. LDGM consistently has a much larger AUC under a precision-recall curve than Glasso, JGL and CNJGL. Bar height represents an AUC under an averaged curve over 30 runs. Error bar represents one standard deviation of AUC under 30 replicated curves.

and precision-recall curves (Figure 2E–F) when ρ is varied with $p = 100$, $n = 300$ and $\rho_1 = 0.1$. Our results suggest that LDGM has clear advantage over other methods in two aspects. First, generally LDGM has better performance based on ROC as compared to other models, and the advantage becomes much clearer when we gradually increase density ρ of G_1 and G_2 (Figure 2A–D). Second, LDGM consistently identifies a remarkably higher proportion of true edges when the estimated differential networks are more sparse, as demonstrated by much larger area under precision-recall curves from LDGM than those from the other models (Figure 2E–H). One main reason for LDGM's better performance is that, as expected, LDGM does not assume individual networks to be sparse while the other graphical models need the sparsity assumption for individual networks.

In addition, the advantages of LDGM on differential network estimation were observed over different combinations of parameters. Area-under-the-curve (AUC) under ROC curves (Figure 3A) and under precision-recall curves (Figure 3B) is computed to numerically summarize the performance of a model under a combination of parameters ρ and ρ_1 , when $p = 100$ and $n = 300$. LDGM has comparable AUC under ROC curves with the other graphical models when $\rho = 0.05, 0.1$. But when ρ gradually increases from 0.05 to 0.3, LDGM has a larger AUC under ROC curve as compared to other methods and the difference becomes more significant (Figure 3A). Besides, LDGM always has a much larger AUC under precision-recall curves than the other methods, regardless of the values of ρ, ρ_1 (Figure 3B).

Furthermore, these advantages of LDGM are robustly held when we vary p, n (Supplementary Figure 1 and Supplementary Figure S2).

Taken together, our simulation results suggest that LDGM outperforms Glasso, JGL and CNJGL in estimating differential networks. When individual graph density is $\rho = 0.05$, which is close to the density of regulatory networks in (4), the estimated sparse differential network by LDGM has a significantly higher proportion of true edges than other methods. When individual graph density increases from 0.05 to 0.3 to represent a more general network sparsity level, the quality of inferred differential networks by LDGM is mildly affected. In contrast, the quality by Glasso, JGL and CNJGL are greatly affected. Overall, our simulation results strongly suggest that LDGM can be robustly applied to a wide range of differential network inference scenarios.

Performance evaluation using the GTEx data sets

We further assessed the performance of estimating differential networks by LDGM and the other graphical models on real data. The GTEx project generated RNA-seq expression data for a large number of human tissues (as of October 2015, there are 8020 samples in more than 60 tissues) (11). However, these samples are not evenly distributed across tissues. Some tissues have a large sample size while others have very limited sample size. Generally, a larger sample size improves performance of graphical lasso models. In this analysis, we estimated differential networks between brain and

