Weighted Fused Pathway Graphical Lasso for Joint Estimation of Multiple Gene Networks

Nuosi Wu[1], Jiang Huang[2], Xiao-Fei Zhang[3], Le Ou-Yang[4,5]*, Shan He[6], Zexuan Zhu[2]* and Weixin Xie[1]

[1] College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China, [2] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, [3] School of Mathematics and Statistics, Central China Normal University, Wuhan, China, [4] Guangdong Key Laboratory of Intelligent Information Processing and Shenzhen Key Laboratory of Media Security, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China, [5] Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, China, [6] School of Computer Science, University of Birmingham, Birmingham, United Kingdom

Gene regulatory networks (GRNs) are often inferred based on Gaussian graphical models that could identify the conditional dependence among genes by estimating the corresponding precision matrix. Classical Gaussian graphical models are usually designed for single network estimation and ignore existing knowledge such as pathway information. Therefore, they can neither make use of the common information shared by multiple networks, nor can they utilize useful prior information to guide the estimation. In this paper, we propose a new weighted fused pathway graphical lasso (WFPGL) to jointly estimate multiple networks by incorporating prior knowledge derived from known pathways and gene interactions. Based on the assumption that two genes are less likely to be connected if they do not participate together in any pathways, a pathway-based constraint is considered in our model. Moreover, we introduce a weighted fused lasso penalty in our model to take into account prior gene interaction data and common information shared by multiple networks. Our model is optimized based on the alternating direction method of multipliers (ADMM). Experiments on synthetic data demonstrate that our method outperforms other five state-of-the-art graphical models. We then apply our model to two real datasets. Hub genes in our identified state-specific networks show some shared and specific patterns, which indicates the efficiency of our model in revealing the underlying mechanisms of complex diseases.

Keywords: Gaussian graphical model, precision matrix, prior information, fused lasso penalty, gene network analysis

INTRODUCTION

Most biological processes within cells involve multiple genes (Schlitt and Brazma, 2007; Zhang et al., 2014). Inferring the regulatory relationships between genes is important for understanding the functional organization within cells and helps to reveal the mechanisms of complex diseases (Liu et al., 2016). In recent years, a large number of works have been proposed for inferring gene regulatory networks (GRNs) from gene expression data (Schlitt and Brazma, 2007; Danaher et al., 2014; Verfaillie et al., 2014; Zhang et al., 2014; Liu et al., 2016; Ou-Yang et al., 2017a, Ou-Yang et al., 2017b). Despite their success in addressing some biological problems, revealing the comprehensive

GRNs is still a challenging task (Zhang et al., 2014; Grechkin et al., 2015).

Gaussian graphical model (GGM) is an attractive paradigm to depict the associations among biomolecules (Yuan and Lin, 2007). In GGM, each node of the graph represents a random variable from a random vector subjected to multivariate normal distribution, and there is an edge between two nodes if the corresponding two random variables are conditionally dependent, which means the corresponding element of the precision matrix (or inverse covariance matrix) is non-zero (Dempster, 1972; Uhler, 2017). This property makes GGM so popular because we are able to get the network structure by just estimating the precision matrix (Yuan and Lin, 2007). Unfortunately, in the analysis of gene expression data, the number of samples is usually far less than the dimension of a random vector, which makes it hard to estimate the precision matrix directly (the empirical covariance is not invertible). By assuming that the precision matrix is sparse and the data samples are drawn independently from the same distribution, several approaches have been proposed to estimate the precision matrix (Meinshausen and Bühlmann, 2006; Yuan and Lin, 2007; Banerjee et al., 2008).

However, as biological systems are highly dynamic (Ideker and Krogan, 2012), we are faced with observations collected from different states (Huang et al., 2018a). For example, the gene expression data can be collected from both the diseased and normal tissues (Tian et al., 2016; Ou-Yang et al., 2017a). Thus, if we estimate each state-specific network separately for each sample set, the common structures within different state-specific networks will be ignored. In contrast, inferring a single network from all sample sets may mask their differences. To address this problem, many works have been proposed in recent years to jointly estimate multiple graphical models (Guo et al., 2011; Danaher et al., 2014; Ou-Yang et al., 2017b). When the focus is to infer the differential network between two different states, instead of inferring two state-specific networks, some works have also been developed to estimate the differential network directly (Zhang and Zou, 2014; Yuan et al., 2017).

Although the above methods for learning the structure of multiple GGMs have been successfully used to estimate the regulatory relationships among genes, their performance may be limited since they do not consider the existing knowledge about genes and their regulatory relationships. For example, a pathway is a set of components that interact with each other to perform specific biological tasks. Researches have found that many diseases arose from the joint action of multiple genes within a pathway (Peng et al., 2010). Therefore, pathway-based learning of gene regulatory networks may yield biological insights that are hard to detect by traditional GGMs (Grechkin et al., 2015). Although a pathway graphical lasso model has been proposed to incorporate pathway-based constraints into GGMs (Grechkin et al., 2015), it is designed for single network estimation and cannot jointly estimate multiple GGMs. Moreover, with the accumulation of high-throughput data, we are able to collect some literature-curated gene interactions from public database (Han et al., 2017; Xu et al., 2018). As the state-specific gene regulatory relationships as well as their changes

across multiple states are more likely to take place between genes that are known to have interactions, incorporating these prior information may help to identify the changes of GRNs more accurately (Xu et al., 2018).

To address the above problems, in this paper, we proposed a novel weighted fused pathway graphical lasso (WFPGL) to jointly estimate multiple gene networks as well as their difference by incorporating prior knowledge derived from known pathways and gene interactions. In particular, given a set of pathways, we first assume that regulatory relationships will not take place between genes that belong to different pathways. Here, the pathway information is assumed to be able to provide us prior knowledge that certain edges are unlikely to be present (Grechkin et al., 2015). Under this assumption, the incorrect links across pathways will be eliminated, as shown in **Figure 1**. To make use of the prior knowledge from public gene interaction database and draw support from multiple sample sets collected from different states, we introduce a weighted fused lasso penalty in our model. The proposed WFPGL is optimized by alternating direction method of multipliers (ADMM) (Boyd, 2010), and we follow the idea of pathway graphical lasso that decomposes the original problem into pathway-based subproblem to accelerate the optimization. To demonstrate the performance of our new algorithm, we first conduct simulation studies and compare our algorithm with other five state-of-the-art graphical models. Experiment results on synthetic data show that our WFPGL outperforms other related methods. We then apply our WFPGL on two real datasets. The first experiment is to estimate the gene regulatory networks of insulin sensitive and insulin resistant type 2 diabetes patients. Experiment results demonstrate that our method could identify some promising candidate genes related to insulin resistance. The second experiment is to jointly estimate the gene regulatory networks of four breast cancer subtypes. We find that our identified subtype-specific networks have some shared and specific structures, which may help to reveal the mechanisms of cancer differentiation. Overall, the experiment results on synthetic and real data demonstrate that our WFPGL could effectively utilize prior knowledge to jointly estimate multiple gene networks. The datasets and source codes of our proposed WFPGL are freely available at https://github.com/NuosiWu/WFPGL.

The rest of this paper is organized as follows. In the section Weighted Fused Pathway Graphic Lasso, we formulate our WFPGL and describe the optimization processes. In the section Simulation Studies, we illustrate the performance of WFPGL and another five state-of-the-art methods on synthetic datasets. The section Real Data Analysis provides studies on two real datasets, and the section Conclusion discusses the utilization of the method.

