A Joint Graphical Model for Inferring Gene Networks Across Multiple Subpopulations and Data Types

Xiao-Fei Zhang, Le Ou-Yang, *Member, IEEE*, Ting Yan, Xiaohua Tony Hu, and Hong Yan, *Fellow, IEEE*

***Abstract*—Reconstructing gene networks from gene expression data is a long-standing challenge. In most applications, the observations can be divided into several distinct but related subpopulations and the gene expression measurements can be collected from multiple data types. Most existing methods are designed to estimate a single gene network from a single dataset. These methods may be suboptimal since they do not exploit the similarities and differences among different subpopulations and data types. In this article, we propose a joint graphical model to estimate the multiple gene networks simultaneously. Our model decomposes each subpopulation-specific gene network as a sum of common and unique components and imposes a group lasso penalty on gene networks corresponding to different data types. The gene network variations across subpopulations can be learned automatically by the decompositions of networks, and the similarities and differences among data types can be captured by the group lasso penalty. The simulation studies demonstrate that our method outperforms the state-of-the-art methods. We also apply our method to the cancer genome atlas breast cancer datasets to reconstruct subtype-specific gene networks. Hub nodes in the estimated subnetworks unique to individual cancer subtypes rediscover well-known genes associated with breast cancer subtypes and provide interesting predictions.**

Manuscript received May 2, 2018; revised September 5, 2019; accepted November 2, 2019. Date of publication November 28, 2019; date of current version January 15, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 11871026, Grant 61402190, Grant 61602309, and Grant 61532008, in part by the Natural Science Foundation of Hubei Province under Grant 2018CFB521, in part by the Fundamental Research Funds for the Central Universities under Grant CCNU18TS026, in part by the Shenzhen Fundamental Research Program under Grant JCYJ20170817095210760, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515011384, and in part by the Hong Kong Research Grants Council under Project C1007-15G and Project 11200818. This article was recommended by Associate Editor Y. Jin. *(Corresponding author: Le Ou-Yang.)*

X.-F. Zhang and T. Yan are with the School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China.

L. Ou-Yang is with the Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen Key Laboratory of Media Security and Guangdong Laboratory of Artificial Intelligence and Digital Economy, College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: leouyang@szu.edu.cn).

X. T. Hu is with the College of Computing and Informatics, Drexel University, Philadelphia, PA 19104 USA.

H. Yan is with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong.

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the author.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2019.2952711

***Index Terms*—Gene network inference, graphical models, group lasso penalty, high-dimensional data, the cancer genome atlas (TCGA).**

I. Introduction

THE BIOLOGICAL systems contain many genes that are connected in complicated relationships [1]. The gene dependency networks, where nodes represent genes and edges represent functional dependencies between genes, provide a mathematical representation of complex biological systems. The analysis of gene networks associated with diseases has the potential to shed light on the causes and mechanisms of diseases. Therefore, reconstructing gene networks at the system-wide scale is an important task in the field of bioinformatics [2]–[4].

The accumulation of gene expression data makes it possible to reconstruct gene networks from high-throughput data. A number of computational methods have been developed recently to address this challenge [5]–[13]. Among these methods, the Gaussian graphical models are particularly popular since they can provide a natural tool for modeling the partial correlations between genes [14]. Compared with the pairwise correlations, the partial correlations are much more likely to represent the direct effects in real systems [15]–[19]. To reconstruct gene networks in high-dimensional cases where the number of genes is larger than the number of observations, various sparse Gaussian graphical models have been proposed [15], [16], [20], [21]. One drawback of the Gaussian graphical models is that the data are, in general, non-Gaussian in reality. The nonparanormal Gaussian graphical models, which are more flexible than the Gaussian ones while retaining the good interpretability of the latter, have been proposed and applied to infer gene networks [22]–[24]. All these methods focus on estimating a single gene network from a single dataset by assuming that the observations are identically distributed.