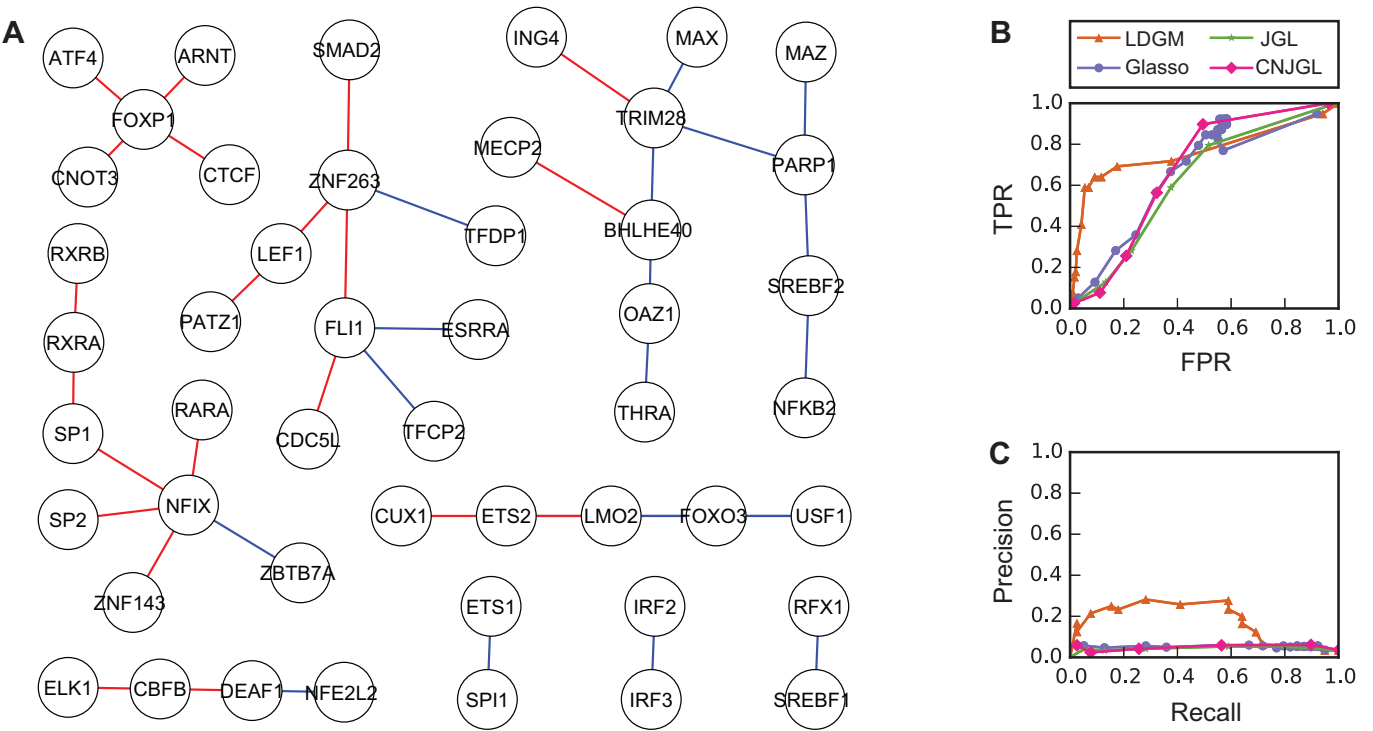


Figure 4. Performance of different methods on the GTEx data (brain and whole blood data). (A) The benchmark network with 48 TFs and 39 tissue-specific interactions. Red edges are interactions specific to brain while blue edges are specific to whole blood. (B) ROC curves and (C) Precision-recall curves of different models to recover the benchmark network. Overall, LDGM outperforms the other models. LDGM, Glasso, JGL and CNJGL have AUC under ROC curves of 0.762, 0.593, 0.617, 0.637 and AUC under precision-recall curves of 0.178, 0.044, 0.048, 0.049, respectively.

blood using the expression data from GTEx and we utilized the network from (4) as a comparison benchmark.

We downloaded the RPKM expression values from 357 samples across 13 human brain tissues and 191 samples for whole blood (dbGaP Accession phs000424.v4.p1). Since human brain tissues are more closely related when compared to other tissues (22), we treated these 357 samples from brain tissues as samples from the human brain. As a common challenge in evaluating the reconstructed GRNs comprehensively, a gold standard differential network is not available for assessing the accuracy of the inferred differential networks, and high-confidence large-scale GRNs are not available to directly construct a benchmark differential network. However, GRNs constructed from high-throughput experimental TF ChIP-seq or DNase-seq open chromatin data for specific tissue types can be used as reasonable benchmark. We therefore constructed our benchmark network based on GRNs from (4).