WEIGHTED FUSED PATHWAY GRAPHICAL LASSO

Problem Formulation

Suppose $X^{(1)}, \ldots, X^{(K)}$ are K groups of sample sets that measure the gene expression levels of p genes across K different states. Each $X^{(k)} \in \mathbb{R}^{p \times n}_k$ may have a different sample size n_k, and
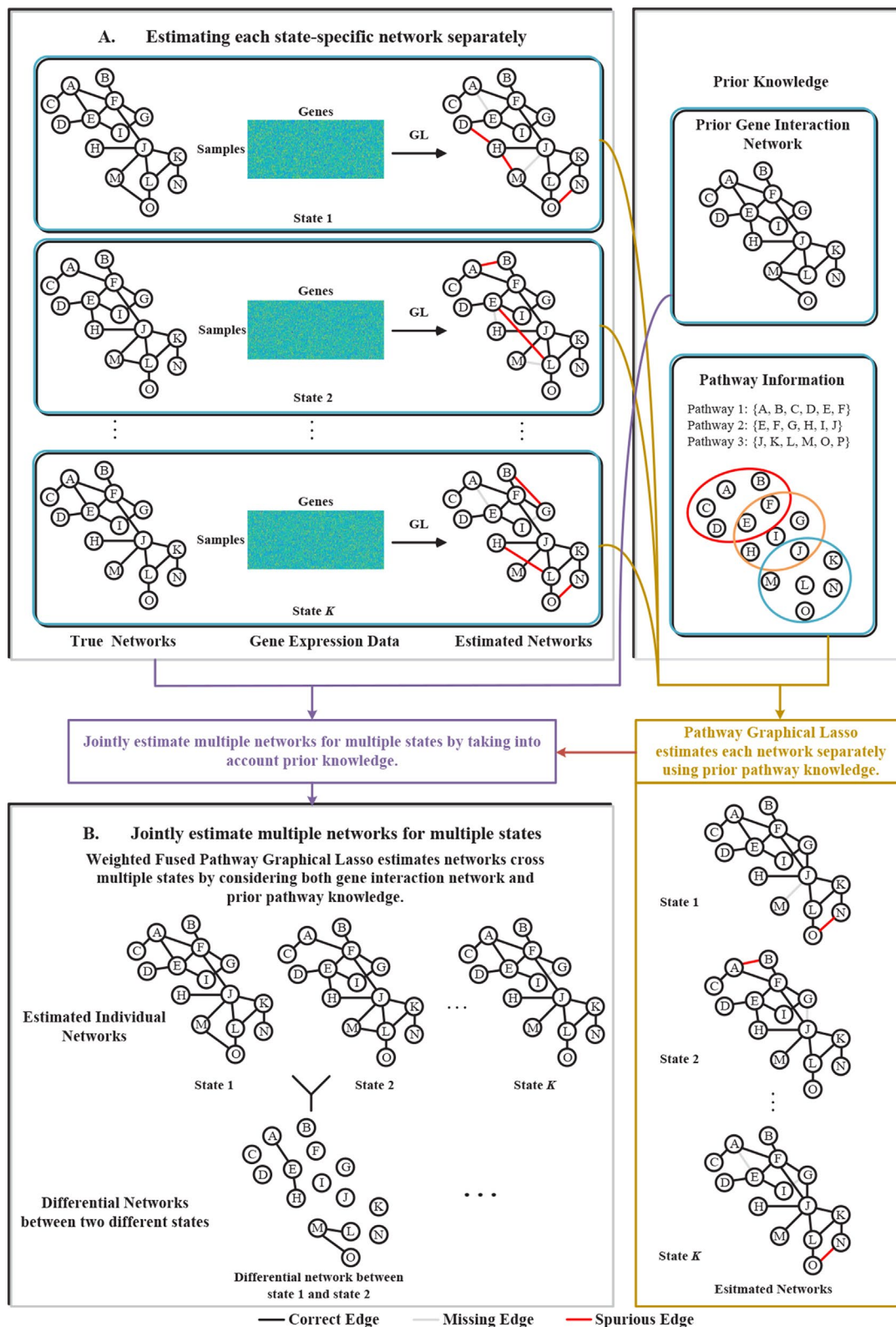
FIGURE 1 | Illustration of the proposed method. **(A)** Graphical lasso just uses gene expression data to separately estimate each state-specific network, leading to incorrect estimation results. **(B)** The proposed weighted fused pathway graphical lasso jointly estimates multiple state-specific networks by considering the prior knowledge of gene interaction networks and pathways, which could eliminate the spurious links between different pathways and results in more accurate estimation of the networks across multiple states.

samples within each group (or state) are independent and follow same multivariate Gaussian distribution. Then maximizing the likelihood functions of multiple Gaussian graphical model is equivalent to solving the following optimization problem (Uhler, 2017):

$$\underset{\Theta}{\text{maximize}} \sum_{k=1}^{K} n_k[\log\{\det(\Theta^{(k)})\} - \text{tr}(S^{(k)}\Theta^{(k)})], \quad (1)$$

where $S^{(k)}$ is the sample covariance matrix for $X^{(k)}$ and $\Theta^{(k)}$ is the inverse covariance matrix (or precision matrix).

To learn the parameters $\Theta^{(k)}$ (which represent networks) for $k = 1, \ldots, K$, additional penalty is often required to guarantee structural recovery especially for high-dimensional setting. Fused lasso penalty, which imposes sparse penalties not only on individual networks but also on the differences between each pair of networks, has been proven to be effective on joint estimation of multiple networks (Danaher et al., 2014). However, traditional fused lasso penalty does not take into account prior information. As the changes of GRNs across different states are more likely to take place between genes that are known to interact with each other (Xu et al., 2018), we encourage the identification of differential edges that appear in the known gene interaction network. Thus, given a prior gene interaction network with adjacency matrix G, we introduce the following weighted fused lasso penalty function:

$$F(\{\Theta^{(k)}\}) = \lambda_1 \sum_{k=1}^{K} \sum_{i \neq j} W_{ij}|\theta_{ij}^{(k)}| + \lambda_2 \sum_{k<k'} \sum_{i,j} W_{ij}|\theta_{ij}^{(k)} - \theta_{ij}^{(k')}|. \quad (2)$$

Here, $\lambda_1 > 0$ is a tuning parameter controlling the sparsity of precision matrices, $\lambda_2 > 0$ is a tuning parameter controlling the sparsity of the differential networks, and W_{ij} is the weight assigned to a pair of genes. In this study, the weight is set as follows:

$$W_{ij} = \begin{cases} 1 & \text{if } G_{ij} = 0, \\ w & \text{if } G_{ij} = 1. \end{cases} \quad (3)$$

where $w \in [0, 1]$ is a predefined parameter. The smaller the value of w, the more likely the corresponding edge will be identified.

As we could make use of pathway information to specify sets of genes that are more likely to work together, incorporating pathway information may help to improve the structure learning of GRNs and achieve more meaningful and interpretable results. Following the ideal of pathway graphical lasso (Grechkin et al., 2015), we constrain the graphical model so that the elements in precision matrices are fixed to 0 if the corresponding gene pairs are not together in any pathways.

Suppose there are J pathways within the p genes (denoted as P_t, $t = 1, \ldots, J$), and considering the pathway constraints and the knowledge of prior gene interaction network, the objective function of our WFPGL model is as follows:

$$\underset{\{\Theta^{(k)}\}}{\text{maximize}} \sum_{k=1}^{K} n_k[\log\{\det(\Theta^{(k)})\} - \text{tr}(S^{(k)}\Theta_{(k)})] - F(\{\Theta^{(k)}\})$$

$$\text{s.t.} \quad \Theta^{(k)} \succeq 0, \ \theta_{ij}^{(k)} = 0, \text{ if } (i,j) \notin \bigcup_{t=1}^{J}\{(i,j)\,|\,i,j \in P_t\},$$

where $F(\{\Theta^{(k)}\})$ takes the form of function (2).

Optimization
Pathway Separable

Given J pathways, let $\Theta_t^{(k)}(t=1,\ldots,J)$ denote a sub-matrix of $\Theta^{(k)}$, which models the sub-graph within the p_t genes in t-th pathway ($p_1 + \cdots + p_J = p$). After some permutations of the rows and columns under the pathway constraints, $\Theta^{(k)}$ can be rearranged into the following form:

$$\Theta^{(k)} = \begin{bmatrix} \Theta_{11}^{(k)} & \Theta_{12}^{(k)} & 0 \\ \Theta_{21}^{(k)} & \Theta_{22}^{(k)} & \Theta_{23}^{(k)} \\ 0 & \Theta_{32}^{(k)} & \Theta_{33}^{(k)} \end{bmatrix} \quad (5)$$

where $\Theta_1^{(k)} = \begin{bmatrix} \Theta_{11}^{(k)} & \Theta_{12}^{(k)} \\ \Theta_{21}^{(k)} & \Theta_{22}^{(k)} \end{bmatrix}$ contains all edges in the first pathway, and $\Theta_{other}^{(k)} = \begin{bmatrix} \Theta_{22}^{(k)} & \Theta_{23}^{(k)} \\ \Theta_{32}^{(k)} & \Theta_{33}^{(k)} \end{bmatrix}$ contains the parameter in the rest of the pathways. $\Theta_{22}^{(k)}$ is the overlapping part of Θ_1 and Θ_{other}, which is a square matrix containing all links among the genes that co-occur in the pathways as well as the remaining pathways. Applying the Schur complement decomposition, we have

$$\det(\Theta^{(k)}) = \det(\Theta_{33}^k) \cdot \det(\Theta_1^k - \Delta_1^{(k)}), \quad (6)$$

where $\Delta_1^{(k)} = [0;\Theta_{23}^k] \cdot (\Theta_{33}^k)^{-1} \cdot [0,\Theta_{32}^{(k)}]$. If we hold the parameters from other pathways fixed and just update the parameters in the first pathway, $\Delta_1^{(k)}$ will be a constant matrix, and problem (1) boils down to maximizing the following function:

$$n_k[\log\{\det(\Theta_1^{(k)}) - \Delta_1^{(k)})\} - \text{tr}(S_1^{(k)}(\Theta_1^{(k)} - \Delta_1^{(k)}))], \quad (7)$$

where the matrix with subscript denotes the sub-matrix corresponding to the first pathway. Notice that (7) is exactly the likelihood function for updating $\Theta_1^{(k)} - \Delta_1^{(k)}$, and $\Delta_1^{(k)}$ is constant and can be pre-calculated. Updating $\Theta_1^{(k)}$ within a certain pathway via maximizing likelihood function only requires information from the corresponding pathway itself. We define a function with such property to be pathway separable.