In many applications, the observations can be divided into several distinct but related subpopulations. Taking the breast cancer datasets studied by the cancer genome atlas (TCGA) as an example [25], there exists four main subtypes of breast cancer. One would expect the gene networks corresponding to different subtypes to be similar to each other since they are

based upon the same type of cancer, but also to have important differences stemming from the fact that gene networks are often rewired across different subtypes [17]–[19], [26]–[32]. That is, there are both edges that are common to all subtypes and edges that are unique to individual subtypes. Using the single network reconstruction methods mentioned above, a naive method is to mix data from multiple subpopulations to estimate a single gene network for all subpopulations. This would ignore the differences among subpopulations that are of great biological interest. An alternative method is to reconstruct a gene network for each subpopulation separately, which would be suboptimal since the similarities among subpopulations are ignored.

In addition, the rapidly evolving technologies allow us to measure gene expression levels for a common set of subjects from different data types (e.g., different experiment platforms). For instance, besides mRNA microarrays, RNA-sequencing can also provide gene expression measurements [33]. Since these multifaceted data are collected from related platforms, they often share certain common information. The gene networks reconstructed from different data types should have similar structures. Analyzing each data type independently ignores the common information shared across different data types. Due to differences in scale, collection bias, and noise in each data type, the edge values of gene networks reconstructed from different data types can be different [34].

To estimate the gene networks for multiple subpopulations, several joint Gaussian graphical models have been developed [26], [27], [35]–[45]. These models use various group penalties in the Gaussian-likelihood framework to exploit the common characteristics shared by different subpopulations. These models can also be applied to gene expression data collected from different data types to learn the similarities among all the data. However, the similarities among subpopulations might be different from the similarities among data types. It is reasonable to assume that the gene networks for different cancer subtypes should consist of common edges and subpopulation-unique edges, while the gene networks for different data types should have similar network structures but different edge values. Most of the previous joint graphical models inherently assume that the gene networks are equally similar to each other and do not consider the heterogeneity in similarities among subpopulations and similarities among data types. Zhang *et al.* [34] have proposed a 2-D joint graphical lasso model to jointly learn subpopulation-specific gene dependency networks from gene expression profiles collected from multiple data types. However, this model is only applied to the case with two subpopulations. Therefore, new methods are needed to integrate data from multiple subpopulations and data types to estimate gene networks jointly.

In this article, we propose a new graphical model for joint estimation of gene networks across multiple subpopulations and data types (JEGN) (Fig. 1). Like most joint graphical models, our model is also based on the penalized Gaussian-likelihood framework. A new penalty function is introduced to simultaneously exploit the similarities among subpopulations and the similarities among data types. Our model decomposes