Neph *et al.* (4) reported the TF regulatory networks of 41 human cell types based on the DNase-seq profiles. There are 7 networks from cell types related to whole blood: B-lymphocyte, B-lymphoblastoid (GM06990 and GM12865), erythroid, haematopoietic stem cell, acute promyelocytic leukemia cell and T-lymphocyte. We constructed the GRN for whole blood from interactions found in at least 6 out of these 7 networks in (4). We also downloaded the brain network. Among the TFs with RPKM >1 in over 80% of the samples in both tissues, 137 TFs are found in both GRNs. These TFs are connected by 2139 interactions that

are common in both tissue types, 412 interactions specific to brain and 356 interactions specific to whole blood. To be more conservative, we used the following rule to keep specific interactions in the benchmark network. Given an interaction $A-B$, we first compute the correlation coefficient of gene expression levels of A and B in brain and whole blood. $A-B$ is kept if the difference between the two correlation coefficients is greater than a threshold (r_b for brain and r_w for blood). The rationale is that a differential interaction $A-B$ indicates that the expression of B is enhanced or inhibited by A only in one tissue. This regulatory relationship changes the expression of B in that tissue, which in turn would change the correlation coefficient between the expression levels of A and B . We set $r_b = 0.8$ and $r_w = 0.7$. Eventually in our benchmark network, brain and whole blood have 20 and 19 tissue specific interactions, respectively. A total of 48 TFs are involved in these 39 specific interactions (Figure 4A). We then generated the correlation matrices of expression levels of these TFs based on RPKM values in brain and whole blood, respectively (see Supplementary Text). Methods with varied tuning parameters were applied to the correlation matrices to estimate a series of differential networks. The inferred differential networks were compared to the benchmark network to evaluate the performance of the models.

Overall, LDGM outperforms the other methods, which is consistent with our simulation results. LDGM has a much higher AUC under ROC curve (Figure 4B). The AUC is 0.762, 0.593, 0.617, 0.637 for LDGM, Glasso, JGL and

CNGJL, respectively. Besides, LDGM identifies a significantly higher proportion of true edges when the estimated differential networks are sparse, as indicated by a much higher precision than the other models when recall is relatively small (Figure 4C). The AUC under precision-recall curves is 0.178, 0.048, 0.044, 0.049 for LDGM, Glasso, JGL and CNGJL, respectively. Moreover, the advantage of LDGM is always observed when r_b and r_w are varied by setting different benchmark data set with different numbers of TFs and tissue-specific interactions (Supplementary Figure S3 and Supplementary Figure S4).

Applying LDGM to TCGA breast cancer data sets

Breast cancer has been classified into five major subtypes based on gene expression: Luminal A, Luminal B, HER2-enriched, Basal-like and normal-like (32). Luminal A subtype is characterized by high expression of estrogen receptor (ER) pathway genes and low expression of proliferation genes, and is associated with a better prognosis (33). Basal-like subtype mostly consists of triple-negative breast cancer that is characterized by low expression levels of ER, progesterone receptor (PR) and HER2, and high expression of genes associated with cell proliferation, and is associated with a poor prognosis (34). As a proof of principle, we applied LDGM using TCGA data to identify differential network between Luminal A and Basal-like subtypes. LDGM, Glasso, JGL and CNGJL with different tuning parameters were performed to identify a series of differential networks with different sparsity levels. Detailed description of the data used here is in Supplementary Text. Note that from these expression values of the genes, we further confirmed that they typically do not completely follow normal distributions (Supplementary Figure S5 and Supplementary Figure S6).

In contrast to the other methods, LDGM consistently identifies ESR1, encoding ER, as one of the genes whose regulatory relationships are rewired greatly between the two subtypes. ER can function as a TF and bind to chromatin directly through estrogen response elements or indirectly by interacting with other TFs, e.g. JUN, SP1, NFKB1. ER can also recruit co-regulators to regulate transcription of target gene expression (35). As demonstrated in Figure 5A and B, ESR1 is consistently among the top genes with highest degrees in differential networks constructed by LDGM. However, it is not among the top 20 genes with highest degrees in differential networks by the other methods. The difference is much greater when the reconstructed networks are more sparse (<150 interactions).