Optimization Work Flow

Problem (4) is converted into the following equivalent problem:

$$\underset{\{\Theta^{(k)}\},\{Z^{(k)}\}}{\text{minimize}} \quad -\sum_{k=1}^{K} n_k[\log\{\det(\Theta^{(k)})\} - \text{tr}(S^{(k)}\Theta^{(k)})] + F(\{Z^{(k)}\}), \quad (8)$$

$$\text{s.t.} \quad Z^{(k)} = \Theta^{(k)}, k = 1, 2, \ldots, K,$$

which can be solved with the ADMM method (Boyd, 2010) by minimizing the following augmented Lagrangian function:

$$-\sum_{k=1}^{K} n_k[\log\{\det(\Theta^{(k)})\} - \text{tr}(S^{(k)}\Theta^{(k)})] + (F\{Z^{(k)}\})$$

$$+ \sum_{k=1}^{K}(\frac{\mu}{2}\|\Theta^{(k)} - Z^{(k)}\|_F^2 + \langle Y^{(k)}, \Theta^k - Z^{(k)}\rangle), \quad (9)$$

where $<A, B> = \text{tr}(AB^T), \|A\|$ denotes the Frobenius norm of matrix A, $\mu > 0$ is a penalty parameter, and $Y^{(k)}$ is a Lagrangian multiplier. We let $V^{(k)} = Y^{(k)}/\mu$, and (9) can be rewritten as the ultimate objective function:

$$L_\mu(\{\Theta^{(k)}\},\{Z^{(k)}\},V^{(k)}) =$$

$$-\sum_{k=1}^{K} n_k[\log\{\det(\Theta^{(k)})\} - \text{tr}(S^{(k)}\Theta^{(k)})] + F(\{Z^{(k)}\})$$

$$+ \sum_{k=1}^{K}(\frac{\mu}{2}\|\Theta^{(k)} - Z^{(k)} + V^{(k)}\|_F^2 - \frac{\mu}{2}\|V^{(k)}\|_F^2).$$

ALGORITHM 1 Framework of alternating direction method of multipliers (ADMM).

Input: $S, n_k, \mu, \lambda_1, \lambda_2$
Initializing: $\mu = 1$, $\Theta_0^{(k)} = I, V_{(0)}^{(k)} = 0, Z_{(0)}^{(k)} = 0$,
for $k = 1, 2, \ldots, K$.
1: **while** not converged **do**
2: $\{\Theta_{(i)}^{(k)}\} \longleftarrow \arg\min_\Theta L_\mu(\{\Theta^{(k)}\},\{Z_{(i-1)}^{(k)}\},\{V_{i-1}^{(k)}\})$.
3: $\{Z_{(i)}^{(k)}\} \longleftarrow \arg\min_\Theta L_\mu(\{\Theta_{(i)}^{(k)}\},\{Z^{(k)}\},\{V_{i-1}^{(k)}\})$.
4: $\{V_{(i)}^{(k)}\} \longleftarrow \{V_{(i)}^{(k)}\} + \{\Theta_{(i)}^{(k)}\} - \{Z_{(i)}^{(k)}\})$.
5: $i \leftarrow i + 1$
6: **end while**

ADMM solves above problem by three steps in every iteration, as shown in Algorithm 1.

ADMM Solver

It is easy to verify that $\|\Theta^{(k)} - Z^{(k)} + V^{(k)}\|_F^2$ is pathway separable. This allows us to use a block-coordinate descent approach for accelerating the updating of $\Theta^{(k)}$ (Tseng, 2001; Grechkin et al., 2015). By updating $\Theta_t^{(k)}$ in each pathway individually while leaving the parameters in other pathway fixed, the complexity of the problem is narrowed down, and the pathway constraints on the precision matrix naturally meet since the parameters outside

the pathway remain 0. Specifically, let $T_t^{(k)} = \Theta_t^{(k)} - \Delta_t^{(k)}$, and we update $\Theta_t^{(k)}$ by solving the subproblem, as follows:

$$\underset{\{\Theta_t^{(k)}\}}{\text{minimize}} \quad -n_k[\log\{\det(T_t^{(k)})\} - \text{tr}(S_t^{(k)}T_t^{(k)})]$$

$$+ \frac{\mu}{2}\|T_t^{(k)} + \Delta_t^{(k)} - Z_t^{(k)} + V_t^{(k)}\|_F^2.$$

Here, the subscript t of a matrix denotes the index of the sub-matrix that corresponds to the t-th pathway. Grechkin et al. (2015) have introduced an efficient message passing algorithm to calculate $\Delta_t^{(k)}$ efficiently. The solution to this problem is UCU^T, where UDU^T represents the eigendecomposition of $S_t^{(k)} + \mu/n_k(\Delta_t^{(k)} - Z_t^{(k)} + V_t^{(k)})$, and C is a diagonal matrix with i-th diagonal element to be $\frac{n_k}{2\mu}\{-D_{ii} + (D_{ii}^2 + 4\mu/n_k)^{1/2}\}$

The second step is to update $Z^{(k)}$ as the minimizer of

$$\sum_{k=1}^{K}(\frac{\mu}{2}\|\Theta^{(k)} - Z^{(k)} + V^{(k)}\|_F^2) + \lambda_1\sum_{k=1}^{K}\sum_{i\neq j}W_{ij}|z_{ij}^{(k)}|$$

$$+ \lambda_2\sum_{k<k'}\sum_{i,j}W_{ij}|z_{ij}^{(k)} - z_{ij}^{(k')}|). \quad (12)$$

The solution to this problem can be obtained by the method introduced in Hoefling (2010). The last step for each iteration is to update $V^{(k)}$ as $V^{(k)} + \Theta^{(k)} - Z^{(k)}$. We stop the algorithm until $\sum_k\|\Theta_{(i)}^{(k)} - \Theta_{(i-1)}^{(k)}\|_F^2 / \sum_k\|\Theta_{(i-1)}^{(k)}\|_F^2 < 10^{-5}$

SIMULATION STUDIES

We compare the performance of WFPGL with that of five state-of-the-art graphical models: 1) graphical lasso (GL) (Meinshausen and Bühlmann, 2006), which is a classical algorithm for precision matrix estimation; 2) pathway graphical lasso (PGL) (Grechkin et al., 2015), which is a framework that uses pathway knowledge to estimate single Gaussian graphical model; 3) fused graphical lasso (FGL) (Danaher et al., 2014), which is a method for joint estimation of multiple precision matrices across multiple states; 4) differential network estimation *via* D-trace loss (Dtrace) (Yuan et al., 2017), which is a method for direct estimation of a differential network between two states; and 5) weighted D-trace loss (WDtrace) (Xu et al., 2018), which is an algorithm proposed for inferring differential network rewiring by integrating static gene regulatory network information. We implement GL and FGL with their R packages and perform PGL and WFPGL in python environment and using Matlab to carry out Dtrace and WDtrace.

WFPGL has three tuning parameters, i.e., the sparsity controllers λ_1 and λ_2 and the weight of prior information w. For w, as suggested by Xu et al. (2018), we set $w = 0.3$ in the following experiments.