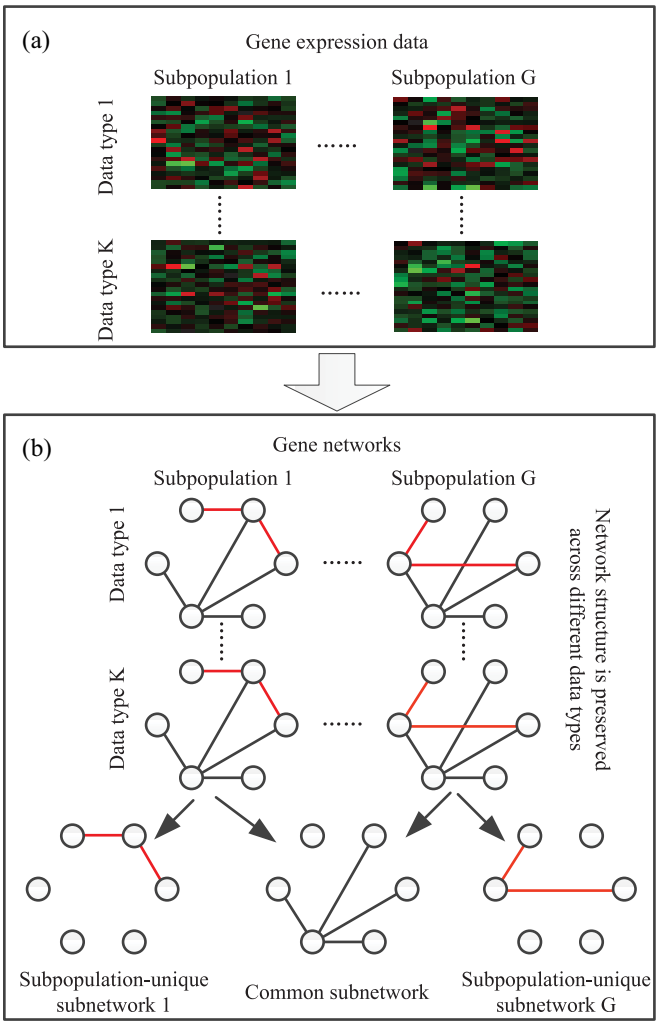


Fig. 1. Overview of JEGN. (a) Gene expression data collected from multiple subpopulations and data types are taken as the input to JEGN. (b) Multiple gene networks corresponding to different subpopulations and data types are inferred jointly by JEGN. Both the similarities among subpopulations and the similarities among data types are exploited. To learn the gene network variations across subpopulations automatically, each subpopulation-specific network is decomposed as a sum of a common subnetwork and a subpopulation-unique subnetwork. A group lasso penalty is applied to gene networks corresponding to different data types so that the network structures are preserved across data types.

each subpopulation-specific network into two parts: 1) a common component capturing the edges shared by all subpopulations and 2) a unique component capturing the edges unique to individual subpopulation. The gene network variations across subpopulations can be learned explicitly via this decomposition. A group lasso penalty is applied to the gene networks across different data types, encouraging an identical network structure across data types. The similarities among data types can be captured by the network structures preserved across the data, and the differences among data types can be quantified by the differences in edge values. We also extend JEGN to deal with the non-Gaussian data which follows the nonparanormal distribution [23], [24], [29], [37], [46]. We first evaluate the empirical performance of the proposed method by comparing with the state-of-the-art methods on simulation data. Then, we apply our method to TCGA data to reconstruct

gene networks for different subtypes of breast cancer from gene expression data collected from different data types. The gene network variations across subtypes are analyzed. We find that the hub nodes in the subtype-unique subnetworks obtained by our method play important roles in characterizing the breast cancer subtypes. The approach to generate simulation data, the used real TCGA data, and our algorithm are implemented as an R package, which is available at https://github.com/Zhangxf-ccnu/JEGN.

II. Methods

A. Brief Review of Sparse Gaussian Graphical Models

The Gaussian graphical models can be used to learn the conditional dependencies among a set of random variables from a set of observations [15], [16], [20]. Let $x_1, \ldots, x_n$ be n observations from a p-dimensional Gaussian random vector $X \coloneqq (X_1, \ldots, X_p)^T \sim N_p(0, \Sigma)$. Without loss of generality, we assume that the data are centered. For Gaussian data, the conditional dependence structure of variables $X_1, \ldots, X_p$ is encoded by the precision matrix (inverse covariance matrix), $\Omega \coloneqq \Sigma^{-1}$. Two variables X_i and X_j are conditionally dependent given the other variables if and only if $\Omega_{ij} \neq 0$. The problem of reconstructing the conditional dependence network is then reduced to the problem of estimating the precision matrix Ω.

A naive way to estimate Ω is via maximum-likelihood, which yields the inverse of the sample covariance matrix, $S = (1/n)\sum_{\ell=1}^{n} x_\ell x_\ell^T$. However, when the number of variables p is larger than the number of observations n, the sample covariance matrix S is not invertible and thus cannot be used to compute an estimator of Ω. To tackle this problem, the sparse Gaussian graphical models have been proposed based on the ℓ_1 penalized maximum-likelihood [16]

$$\min_{\Omega \in S_{++}^p} \; n\{\mathrm{tr}(S\Omega) - \log\det(\Omega)\} + \lambda \sum_{i \neq j} \left|\Omega_{ij}\right| \tag{1}$$

where n is the sample size, $\Omega \in S_{++}^p$ is the positive-definite constraint, $\mathrm{tr}(\cdot)$ is the trace of the matrix, $\det(\cdot)$ is the determinant of the matrix, and λ is a non-negative tuning parameter. The solution to this optimization problem provides a sparse estimator of the precision matrix Ω.