The majority of interactions involved by ESR1 in differential networks by LDGM are from Luminal A subtype, inferred by a principle of majority method based on Glasso (see Supplementary Text). For example, 9 out of 11 interactions involved in ESR1 are from Luminal A subtype in a differential network produced by LDGM with $\lambda = 0.362$ (Figure 5C). This observation is consistent with the high expression of ESR1 in Luminal A subtype and low expression of ESR1 in Basal-like subtype.

To further explore the functions of the reconstructed differential interactions, we performed pathway enrichment analysis by DAVID (36,37) on two exclusive sets of genes in

the reconstructed differential network by LDGM (shown in Figure 5C). One set of genes contains 31 genes where the majority (>50%) of differential interactions are from Basal-like subtype. These genes are significantly enriched in a pathway cadmium-induced DNA synthesis and proliferation in macrophages (FDR = 1.44E-03), including MAPK1, HRAS, MAP2K1, JUN, PLCB. Another set of genes contains 25 genes where the majority of differential interactions (>50%) are from Luminal A subtype. These genes are enriched with a pathway PTEN-dependent cell cycle arrest and apoptosis (FDR=1.40E-02), including AKT1, SOS1, PIK3CA, SHC1 and PIK3R1. Also, AKT1, PIK3CB and SOS1 are involved in a pathway of inhibition of cellular proliferation. Our results are consistent with the characteristics of Luminal A and Basal-like breast cancer subtypes. For example, proliferation related genes are known to have high expressions in Basal-like subtype and low expressions in Luminal A subtype, and PTEN loss frequently occurs in Basal-like subtype (38). All enriched pathways (FDR < 0.05) are reported in Supplementary Table S1.

In addition, we performed analysis to evaluate the differential network related to ESR1 using ChIP-seq data sets generated in MCF-7 breast cancer cell line which has the luminal phenotype. A total of 54 ChIP-seq experiments on ESR1 from MCF-7 cell line were downloaded from CistromeDB (39,40). A putative target gene of ESR1 in MCF-7 cell lines is defined as a gene where there is at least one ESR1 ChIP-seq peak within 5 kbp of the gene in at least 10 out of 54 ChIP-seq experiments. We found that when ESR1 has at least 10 neighbors, over 50% of the neighboring genes in the differential networks are putative target genes of ESR1 in at least 10 ChIP-seq experiments on MCF-7 (Supplementary Figure S7).

These analyses suggest that LDGM has great potential to identify specific differential networks between different cancer subtypes to help better understand molecular mechanisms of tumor heterogeneity.

DISCUSSION

In this paper, we introduced a new method LDGM to infer differential network among different tissues. The novelty of our method is that we now can estimate the differential network between two tissue types directly, without inferring the network for individual tissues and without assuming normal distribution of the gene expression values. This approach also has a clear advantage of utilizing much smaller sample size to achieve reliable differential network estimation. Unlike other Gaussian graphical model based methods that need to assume the GRN is sparse, our method only assumes that the differential network between two tissue types is sparse. In addition, we do not have the assumption on certain topological similarity of the GRNs between two tissues. For example, some previous graphical model based methods assume that the two GRNs have similar hub nodes (e.g. CNJGL), which do not hold for tissues with very different key regulatory proteins that regulate many downstream genes. Our comprehensive simulation results demonstrated that LDGM consistently outperforms other Gaussian graphical model-based meth-