To evaluate the performance of various, we adopt four evaluation metrics named true-positive rate (TPR), false-positive rate (FPR), true-positive differential rate (TPDR), and false-positive

differential rate (FPDR). Let θ_{ij} and $\hat{\theta}_{ij}$ denote the elements in true precision matrix and the estimated precision matrix, respectively. The definitions of these four metrics are as follows:

$$TPR = \frac{\sum_{k=1}^{K} \sum_{i<j} I\{\hat{\theta}_{ij}^{(k)} \neq 0 \quad \text{and} \quad \theta_{ij}^{(k)} \neq 0\}}{\sum_{k=1}^{K} \sum_{i<j} I\{\theta_{ij}^{(k)} \neq 0\}},$$

$$FPR = \frac{\sum_{k=1}^{K} \sum_{i<j} I\{\hat{\theta}_{ij}^{(k)} \neq 0 \quad \text{and} \quad \theta_{ij}^{(k)} = 0\}}{\sum_{k=1}^{K} \sum_{i<j} I\{\theta_{ij}^{(k)} = 0\}},$$

$$TPDR = \frac{\sum_{k<k'} \sum_{i<j} I\{\hat{\theta}_{ij}^{(k)} \neq \hat{\theta}_{ij}^{(k')} \quad \text{and} \quad \theta_{ij}^{(k)} \neq \theta_{ij}^{(k')}\}}{\sum_{k<k'} \sum_{i<j} I\{\theta_{ij}^{(k)} \neq \theta_{ij}^{(k')}\}},$$

$$FPDR = \frac{\sum_{k<k'} \sum_{i<j} I\{\hat{\theta}_{ij}^{(k)} \neq \hat{\theta}_{ij}^{(k')} \quad \text{and} \quad \theta_{ij}^{(k)} = \theta_{ij}^{(k')}\}}{\sum_{k<k'} \sum_{i<j} I\{\theta_{ij}^{(k)} = \theta_{ij}^{(k')}\}},$$

According to the above definitions, TPR and FPR measure the accuracy of network estimation, whereas TPDR and FPDR measure the accuracy of differential network estimation.

Experiments on Two Groups of Samples

In this section, we first consider the situation when there are two groups of samples corresponding to two different states. The main generation procedure of the synthetic datasets is described as follows.

STEP 1: PATHWAY DEFINITION

For ease of simulation study, we simply put successive features into one pathway and let the intersection of two pathways to be non-empty set only when the pathways are "neighbors." We create 10 pathways with same size that covers all 400 features with $n_{ol} = 5$ features overlapped in neighbor pathway. Let P_t ($t = 1, \ldots, 10$) represent a pathway set in which the element indicates the index of feature. In such configuration, pathways are generated as $P_1 = \{1, 2, \ldots, 45\}$, $P_2 = (41, 42, \ldots, 85)$, $\ldots$, $P_{10} = (361, 362, \ldots, 400)$.

STEP 2: NETWORK CONSTRUCTION

We first build a random scale-free network for the state 1, denoted by its adjacency binary matrix $M^{(1)}$. $\tilde{M}^{(1)}$ is a copy of $M^{(1)}$ with each non-zero element substituted by a uniform distribution value on $[-0.6, -0.3] \cup [0.3, 0.6]$. Then $\tilde{M}^{(2)}$ is generated by the copy of $\tilde{M}^{(1)}$, with about $r = 30\%$ non-zero elements vanished.

STEP 3: PRIOR NETWORK GENERATION

We select a proportion (η) of edges from $M^{(1)}$ and connect the corresponding features in prior gene interaction network G.

STEP 4: PRECISION MATRIX CALCULATION

To ensure the positive definiteness of the covariance matrix, we get the real precision matrix $\tilde{\Theta}^{(k)}$ as

$$\tilde{\Theta}^{(k)} = \tilde{M}^{(k)} \circ Q + \sigma I,$$

where $\circ$ is Hadamard product operator, Q is a $p \times p$ binary "pathway network" matrix in which the "1" element indicates the corresponding feature pairs that co-occur in any of the pathways, σ is the absolute value of the minimum eigenvalues of $(\tilde{M}^{(k)} \circ Q)$ for $k = 1, 2$, and I is identity matrix.

After the precision matrix $\tilde{\Theta}^{(k)}$ for each state is settled in Step 4, the synthetic gene expression data could be generated with zero means and $\Sigma = \tilde{\Theta}^{-1}$.
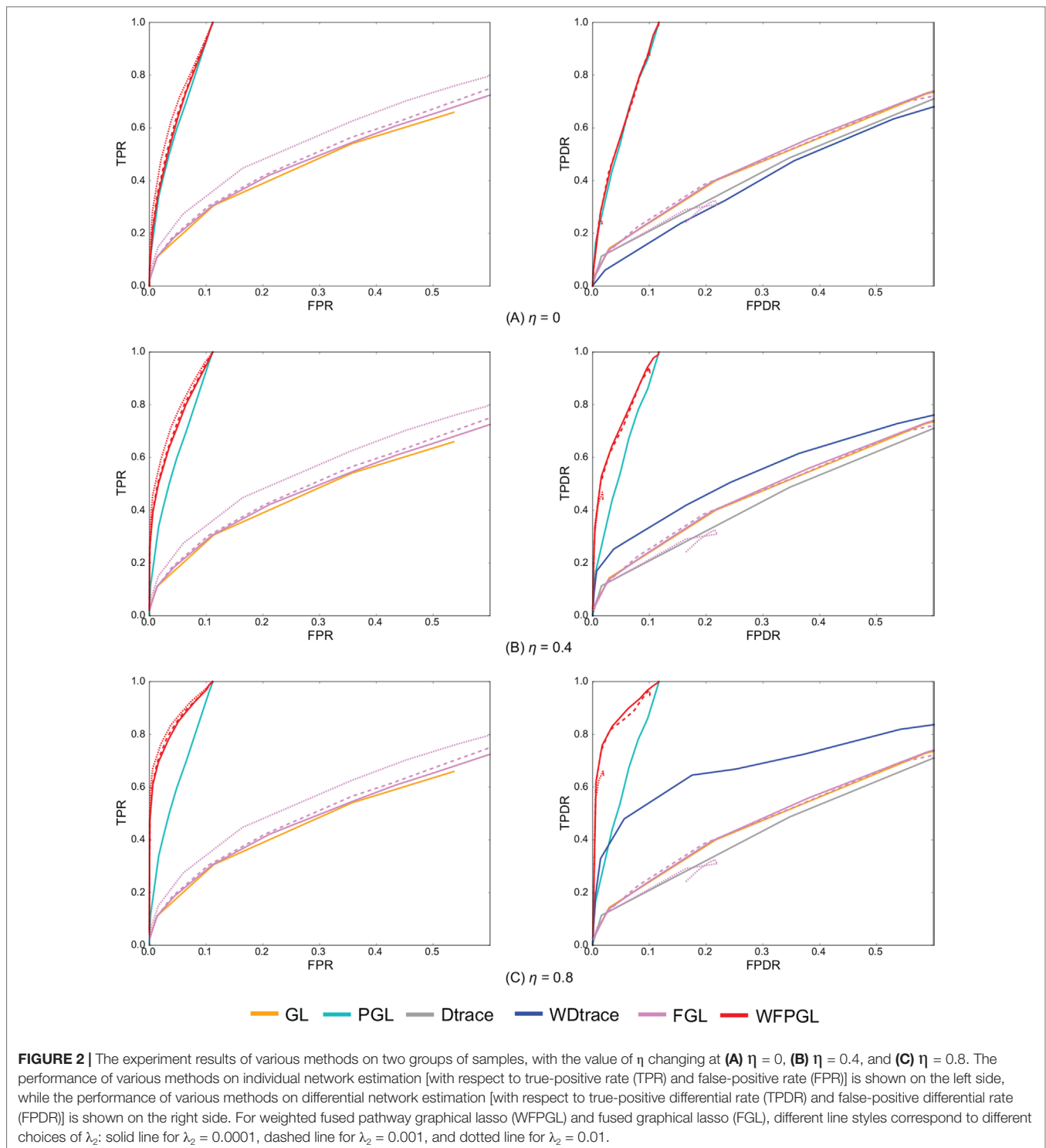
We generate 50 random two-state datasets with the same setting: $n_k = 100$ samples for both two states, 10 pathways with an overlapping number $n_{ol} = 5$ covering all $p = 400$ genes as introduced in Step 1.

As defined in Step 3 of dataset generation, prior rate η controls the proportion of true edges that is covered by prior gene interaction network. Note that the edges in the prior gene interaction network are not necessarily differential edges. We set η to be 0, 0.4, and 0.8 to see its impact on the performance of WFPGL.

Figure 2 presents the average performance of various methods over 50 random generations of data with different values of parameters. In particular, for GL, PGL, Dtrace, and WDtrace that have one sparsity-controlling parameter, we vary the value of the parameter and show their performance. For FGL and WFPGL that have two tuning parameters, i.e., λ_1 and λ_2, which control the sparsity of individual networks and their difference, respectively, we fix the value of λ_2 and show their performance with the value of λ_1 varied. The performance of different methods is shown in distinct colors. Dtrace and WDtrace do not have TPR and FPR, since these two methods directly estimate the differential network between two states and do not predict the individual networks. For GL, PGL, Dtrace, and WDtrace, each curve depicts the performance from a wide range of the sparsity-controlling parameter. For FGL and WFPGL, we pick up three curves here to show their performance. Each curve presents the results with λ_2 being fixed to a certain value: solid line for $\lambda_2 = 0.0001$, dashed line for $\lambda_2 = 0.001$, and dotted line for $\lambda_2 = 0.01$.