B. Problem Formulation

Suppose that we are given gene expression measurements on a common set of p genes for observations from G subpopulations (e.g., different cancer subtypes), and for each observation, the gene expression measurements are collected from K different data types (e.g., different experimental platforms) [Fig. 1(a)], our goal is to jointly reconstruct subpopulation-specific gene networks by integrating different data types [Fig. 1(b)]. For the gth subpopulation ($g = 1, \ldots, G$) and the kth data type ($k = 1, \ldots, K$), we assume that the observations are independently and identically distributed: $x_1^{kg}, \ldots, x_{n_g}^{kg} \sim N_p(0, \Sigma^{kg})$, where n_g is the number of observations for the gth subpopulation. Let $\Omega^{kg} \coloneqq (\Sigma^{kg})^{-1}$ be the precision matrix for the gth subpopulation and the kth data type. Based on the Gaussian graphical models, the problem of reconstructing subpopulation-specific gene networks from different data types can be formulated as the problem of estimating the KG precision matrices Ω^{kg} for $k = 1, \ldots, K$ and $g = 1, \ldots, G$.

C. Graphical Model for Joint Estimation of Gene Networks

Let $S^{kg} = (1/n_g)\sum_{\ell=1}^{n_g} x_\ell^{kg}(x_\ell^{kg})^T$ be the sample covariance matrix for the gth subpopulation and the kth data type. To make full use of the similarities among subpopulations and the similarities among data types, we extend the sparse Gaussian graphical model (1) to the case of multiple subpopulations and data types following the penalized Gaussian-likelihood framework [26], [34], [36]:

$$\min_{\{\Omega\}} \underbrace{\sum_{k=1}^{K}\sum_{g=1}^{G} n_g \left\{\mathrm{tr}\left(S^{kg}\Omega^{kg}\right) - \log\det\left(\Omega^{kg}\right)\right\}}_{\text{negative log-likelihood}} + \underbrace{P(\{\Omega\})}_{\text{penalty}}. \tag{2}$$

Here, we denote $\{\Omega\} = \{\Omega^{kg} : k = 1, \ldots, K,\ g = 1, \ldots, G\}$ for the sake of simplicity. The first term is the negative log-likelihood function (up to a constant), and the second term $P(\{\Omega\})$ is a penalty function that controls the sparsity level of gene networks and the extent of similarity among gene networks.

One would expect that the gene networks corresponding to different subpopulations should share certain common edges that can capture the similarities among subpopulations, and also have subpopulation-unique edges that can explain the network variations across subpopulations. To explicitly identify the common edges and the subpopulation-unique edges, for data type k, we model Ω^{kg} as $\Omega^{kg} = R^k + M^{kg}$, where R^k represents the common component which captures the edges common to all subpopulations, and M^{kg} represents the subpopulation-unique component which captures the edges unique to the gth subpopulation [Fig. 1(b)]. If the G gene networks, $\Omega^{k1}, \ldots, \Omega^{kG}$, are similar to each other, the subpopulation-unique component M^{kg} would be sparse. If the G gene networks are different from each other, the common component R^k would be sparse. Since different data types measure the activities of genes in a multiview manner, one would expect that the gene networks estimated from different data types should have an identical structure. However, due to differences in scale, collection bias, and noise, the edge values could be different across data types.

To simultaneously learn the similarities among subpopulations and the similarities among data types, we develop a new penalty function

$$\begin{aligned} P(\{\Omega\}) = {} & \lambda\alpha G \sum_{i \neq j} \sqrt{\sum_{k=1}^{K}\left(R_{ij}^k\right)^2} \\ & + \lambda(1-\alpha)\sum_{g=1}^{G}\sum_{i \neq j}\sqrt{\sum_{k=1}^{K}\left(M_{ij}^{kg}\right)^2} \end{aligned} \tag{3}$$

with $\Omega^{kg} = R^k + M^{kg}$. The first term applies a group lasso penalty to the (i, j)th element across the K common components for different data types, encouraging a similar pattern of sparsity across the K common matrices, $\{R^1, \ldots, R^K\}$. The second term applies a group lasso penalty to the (i, j)th element across the K subpopulation-unique components for different data types, encouraging a shared network structure, $\{M^{1g}, \ldots, M^{Kg}\}$. Since the network structures of both common and unique components are preserved across all data types, for each subpopulation, the K estimated networks, $\{\Omega^{1g}, \ldots, \Omega^{Kg}\}$, will have an identical network structure. Here, we decompose Ω^{kg} as $\Omega^{kg} = R^k + M^{kg}$ so that we can automatically identify the common subnetworks shared across all subpopulations and the subpopulation-unique subnetworks that only belong to individual subpopulations. We use the group lasso penalties so that our model can identify the network structures shared by different data types and allows differences in edge values. The tuning parameter λ controls the degree of sparsity of the estimated gene networks. If the resulting networks are expected to be very sparse, a large value of λ is preferred. The tuning parameter α $(0 < \alpha < 1)$ reflects the difference of sparsity level between the common and unique components. If the resulting networks are expected to be very similar across different subpopulations, a small value of α is preferred. Our method for selecting the tuning parameters is discussed in Section II-F.