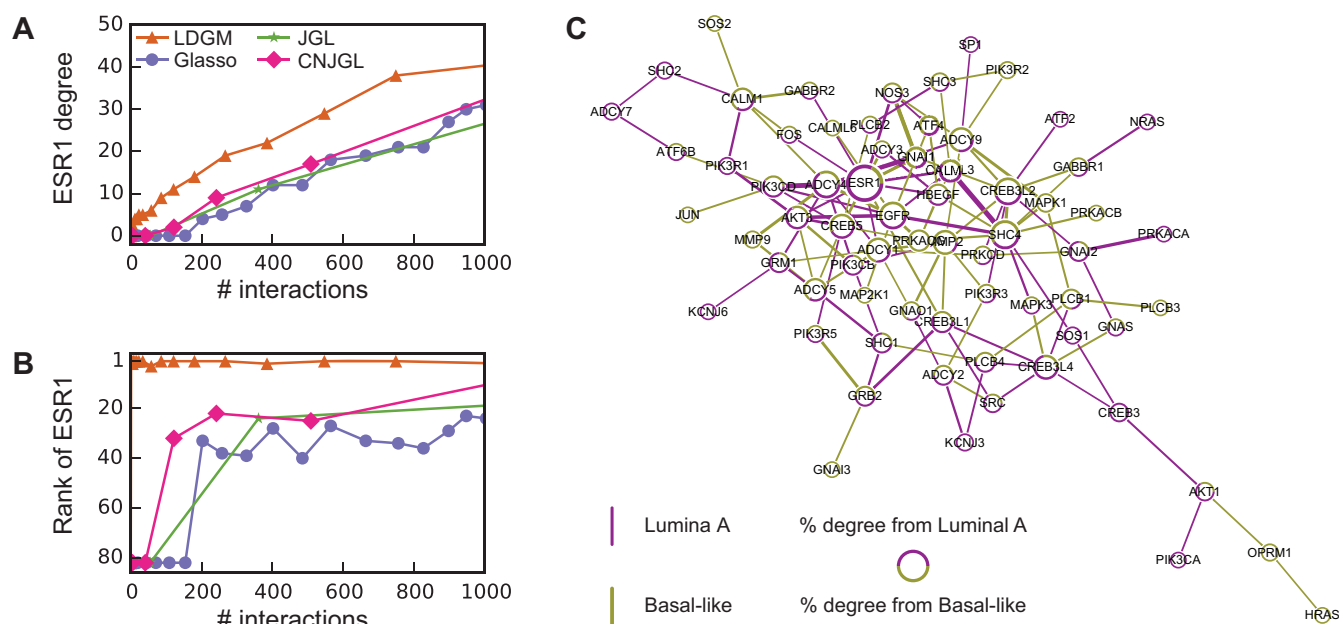


Figure 5. Differential networks on estrogen signaling pathway reconstructed based on gene expression data from breast cancer Luminal A and Basal-like subtypes. (A) The degree of ESR1 in estimated differential networks with increased number of interactions. (B) The rank of ESR1 by its degree in differential networks. The number of interactions is up to 1000 in (A) and (B). (C) A differential network $\hat{\Delta}$ estimated by LDGM with $\lambda = 0.362$. Node size is proportional to the node's degree. Width of an interaction $i - j$ is proportional to the score $|\hat{\Delta}_{ij}|$. The origin of interactions in the differential network is inferred by a principle of majority approach based on Glasso (see Supplementary Text).

ods. This is further confirmed by the evaluation using GTEx data. Finally, we applied our method to the TCGA breast cancer samples to study network rewiring between cancer subtypes. We demonstrated the potential of LDGM to identify subtype specific network interactions that could provide insight into the molecular mechanisms of inter-tumor heterogeneity.

One limitation of LDGM is that it cannot directly tell which tissue type a differential interaction comes from. However, once LDGM identifies the differential network, we can use other method as a subsequent step to distinguish that. For example, in our analysis of the TCGA breast cancer data, we used a principle of majority method based on Glasso to help infer the origin of a differential interactions. It would be an interesting future work to reconstruct differential interactions and their origin simultaneously.