The plots on the left side of **Figure 2** show the performance of various methods for individual network estimation. We can find that pathway-based methods WFPGL and PGL dominate other methods that do not use pathway information. Joint estimation methods outperform single network estimation methods (WFPGL performs better than PGL, and FGL performs better than GL). This may be because the joint estimation methods can draw support from multiple sample sets to achieve a more accurate estimation. Though λ_2 controls the sparsity of differential networks, it also affects the accuracy of individual network estimation with larger values result in a better prediction.

The plots on the right side of **Figure 2** show the results of differential network prediction. WFPGL also outperforms other

FIGURE 2 | The experiment results of various methods on two groups of samples, with the value of η changing at **(A)** η = 0, **(B)** η = 0.4, and **(C)** η = 0.8. The performance of various methods on individual network estimation [with respect to true-positive rate (TPR) and false-positive rate (FPR)] is shown on the left side, while the performance of various methods on differential network estimation [with respect to true-positive differential rate (TPDR) and false-positive differential rate (FPDR)] is shown on the right side. For weighted fused pathway graphical lasso (WFPGL) and fused graphical lasso (FGL), different line styles correspond to different choices of λ_2: solid line for $\lambda_2 = 0.0001$, dashed line for $\lambda_2 = 0.001$, and dotted line for $\lambda_2 = 0.01$.

methods, since it uses both two views of prior knowledge, followed by PGL that only uses pathway information and WDtrace that only uses prior gene interaction network knowledge. GL, FGL, and Dtrace, which do not use any prior knowledge, cannot identify differential networks accurately. There is a distinct improvement on the performance of WFPGL and WDtrace

when the value of η increases, which indicates the effectiveness of using prior gene interactions.

To evaluate the performance of various methods without prior network information or prior pathway information, we show the results of various methods without prior network information (i.e., $\eta = 0$) in **Figure 2**, and we show the results

of various methods without prior pathway information in the **Supplementary Material (Figure S1)**. In fact, when there is no prior network information, our method degenerates to a fused pathway graphical lasso model. When there is no prior pathway information, our method degenerates to a weighted fused graphical lasso model. We can find from **Figure 2** and **Figure S1** that when there is no prior network or prior pathway information, our method could still achieve competitive performance with other comparative methods.
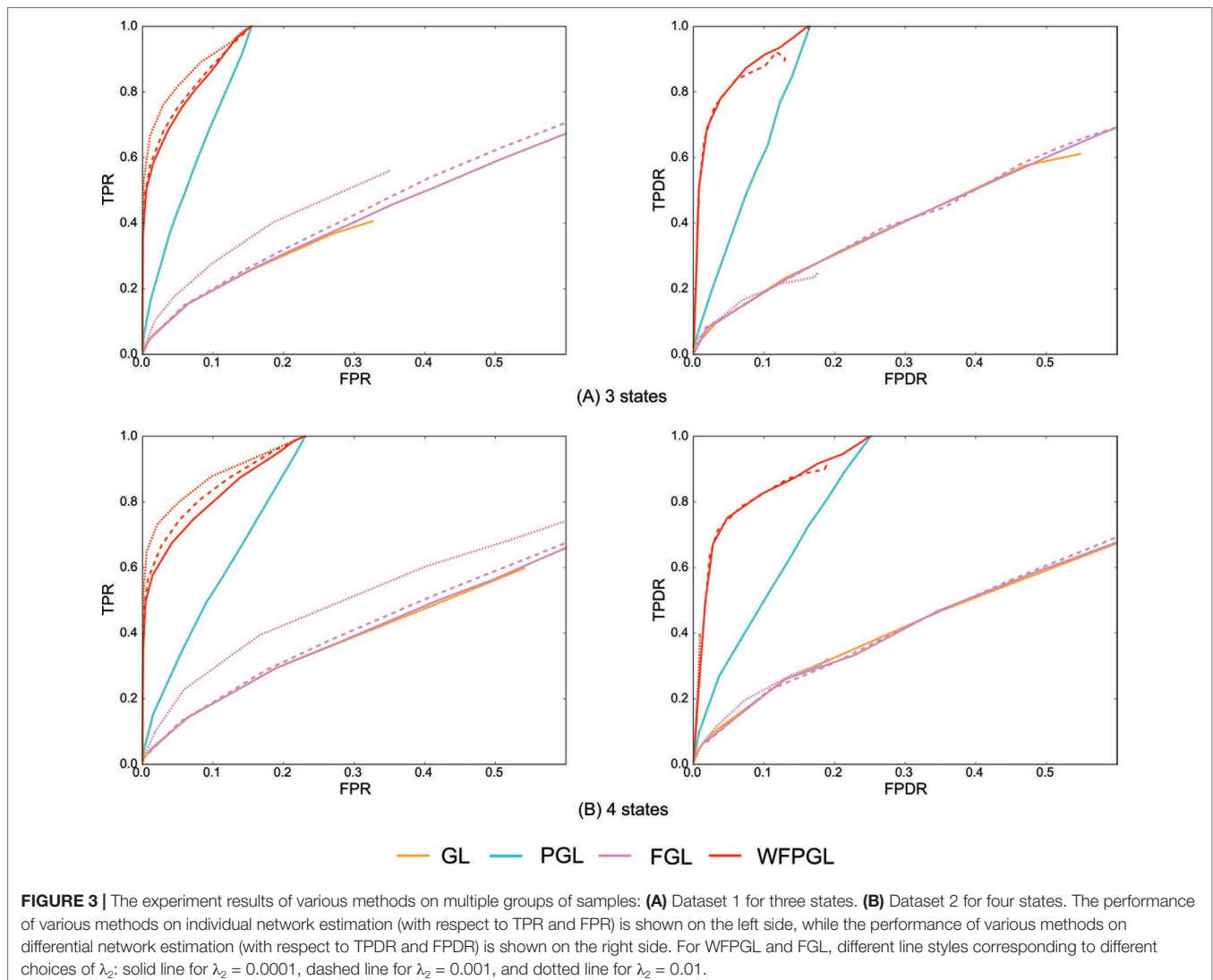
Experiments on Three and Four Groups of Samples

We generate three groups of samples (corresponding to three different states) with $p = 400$ features and $n = 50$ samples per group, and we assume that there are eight pathways covering all features with the overlapping $n_{ol} = 10$. The first two networks are generated in the same way as the above section, except that only

$r = 10\%$ edges are deleted from state 1 to form the network of state 2 (Step 2). The precision matrix for state 3 is copied from state 2, with the removed elements in state 2 being reassigned by uniform distributed values $[-0.6, -0.3] \cup [0.3, 0.6]$ with the possibility of 0.5.

To further test the robustness of WFPGL, we generated a dataset with four different states. In particular, we generated four groups of samples with $p = 200$ features and $n = 50$ samples per group, covered by five pathways with $n_{ov} = 8$. The first three networks are generated the same way as the above section except that the different rate is set to be $r = 20\%$. Then we randomly deleted 20% non-zero elements from state 1 to build the network of state 4.

Similarly, 50 random datasets are generated for the above two situations, and **Figure 3** shows the averaged results with $\eta = 0.8$. Because Dtrace and WDtrace cannot handle the estimation of multiple differential networks, we compare WFPGL with GL, PGL, and FGL. WFPGL still outperforms others on both

FIGURE 3 | The experiment results of various methods on multiple groups of samples: **(A)** Dataset 1 for three states. **(B)** Dataset 2 for four states. The performance of various methods on individual network estimation (with respect to TPR and FPR) is shown on the left side, while the performance of various methods on differential network estimation (with respect to TPDR and FPDR) is shown on the right side. For WFPGL and FGL, different line styles corresponding to different choices of λ_2: solid line for $\lambda_2 = 0.0001$, dashed line for $\lambda_2 = 0.001$, and dotted line for $\lambda_2 = 0.01$.

individual and differential network estimations, followed by PGL, which uses pathway information only. FGL and GL do not perform well on these datasets, though FGL performs slightly better than GL on individual network estimation.