Note that the network structures captured by different data types may be distinct due to collection and technical bias. To learn the variations of network structures across data types, a sparse group lasso penalty can be used, $P_s(\{\Omega\}) = \alpha G\{\lambda \sum_{i\neq j} \sqrt{\sum_{k=1}^K (R_{ij}^k)^2} + \gamma \sum_{k=1}^K \sum_{i\neq j} |R_{ij}^k|\} + (1-\alpha) \sum_{g=1}^G \{\lambda \sum_{i\neq j} \sqrt{\sum_{k=1}^K (M_{ij}^{kg})^2} + \gamma \sum_{k=1}^K \sum_{i\neq j} |M_{ij}^{kg}|\}$. That is, an ℓ_1 penalty can be added to allow changes of network structures across data types. We do not use the sparse group lasso penalty to learn the network changes across data types in this article due to the following reasons. First, even though the network structures captured by different data types may be different from each other, the true subpopulation-specific networks do not depend on data types. Our goal is to learn the networks that represent the true gene relationships in cells. Compared with the edges unique to individual data types, the edges captured by multiple data types are more likely to be true gene relationships. Therefore, we use the group lasso penalty to learn the network structures preserved across data types. Second, another tuning parameter γ will be added if the sparse group lasso penalty is used. A method that includes too many tuning parameters cannot be used easily in practice.

By substituting the penalty function (3) into (2), we develop a new graphical model for JEGN

$$\begin{aligned}\min_{\{\Omega\},\{R\},\{M\}} \; & \sum_{k=1}^K \sum_{g=1}^G n_g \left\{ \operatorname{tr}\left(S^{kg}\Omega^{kg}\right) - \log\det\left(\Omega^{kg}\right)\right\} \\ & + \lambda\alpha G \sum_{i\neq j} \sqrt{\sum_{k=1}^K \left(R_{ij}^k\right)^2} \\ & + \lambda(1-\alpha) \sum_{g=1}^G \sum_{i\neq j} \sqrt{\sum_{k=1}^K \left(M_{ij}^{kg}\right)^2} \\ \text{s.t.} \quad & \Omega^{kg} = R^k + M^{kg}, \; \Omega^{kg} \in S_{++}^p \end{aligned} \tag{4}$$

where $\{R\} = \{R^k : k = 1, \ldots, K\}$, $\{M\} = \{M^{kg} : k = 1, \ldots, K, \; g = 1, \ldots, G\}$, and $\Omega^{kg} \in S_{++}^p$ are the positive-definite constraints.

The solution to (4), $\hat{\Omega}^{kg}$ for $k = 1, \ldots, K$ and $g = 1, \ldots G$, can be used to reconstruct gene networks. For the kth data type and the gth subpopulation, two genes i and j are connected in the estimated gene network if and only if $\hat{\Omega}_{ij}^{kg} \neq 0$. Note that we apply a group lasso penalty to gene networks for different data types. For each subpopulation, the gene networks estimated from different data types have an identical network structure. Methodological comparisons with related works are provided in Sections S3.1 and S3.2 in the supplementary material.