A common challenge in evaluating GRN inference comprehensively using real data is the lack of gold standard. In this work, we made effort to construct a benchmark differential network when comparing LDGM with graphical model based methods using GTEx data. The benchmark network was from (4) where the authors built the network interactions based on the presence of TF binding site motif within DNaseI hypersensitive sites close to the genes. However, this type of network for the entire transcriptome is reasonable but still not perfect. As a matter of fact, it is not always true that TF A regulates gene B whenever there are binding motifs of A within the promoter regions of gene B . This could also be the reason why the average AUC for LDGM and Glasso in the GTEx evaluation are both not very high, aside from the possibility of tissue variation and difference between the two studies. Nevertheless, with more

data from large-scale projects such as the ENCODE project (41) and the Roadmap Epigenomics project (42), we now have access to comprehensive functional genomic profiles to characterize regulatory regions in the human genome across various cell lines and tissues. The data from such high-throughput assays in a given cellular context is very informative to study gene regulation across cell/tissue types even though the sample size for each tissue type is typically very limited. The LDGM method developed in this work may provide a unique way of integrating network inference from large gene expression data sets such as GTEx and regulatory genomics data sets from ENCODE and Roadmap Epigenomics projects to better ascertain the GRN dynamics globally across different tissue types and cell types.

AVAILABILITY OF DATA AND MATERIAL

The source code of our LDGM method can be found at <https://github.com/ma-compbio/LDGM>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for helpful suggestions that improved the manuscript. *Authors' contributions:* Q.G. and J.M. conceived the project and designed the research; Q.G. and D.T. developed the code; D.T. analyzed the data; D.T., Q.G. and J.M. wrote the paper.

FUNDING

National Institutes of Health [HG007352, CA182360 and DK107965 to J.M., in part]; National Science Foundation [1054309 and 1262575 to J.M., in part]. Funding for open access charge: National Science Foundation [1262575].
Conflict of interest statement. None declared.