REAL DATA ANALYSIS

In this section, we use WFPGL to estimate the gene regulatory networks of type 2 diabetes patients and breast cancer patients, with 8,444 known gene interactions from TRRUST database (Han et al., 2017) as prior network information and pathways collected from Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa, 2000) as prior pathway information. The tuning parameters (λ_1, λ_2) are determined based on the following Akaike information criterion (AIC):

$$\text{AIC}(\lambda_1, \lambda_2) = \sum_{k=1}^{K} [n_k \, \text{tr}(S^{(k)} \hat{\Theta}^{(k)}) - n_k \log \det(\hat{\Theta}^{(k)}) + 2n_e^{(k)}],$$

where $\hat{\Theta}^{(k)}$ is the estimated precision matrix by using the tuning parameters λ_1 and λ_2, and $2n_e^{(k)}$ is the number of non-zero entities in $\hat{\Theta}^{(k)}$. We select the parameters λ_1 and λ_2, which obtain the outputs that minimize $\text{AIC}(\lambda_1, \lambda_2)$ to generate the final estimation of $\Theta^{(k)}$.

Insulin Resistance in Type 2 Diabetes

We downloaded the gene expression data of type 2 diabetes patients from (Elbein et al., 2012), which contain 31 insulin-resistant (IR) samples and 31 insulin-sensitive (IS) samples. Four relevant pathways are used as prior pathway information, i.e., type 2 diabetes mellitus pathway, Wnt signaling pathway, AMPK signaling pathway, and PI3K–AKT signaling pathway. These four pathways are all implicated in type 2 diabetes. Wnt co-receptor (LRP-5) and the Wnt pathway effector TCF7L2 have been revealed in the development of diabetes (Ip et al., 2012). Targeted drugs have been designed to treat type 2 diabetes by activating AMPK signaling pathway (Jeon, 2016) or PI3K–AKT signaling pathway (Huang et al., 2018b).

We normalize each gene to have 0 mean and 1 standard deviation. By merging rows with the same gene name based on their average, 442 genes are covered by the prior pathway information. According to the experiences in simulation studies, we set $w = 0.3$ and choose λ_1 (ranging from 0.05 to 0.5) and λ_2 (ranging from 0.1 to 1) based on AIC. Networks for both IR and IS samples are built, and we calculate their difference to identify the network rewiring associated with insulin resistance.

For real data analysis, due to the lack of ground truth, it is hard to know whether the predicted regulatory relationships are real. As the changes of gene regulatory relationships are often derived by the aberrations of certain genes, we assess the performance of our method by quantifying how hub nodes in our estimated differential networks capture known functionally important genes. **Table 1** shows the top 10 hub genes in our predicted differential network between IR and IS patients. Among these genes, overexpression of MYC is verified to prevent insulin resistance (Riu et al., 2003); TP53 is found to be responsible for the formation of tissue-specific insulin resistance (Strycharz et al., 2017). In liver, inactivation of RELA leads to the improvement of insulin sensitivity (Ke et al., 2015); whole-body insulin sensitivity improves when VEGFA is overexpressed in adipose tissue (Fatima et al., 2017); PRKAA2 gene interfering insulin resistance has been observed in the Japanese population (Horikoshi et al., 2006). Although not confirmed, indirect evidence like statistics (Naderpoor et al., 2017) or mouse experiments (Williams et al., 2015) indicate that EGFR and ITGA1 may be good candidates for insulin resistance. Besides, how RAC1 relates to insulin sensitivity after exercise has attracted a lot of attention (Maarbjerg et al., 2011).

Breast Cancer Subtypes

We consider the gene expression data from The Cancer Genome Atlas (Cancer Genome Atlas Research Network et al., 2013) that measure the expression levels of 17,327 genes in 511 patients with breast cancer. The observations are classified into four subtypes: 95 for basal-like, 58 for HER2 enriched, and 231 and 127 for luminal A and luminal B, respectively. In this study, we consider seven pathways, namely, apoptosis pathway, hedgehog signaling pathway, homologous recombination pathway, notch signaling pathway, TGF-5 signaling pathway, mTOR signaling pathway, and p53 signaling pathway, which cover 360 genes in the gene expression data. All these pathways are related to breast cancer. Strong evidence shows that reduced apoptosis will cause breast tumor growth, and high levels of apoptosis in a breast tumor are likely to predict worse survival (Parton et al., 2001). DNA repair pathways that targeted genes involved in homologous recombination were discovered to be associated with hereditary breast cancer, while almost 40% of familial and sporadic breast cancers are homologous recombination deficient (den Brok et al., 2017). High notch1 expression was found to be associated with not only high-grade tumors but also poor prognosis for breast cancer (Zardawi et al., 2009). Changes in regulators of p53 activity were demonstrated to be predictive of early relapse in breast cancer (Gasco et al., 2002). mTOR pathway is implicated in endocrine resistance in ER-positive tumors, and the targeted drugs may be

---

TABLE 1 | Top 10 nodes with the highest degree in the predicted differential network between insulin-resistant (IR) and insulin-sensitive (IS) patients.

Rank	1	2	3	4	5	6	7	8	9	10
Name	MYC	TP53	RELA	EGFR	NFKB1	RAC1	VEGFA	CREB3L1	ITGA1	PRKAA2

used to treat brain metastases (Paplomata and O'Regan, 2014). Dysregulation of hedgehog and TGF-5 signaling pathway have been identified in the development and progression of breast cancer (Habib and O'Shaughnessy, 2016; Skoda et al., 2018).

We build four state-specific networks corresponding to four breast cancer subtypes. We still set $w = 0.3$ but choose λ_1 ranging from 0.01 to 0.1 and λ_2 ranging from 0.001 to 0.01. **Table 2** shows the nodes with the highest degree in our inferred networks. Two types of genes are categorized by the fused penalty—cancer-related genes and subtype-specific genes.

The genes appeared in all subtypes are significantly associated with breast cancer. For example, IGF-1 gene is identified as a key player in major signaling pathways involved in breast cancer growth (Christopoulos et al., 2015); TP53 gene mutations have been found in almost all subtypes (Bertheau et al., 2013); TNF gene is highly expressed in breast carcinomas, and its chronic expression supports tumor growth (Kamel et al., 2012); BAMBI and interferon gamma protein (encoded by IFNG gene) are found to inhibit the tumor growth of breast cancer (Shangguan et al., 2012; Ning et al., 2010). In detecting breast cancer, experiments have revealed the diagnostic value of THBS1 protein (Suh et al., 2012). The overexpression and amplification of RPS6KB2 gene as well as FAS gene are reported to be associated with breast cancer prognosis (Pérez-Tenorio et al., 2011; Bębenek et al., 2013); BMP2 is considered as a driving factor for promoting breast

cancer stemness, and BMP4 is a potent suppressor of breast cancer metastasis (Huang et al., 2017; Ampuja et al., 2013).

In addition, the genes that emerged only in one subtype are regarded as potential subtype-specific genes, which may have diverse functions across subtypes. In this result, there are three subtype-specific hub genes identified by our method, i.e., CDKN2B in basal-like, and IL1R1 and TNFRSF10B in luminal B. CDKN2B protein is a cyclin-dependent kinase inhibitor that functions as a cell growth regulator, and its methylation is part of triple negative breast cancer (TNBC) profile (Branham et al., 2012). This result is also in accordance with a recent study (Bareche et al., 2018) that declared that the copy number aberrations of CDKN2B gene suffer a high gain in basal-like 1 subtype. The high expression of IL-1a has been found to be correlated with better prognosis in luminal B breast cancer (Dagenais et al., 2017). Since IL1R1 protein is the receptor for IL-1α, there may be a latent relationship between the IL1R1 gene expression level and luminal B breast cancer. TNFRSF10B (also named as DR5) is a cell surface receptor that can be activated by tumor necrosis factor-related apoptosis inducing ligand (TNFSF10/TRAIL) (Benson et al., 2018). TRAIL kills tumor cells while sparing normal cells and becomes a drug target. However, TRAIL may be selective to patients, since only a small subset of patients respond well to TRAIL in previous clinical trials (Dimberg et al., 2013). Considering that the lack of surface DR5 is sufficient to render tumors resistant to the targeted therapies (Twomey et al., 2015), the correlation between DR5 and luminal B discovered by this paper could provide a new insight of TRAIL therapies on their receivers.

CONCLUSION

In this paper, we propose a novel weighted fused pathway graphical lasso (WFPGL) that can effectively incorporate additional knowledge including pathway information and gene interaction networks to jointly estimate multiple gene regulatory networks. These two kinds of prior information have different effects on our algorithm. We incorporate gene interaction priors by assigning a weight matrix to the estimated individual networks and differential networks. When there is no prior gene interaction information, all elements in the weight matrix are set to 1, and our model degenerates to a fused pathway graphical lasso model. For prior pathway information, we utilize the information by imposing constraints that a pair of genes can be connected to each other only if they co-occur in at least one pathway. The constraints have the potential of improving structure learning of gene regulatory networks. First, it can accelerate the optimization of the algorithm, leading to acceptable results when dealing with high-dimensional data. Second, making use of such prior information in learning the structures of networks can yield results that are more meaningful and interpretable. Moreover, as our algorithm is a flexible framework, the "pathway" here does not need to be an exact biological pathway. The "pathway" here stands for a partition of genes such that genes that belong to different "pathways" are less likely to have regulatory relationships. On the one hand, if we can collect additional information, such as the transcript factors (TFs) of genes in a given pathway, we could combine these TFs and their regulated genes into a new

TABLE 2 | Top 30 nodes with the highest degree in four breast cancer subtypes.