D. Algorithm

We propose an alternating direction method of multipliers (ADMMs) algorithm [47] to solve the optimization problem (4). The augmented Lagrangian to (4) is

$$\begin{aligned} & L_\rho(\{\Omega\}, \{R\}, \{M\}, \{Q\}) \\ & = \sum_{k=1}^K \sum_{g=1}^G n_g \left\{ \operatorname{tr}\left(S^{kg}\Omega^{kg}\right) - \log\det\left(\Omega^{kg}\right)\right\} \\ & \quad + \lambda\left(\alpha G \sum_{i\neq j} \sqrt{\sum_{k=1}^K \left(R_{ij}^k\right)^2} + (1-\alpha) \sum_{g=1}^G \sum_{i\neq j} \sqrt{\sum_{k=1}^K \left(M_{ij}^{kg}\right)^2}\right) \\ & \quad + \sum_{k=1}^K \sum_{g=1}^G \left\langle Q^{kg}, \Omega^{kg} - \left(R^k + M^{kg}\right)\right\rangle \\ & \quad + \frac{\rho}{2} \sum_{k=1}^K \sum_{g=1}^G \left\| \Omega^{kg} - \left(R^k + M^{kg}\right)\right\|_F^2 \end{aligned} \tag{5}$$

where $\{Q\} = \{Q^{kg} : k = 1, \ldots, K, g = 1, \ldots, G\}$ are dual variables (Lagrange multipliers), and ρ is the penalty parameter.

The ADMM algorithm updates each primal variable $(\{\Omega\}, \{R\}, \{M\})$ in turn while holding all other variables fixed. Then, the dual variables $(\{Q\})$ are updated using a dual-ascent rule. The complete ADMM algorithm for problem (4) is presented in Algorithm 1, in which the operator Expand [26], [36] is given by

$$\begin{aligned} \text{Expand}(A, \rho, n) &= \arg\min_{\Omega\in S_{++}^p} -n\log\det(\Omega) + \frac{\rho}{2}\|\Omega - A\|_F^2 \\ &= \frac{1}{2}U\left(D + \sqrt{D^2 + \frac{4n}{\rho}I}\right)U^T \end{aligned} \tag{6}$$

where UDU^T is the eigenvalue decomposition of a symmetric matrix A. Here, the second equality is given in [48].

The operator Γ is given by

$$\Gamma\Big(\big\{A^1,\ldots,A^K\big\},\lambda\Big)=\underset{\{X^1,\ldots,X^K\}}{\arg\min}\ \frac{1}{2}\sum_{k=1}^{K}\Big\|X^k-A^k\Big\|_F^2+\lambda\sum_{i\neq j}\sqrt{\sum_{k=1}^{K}\Big(X_{ij}^k\Big)^2} \tag{7}$$

which takes a simply exactly closed-form solution [26]. We accelerate this algorithm by adaptively changing ρ in iterations [49]. We set $\rho=0.1$ and increase it in the next iteration as $\rho\cdot\rho^{\text{incr}}$, and we set $\rho^{\text{incr}}=1.2$ in our algorithm. We consider the process converges if the following two conditions are satisfied:

$$\sum_{k=1}^{K}\sum_{g=1}^{G}\sum_{ij}\left|\Big(\Omega_{ij}^{kg}\Big)^{(t)}-\Big(\Omega_{ij}^{kg}\Big)^{(t-1)}\right|\leq 10^{-5}\sum_{k=1}^{K}\sum_{g=1}^{G}\sum_{ij}\left|\Big(\Omega_{ij}^{kg}\Big)^{(t)}\right| \tag{8}$$

$$\sum_{k=1}^{K}\sum_{g=1}^{G}\sum_{ij}\left|\Big(\Omega_{ij}^{kg}\Big)^{(t)}-\Big(\Big(R_{ij}^{k}\Big)^{(t)}+\Big(M_{ij}^{kg}\Big)^{(t)}\Big)\right|\leq 10^{-5}\sum_{k=1}^{K}\sum_{g=1}^{G}\sum_{ij}\left|\Big(\Omega_{ij}^{kg}\Big)^{(t)}\right| \tag{9}$$

where $(\Omega^{kg})^{(t)}$, $(R^k)^{(t)}$, and $(M^{kg})^{(t)}$ denote the estimates of parameters at the t-th iteration. Please refer to Section S3.3 in the supplementary material for details of our algorithm. The complete procedure for inferring gene networks from gene expression data is presented in Section S3.4 in the supplementary material.

We now discuss the computational complexity of the ADMM algorithm. For each ADMM iteration, the operator Expand for updating Ω^{kg} involves the eigenvalue decomposition of a symmetric matrix (6), which requires $O(p^3)$ operations. Therefore, updating the KG matrices $\{\Omega^{kg}\}$ requires $O(KGp^3)$ operations. The complexity of the operator Γ (7) is $O(Kp^2)$. Updating $\{R^k\}$ and $\{M^{kg}\}$ involves $O(Kp^2)$ and $O(KGp^2)$ operations, respectively. Thus, each ADMM iteration requires $O(KGp^3)$ operations, and total complexity of our algorithm is $O(TKGp^3)$, where T is the number of iterations.