REFERENCES

- Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N. *et al.* (2010) An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*, **140**, 744–752.
- Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. and Luscombe, N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.
- Davidson, E.H. (2006) *The regulatory genome: gene regulatory networks in development and evolution*. Academic Press, San Diego.
- Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E. and Stamatiou, J.A. (2012) Circuitry and dynamics of human transcription factor regulatory networks. *Cell*, **150**, 1274–1286.
- Ardlie, K.G., Deluca, D.S., Segrè, A.V., Sullivan, T.J., Young, T.R., Gelfand, E.T., Trowbridge, C.A., Maller, J.B., Tukiainen, T., Lek, M. *et al.* (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, **348**, 648–660.
- Melè, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J. *et al.* (2015) The human transcriptome across tissues and individuals. *Science*, **348**, 660–665.
- Ideker, T. and Krogan, N.J. (2012) Differential network biology. *Mol. Syst. Biol.*, **8**, 565.
- Mitra, K., Carvunis, A.-R., Ramesh, S.K. and Ideker, T. (2013) Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet.*, **14**, 719–732.
- Zhang, S., Tian, D., Tran, N.H., Choi, K.P. and Zhang, L. (2014) Profiling the transcription factor regulatory networks of human cell types. *Nucleic Acids Res.*, **42**, 12380–12387.
- Vidal, M., Cusick, M.E. and Barabási, A.-L. (2011) Interactome networks and human disease. *Cell*, **144**, 986–998.
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. *et al.* (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
- Dobra, A., Hans, C., Jones, B., Nevins, J.R., Yao, G. and West, M. (2004) Sparse graphical models for exploring gene expression data. *J. Multivariate Anal.*, **90**, 196–212.
- Yuan, M. and Lin, Y. (2007) Model selection and estimation in the Gaussian graphical model. *Biometrika*, **94**, 19–35.
- Friedman, J., Hastie, T. and Tibshirani, R. (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**, 432–441.
- Mohan, K., London, P., Fazel, M., Witten, D. and Lee, S.-I. (2014) Node-based learning of multiple gaussian graphical models. *J. Mach. Learn. Res.*, **15**, 445–488.
- Danaher, P., Wang, P. and Witten, D.M. (2014) The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. B*, **76**, 373–397.
- Oshlack, A., Robinson, M.D., Young, M.D. *et al.* (2010) From RNA-seq reads to differential expression results. *Genome Biol.*, **11**, 220.
- Li, P., Piao, Y., Shon, H.S. and Ryu, K.H. (2015) Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. *BMC Bioinform.*, **16**, 1.
- Liu, H., Lafferty, J. and Wasserman, L. (2009) The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.*, **10**, 2295–2328.
- Liu, H., Han, F., Yuan, M., Lafferty, J., Wasserman, L. *et al.* (2012) High-dimensional semiparametric Gaussian copula graphical models. *Ann. Stat.*, **40**, 2293–2326.
- Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C. *et al.* (2015) Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.*, **47**, 569–576.
- Pierson, E., Koller, D., Battle, A., Mostafavi, S., Consortium, G. *et al.* (2015) Sharing and specificity of co-expression networks across 35 human tissues. *PLoS Comput. Biol.*, **13**, e1004220.
- Zhao, S.D., Cai, T.T. and Li, H. (2014) Direct estimation of differential networks. *Biometrika*, **101**, 253–268.
- Beck, A. and Teboulle, M. (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, **2**, 183–202.
- Lauritzen, S.L. (1996) *Graphical models*, Oxford University Press, NY.
- Van der Vaart, A.W. (2000) *Asymptotic statistics*. Cambridge University Press, Cambridge, Vol. 3.
- Cai, T., Liu, W. and Luo, X. (2011) A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Am. Stat. Assoc.*, **106**, 594–607.
- Kruskal, W. (1958) Ordinal Measures of Association. **53**, 814–861.
- Golub, G.H. and Van Loan, C.F. (2012) *Matrix computations*. Johns Hopkins University Press, London, Vol. 3.
- Zhao, T., Liu, H., Roeder, K., Lafferty, J. and Wasserman, L. (2012) The huge package for high-dimensional undirected graph estimation in R. *J. Mach. Learn. Res.*, **13**, 1059–1062.
- Barabási, A.-L. and Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.
- Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S. *et al.* (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 8418–8423.
- Ma, C.X., Reinert, T., Chmielewska, I. and Ellis, M.J. (2015) Mechanisms of aromatase inhibitor resistance. *Nat. Rev. Cancer*, **15**, 261–275.
- Rakha, E.A., Reis-Filho, J.S. and Ellis, I.O. (2008) Basal-like breast cancer: a critical review. *J. Clin. Oncol.*, **26**, 2568–2581.
- Hah, N. and Kraus, W.L. (2014) Hormone-regulated transcriptomes: lessons learned from estrogen signaling pathways in breast cancer cells. *Mol. Cell. Endocrinol.*, **382**, 652–664.
- Alvord, G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C. and Lempicki, R.A. (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.*, **8**, 183.
- Sun, H., Qin, B., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C. *et al.* (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.*, **35**(Suppl 2), W169–W175.
- Lehmann, B.D. and Pietenpol, J.A. (2014) Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. *J. Pathol.*, **232**, 142–150.
- Sun, H., Qin, B., Liu, T., Wang, Q., Liu, J., Wang, J., Lin, X., Yang, Y., Taing, L., Rao, P.K. *et al.* (2013) CistromeFinder for ChIP-seq and DNase-seq data reuse. *Bioinformatics*, **29**, 1352–1354.
- Qin, B., Zhou, M., Ge, Y., Taing, L., Liu, T., Wang, Q., Wang, S., Chen, J., Shen, L., Duan, X. *et al.* (2012) CistromeMap: a knowledgebase and web server for ChIP-Seq and DNase-Seq studies in mouse and human. *Bioinformatics*, **28**, 1411–1412.
- Consortium, E.P. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.