Rank	Basal like	HER2 enriched	Luminal A	Luminal B
1	IGF1	BMP4	PRKACB	BMP4
2	TP53	IGF1	IGF1	IGF1
3	TNF	ID4	BMP4	IFNG
4	BCL2	IFNG	BMP2	FAS
5	FAS	BCL2	TNF	RPS6KB2
6	BMP2	BMP2	RPS6KB2	TP53
7	THBS1	TP53	THBS1	PRKACB
8	PIK3CG	BIRC3	FAS	BAMBI
9	BMP4	RPS6KB1	TP53	THBS1
10	IFNG	FAS	TGFB2	BMP2
11	AKT3	BMPR1B	BIRC3	TNF
12	BAMBI	MYC	BAMBI	PRKAR2B
13	ID4	PIK3CG	BMPR1B	TNFRSF10B
14	CDKN2B	RPS6KB2	RPS6KB1	BMPR1B
15	TGFB2	TNF	AKT3	BMP7
16	INHBB	BMP7	BCL2	ACVR1C
17	RPS6KB2	THBS1	ACVR1C	IL1R1
18	BIRC3	TGFB2	APAF1	BCL2
19	PITX2	INHBB	AKT1	PIK3R1
20	BMP7	DCN	ID4	ID4
21	BMP5	AKT1	LEFTY1	APAF1
22	LEFTY1	PRKACB	FST	BIRC3
23	INHBA	BAMBI	PIK3R1	PITX2
24	MYC	ACVR1C	IFNG	MYC
25	CCNB3	LEFTY1	INHBB	PIK3CD
26	PRKACB	CDKN2A	PRKAR2B	PMAIP1
27	LEFTY2	SMAD9	PIK3CG	DCN
28	BMP6	GADD45A	PIK3R5	AKT1
29	PIK3CD	INHBA	INHBA	INHBB
30	DCN	SERPINB5	PIK3R3	FASLG

"pathway" such that the regulatory relationships are more likely to take place between genes within the same "pathways." On the other hand, if the pathway information is not comprehensive or remains unknown to us, we could treat all genes as a pathway, and our model degenerates to a weighted fused graphical lasso model.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: https://www.genome.jp/; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40234; https://portal.gdc.cancer.gov/

AUTHOR CONTRIBUTIONS

NW conceived and designed the study, performed the statistical analysis, and drafted the manuscript. LO-Y and ZZ conceived the study, and participated in its design and coordination, and helped to draft the manuscript. JH and X-FZ participated in the design of the study, performed the statistical analysis, and helped to revise the manuscript. SH and WX participated in the design of the study and helped to revise the manuscript. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00623/full#supplementary-material

REFERENCES

Ampuja, M., Jokimäki, R., Juuti-Uusitalo, K., Rodriguez-Martinez, A., Alarmo, E.-L., and Kallioniemi, A. (2013). BMP4 inhibits the proliferation of breast cancer cells and induces an MMP-dependent migratory phenotype in MDA-MB-231 cells in 3D environment. *BMC Cancer* 13, 429. doi: 10.1186/1471-2407-13-429

Banerjee, O., Ghaoui, L. E., and d' Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* 9, 485–516. doi: 10.1145/1390681.1390696

Bareche, Y., Venet, D., Ignatiadis, M., Aftimos, P., Piccart, M., Rothe, F., et al. (2018). Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann. Oncol.* 29, 895–902. doi: 10.1093/annonc/mdy024

Bębenek, M., Duś, D., and Koźlak, J. (2013). Prognostic value of the Fas/Fas ligand system in breast cancer. *Contemp. Oncol.* 17, 120. doi: 10.5114/wo.2013.34612

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., et al. (2018). Genbank. *Nucleic Acids Res.* 46, D41–D47. doi: 10.1093/nar/gkx1094

Bertheau, P., Lehmann-Che, J., Varna, M., Dumay, A., Poirot, B., Porcher, R., et al. (2013). p53 in breast cancer subtypes and new insights into response to chemotherapy. *Breast* 22, S27–S29. doi: 10.1016/j.breast.2013.07.005

Boyd, S. (2010). Distributed optimization and statistical learning *via the* alternating direction method of multipliers. *Found. Trends® Mach. Learn.* 3, 1–122. doi: 10.1561/2200000016

Branham, M., Marzese, D., Laurito, S., Gago, F., Orozco, J., Tello, O., et al. (2012). Methylation profile of triple-negative breast carcinomas. *Oncogenesis* 1, e17. doi: 10.1038/oncsis.2012.17

Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., et al. (2013). The cancer genome Atlas pan-cancer analysis project. *Nat. Genet.* 45 (10), 1113–1120. doi: 10.1038/ng.2764

Christopoulos, P. F., Msaouel, P., and Koutsilieris, M. (2015). The role of the insulin-like growth factor-1 system in breast cancer. *Mol. Cancer* 14, 1. doi: 10.1186/s12943-015-0291-7

Dagenais, M., Dupaul-Chicoine, J., Douglas, T., Champagne, C., Morizot, A., and Saleh, M. (2017). The interleukin (IL)-1R1 pathway is a critical negative regulator of PyMT-mediated mammary tumorigenesis and pulmonary metastasis. *Oncoimmunology* 6, e1287247. doi: 10.1080/2162402X.2017.1287247

Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Series B. Stat. Methodol.* 76, 373–397. doi: 10.1111/rssb.12033

Dempster, A. P. (1972). Covariance selection. *Biometrics*, 28, 157–175. doi: 10.2307/2528966

den Brok, W. D., Schrader, K. A., Sun, S., Tinker, A. V., Zhao, E. Y., Aparicio, S., et al. (2017). Homologous recombination deficiency in breast cancer: a clinical review. *JCO Precis. Oncol.* 1, 1–13. doi: 10.1200/PO.16.00031

Dimberg, L. Y., Anderson, C. K., Camidge, R., Behbakht, K., Thorburn, A., and Ford, H. L. (2013). On the trail to successful cancer therapy? Predicting and counteracting resistance against trail-based therapeutics. *Oncogene* 32, 1341. doi: 10.1038/onc.2012.164

Elbein, S. C., Gamazon, E. R., Das, S. K., Rasouli, N., Kern, P. A., and Cox, N. J. (2012). Genetic risk factors for type 2 diabetes: a trans-regulatory genetic architecture? *Am. J. Human Genet.* 91, 466–477. doi: 10.1016/j.ajhg.2012.08.002

Fatima, L., Campello, R., de Souza Santos, R., Freitas, H., Frank, A., Machado, U., et al. (2017). Estrogen receptor 1 (ESR1) regulates VEGFA in adipose tissue. *Sci. Rep.* 7, 16716. doi: 10.1038/s41598-017-16686-7

Gasco, M., Shami, S., and Crook, T. (2002). The p53 pathway in breast cancer. *Breast Cancer Res.* 4, 70. doi: 10.1186/bcr426

Grechkin, M., Fazel, M., Witten, D., and Lee, S., (2015). Pathway graphical lasso, in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, vol. 2015. (Austin, Texas: NIH Public Access), 2617–2623.

Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika* 98, 1–15. doi: 10.1093/biomet/asq060

Habib, J. G., and O'Shaughnessy, J. A. (2016). The hedgehog pathway in triple-negative breast cancer. *Cancer Med.* 5, 2989–3006. doi: 10.1002/cam4.833

Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., et al. (2017). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 46, D380–D386. doi: 10.1093/nar/gkx1013

Hoefling, H. (2010). A path algorithm for the fused lasso signal approximator. *J. Comput. Graph. Stat.* 19, 984–1006. doi: 10.1198/jcgs.2010.09208

Horikoshi, M., Hara, K., Ohashi, J., Miyake, K., Tokunaga, K., Ito, C., et al. (2006). A polymorphism in the AMPKA2 subunit gene is associated with insulin

resistance and type 2 diabetes in the Japanese population. *Diabetes* 55, 919–923. doi: 10.2337/diabetes.55.04.06.db05-0727

Huang, F., Chen, S., and Huang, S.-J. (2018a). Joint estimation of multiple conditional Gaussian graphical models. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 3034–3046. doi: 10.1109/tnnls.2017.2710090

Huang, P., Chen, A., He, W., Li, Z., Zhang, G., Liu, Z., et al. (2017). BMP-2 induces EMT and breast cancer stemness through Rb and CD44. *Cell Death Discov.* 3, 17039. doi: 10.1038/cddiscovery.2017.39

Huang, X., Liu, G., Guo, J., and Su, Z. (2018b). The PI3K/AKT pathway in obesity and type 2 diabetes. *Int. J. Biol. Sci.* 14, 1483. doi: 10.7150/ijbs.27173

Ideker, T., and Krogan, N. J. (2012). Differential network biology. *Mol. Syst. Biol.* 8, 565. doi: 10.1038/msb.2011.99

Ip, W., Chiang, Y.-t. A., and Jin, T. (2012). The involvement of the Wnt signaling pathway and TCF7L2 in diabetes mellitus: the current understanding, dispute, and perspective. *Cell Biosci.* 2, 28. doi: 10.1186/2045-3701-2-28

Jeon, S.-M. (2016). Regulation and function of AMPK in physiology and diseases. *Exp. Mol. Med.* 48, e245. doi: 10.1038/emm.2016.81

Kamel, M., Shouman, S., El-Merzabany, M., Kilic, G., Veenstra, T., Saeed, M., et al. (2012). Effect of tumour necrosis factor-alpha on estrogen metabolic pathways in breast cancer cells. *J. Cancer.* 3, 310. doi: 10.7150/jca.4584

Kanehisa, M. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Ke, B., Zhao, Z., Ye, X., Gao, Z., Manganiello, V., Wu, B., et al. (2015). Inactivation of NF-κB p65 (rela) in liver improves insulin sensitivity and inhibits cAMP/PKA pathway. *Diabetes* 64, 3355–3362. doi: 10.2337/db15-0242

Liu, X., Wang, Y., Ji, H., Aihara, K., and Chen, L. (2016). Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res.* 44, e164–e164. doi: 10.1093/nar/gkw772

Maarbjerg, S. J., Sylow, L., and Richter, E. (2011). Current understanding of increased insulin sensitivity after exercise-emerging candidates. *Acta Physiol.* 202, 323–335. doi: 10.1111/j.1748-1716.2011.02267.x

Meinshausen, N., and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Stat.*, 34, 1436–1462. doi: 10.1214/009053606000000281

Naderpoor, N., Lyons, J. G., Mousa, A., Ranasinha, S., De Courten, M. P., Soldatos, G., et al. (2017). Higher glomerular filtration rate is related to insulin resistance but not to obesity in a predominantly obese non-diabetic cohort. *Sci. Rep.* 7, 45522. doi: 10.1038/srep45522

Ning, Y., Riggins, R. B., Mulla, J. E., Chung, H., Zwart, A., and Clarke, R. (2010). IFNγ restores breast cancer sensitivity to fulvestrant by regulating STAT1, IFN regulatory factor 1, NF-κB, BCL2 family members, and signaling to caspase-dependent apoptosis. *Mol. Cancer Ther.* 9, 1274–1285. doi: 10.1158/1535-7163.MCT-09-1169

Ou-Yang, L., Yan, H., and Zhang, X. F. (2017a). Identifying differential networks based on multi-platform gene expression data. *Mol. Biosyst.* 13, 183–192. doi: 10.1039/C6MB00619A

Ou-Yang, L., Zhang, X.-F., Wu, M., and Li, X.-L. (2017b). Node-based learning of differential networks from multi-platform gene expression data. *Methods* 129, 41–49. doi: 10.1016/j.ymeth.2017.05.014

Paplomata, E., and O'Regan, R. (2014). The PI3K/AKT/MTOR pathway in breast cancer: targets, trials and biomarkers. *Ther. Adv. Med. Oncol.* 6, 154–166. doi: 10.1177/1758834014530023

Parton, M., Dowsett, M., and Smith, I. (2001). Studies of apoptosis in breast cancer. *BMJ* 322, 1528–1532. doi: 10.1136/bmj.322.7301.1528

Peng, G., Luo, L., Siu, H., Zhu, Y., Hu, P., Hong, S., et al. (2010). Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur. J. Hum. Genet.* 18, 111. doi: 10.1038/ejhg.2009.115

Pérez-Tenorio, G., Karlsson, E., Waltersson, M. A., Olsson, B., Holmlund, B., Nordenskjold, B., et al. (2011). Clinical potential of the mTOR targets S6K1 and S6K2 in breast cancer. *Breast Cancer Res. Treat.* 128, 713–723. doi: 10.1007/s10549-010-1058-x

Riu, E., Ferre, T., Hidalgo, A., Mas, A., Franckhauser, S., Otaegui, P., et al. (2003). Overexpression of c-*myc* in the liver prevents obesity and insulin resistance. *FASEB J.* 17, 1715–1717. doi: 10.1096/fj.02-1163fje

Schlitt, T., and Brazma, A. (2007). Current approaches to gene regulatory network modelling. *BMC Bioinf.* 8, S9. doi: 10.1186/1471-2105-8-S6-S9

Shangguan, L., Ti, X., Krause, U., Hai, B., Zhao, Y., Yang, Z., et al. (2012). Inhibition of TGF-β/Smad signaling by BAMBI blocks differentiation of human mesenchymal stem cells to carcinoma-associated fibroblasts and abolishes their protumor effects. *Stem cells* 30, 2810–2819. doi: 10.1002/stem.1251

Skoda, A. M., Simovic, D., Karin, V., Kardum, V., Vranic, S., and Serman, L. (2018). The role of the hedgehog signaling pathway in cancer: a comprehensive review. *Bosn. J. Basic Med. Sci.* 18, 8. doi: 10.17305/bjbms.2018.2756

Strycharz, J., Drzewoski, J., Szemraj, J., and Sliwinska, A. (2017). Is p53 involved in tissue-specific insulin resistance formation? *Oxid. Med. Cellular Longev.* 2017, 15. doi: 10.1155/2017/9270549

Suh, E. J., Kabir, M. H., Kang, U.-B., Lee, J. W., Yu, J., Noh, D.-Y., et al. (2012). Comparative profiling of plasma proteome from breast cancer patients reveals thrombospondin-1 and BRWD3 as serological biomarkers. *Exp. Mol. Med.* 44, 36. doi: 10.3858/emm.2012.44.1.003

Tian, D., Gu, Q., and Ma, J. (2016). Identifying gene regulatory network rewiring using latent differential graphical models. *Nucleic Acids Res.* 44, e140. doi: 10.1093/nar/gkw581

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* 109, 475–494. doi: 10.1023/A:1017501703105

Twomey, J. D., Kim, S.-R., Zhao, L., Bozza, W. P., and Zhang, B. (2015). Spatial dynamics of trail death receptors in cancer cells. *Drug Resist. Updat.* 19, 13–21. doi: 10.1016/j.drup.2015.02.001

Uhler, C. (2017). Gaussian graphical models: an algebraic and geometric perspective. *arXiv preprint arXiv:1707.04345*.

Verfaillie, A., Imrichova, H., Van de Sande, B., Standaert, L., Christiaens, V., Hulselmans, G., et al. (2014). iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput. Biol.* 10, e1003731. doi: 10.1371/journal.pcbi.1003731

Williams, A. S., Kang, L., Zheng, J., Grueter, C., Bracy, D. P., James, F. D., et al. (2015). Integrin a1-null mice exhibit improved fatty liver when fed a high fat diet despite severe hepatic insulin resistance. *J. Biol. Chem.* 290, 6546–6557. doi: 10.1074/jbc.M114.615716

Xu, T., Ou-Yang, L., Hu, X., and Zhang, X.-F. (2018). Identifying gene network rewiring by integrating gene expression and gene network data. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 15, 2079–2085. doi: 10.1109/TCBB.2018.2809603

Yuan, H., Xi, R., Chen, C., and Deng, M. (2017). Differential network analysis *via* lasso penalized D-trace loss. *Biometrika* 104, 755–770. doi: 10.1093/biomet/asx049

Yuan, M., and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* 94, 19–35. doi: 10.1093/biomet/asm018

Zardawi, S. J., O'Toole, S. A., Sutherland, R. L., and Musgrove, E. A. (2009). Dysregulation of hedgehog, Wnt and notch signalling pathways in breast cancer. *Histol. Histopathol.* 24 (3), 385–398. doi: 10.14670/HH-24.385

Zhang, T., and Zou, H. (2014). Sparse precision matrix estimation *via* lasso penalized D-trace loss. *Biometrika* 101, 103–120. doi: 10.1093/biomet/ast059

Zhang, X., Zhao, J., Hao, J.-K., Zhao, X.-M., and Chen, L. (2014). Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks. *Nucleic Acids Res.* 43, e31–e31. doi: 10.1093/nar/gku1315