E. Nonparanormal JEGN

Although our JEGN is derived for Gaussian data, both the theoretical and simulation results in several recent works have shown that the Gaussian graphical models can be easily extended to fit nonparanormal data [23], [24], [29], [37], [46]. A random vector is said to belong to a nonparanormal family if there exist a set of unknown monotone univariate transformations such that the transformed data follow a multivariate Gaussian distribution. To learn gene networks from the non-Gaussian data, we extend JEGN following the method of the nonparanormal graphical models [23], [24], [29], [37]. We first compute a rank correlation (e.g., Kendall's tau correlation) to obtain a nonparametric sample estimate of correlation matrix,

Algorithm 1 ADMM Algorithm for JEGN (4)

- **Inputs:** Sample covariance matrices $\{S^{kg}: k=1,\ldots,K,\ \ g=1,\ldots,G\}$, tuning parameters λ and α, sample sizes n_g for $g=1,\ldots G$.
- **Output:** Estimated precision matrices $\{\hat{\Omega}^{kg}: k=1,\ldots,K,\ \ g=1,\ldots,G\}$, estimated common components $\{\hat{R}^k: k=1,\ldots,K\}$, and estimated subpopulation-unique components $\{\hat{M}^{kg}: k=1,\ldots,K,\ \ g=1,\ldots,G\}$.

1: Initialize: $\Omega^{kg}=I$ for $k=1,\ldots,K$ and $g=1,\ldots,G$, $R^k=I$ for $k=1,\ldots,K$, $M^{kg}=I$ for $k=1,\ldots,K$ and $g=1,\ldots,G$, $Q^{kg}=0$ for $k=1,\ldots,K$ and $g=1,\ldots,G$, $\rho=0.1$, $\rho^{\text{incr}}=1.2$, $\rho^{\max}=10^{10}$.
2: **while** not converged **do**
3: $\Omega^{kg}\leftarrow Expand\Big(R^k+M^{kg}-\frac{1}{\rho}\Big(n_gS^{kg}+Q^{kg}\Big),\rho,\ n_g\Big)$;
4: $A^k\leftarrow\frac{1}{G}\sum_{g=1}^{G}\Big(\Omega^{kg}-M^{kg}+\frac{Q^{kg}}{\rho}\Big)$;
5: $\{R^1,\ldots,R^K\}\leftarrow\Gamma\Big(\{A^1,\ldots,A^K\},\frac{\lambda\alpha}{\rho}\Big)$;
6: $B^{kg}\leftarrow\Omega^{kg}-R^k+\frac{Q^{kg}}{\rho}$;
7: $\{M^{1g},\ldots,M^{Kg}\}\leftarrow\Gamma\Big(\{B^{1g},\ldots,B^{Kg}\},\frac{\lambda(1-\alpha)}{\rho}\Big)$;
8: $Q^{kg}\leftarrow Q^{kg}+\rho\Big(\Omega^{kg}-\Big(R^k+M^{kg}\Big)\Big)$;
9: Check the convergence condition;
10: $\rho\leftarrow\min(\rho\cdot\rho_{\text{incr}},\rho_{\max})$;
11: **end while**
12: **return** $\hat{\Omega}^{kg}=R^k+M^{kg}$, $\hat{R}^k=R^k$, $\hat{M}^{kg}=M^{kg}$.

$\hat{\tau}^{kg}$, for each subpopulation and each data type

$$\hat{\tau}_{ij}^{kg}=\frac{2}{n_g(n_g-1)}\sum_{1\leq\ell<\ell'\leq n_g}\operatorname{sign}\Big(\Big(x_{\ell i}^{kg}-x_{\ell' i}^{kg}\Big)\Big(x_{\ell j}^{kg}-x_{\ell' j}^{kg}\Big)\Big). \tag{10}$$

Then, a bridge function is used to connect Kendall's tau correlation to Pearson's correlation

$$\hat{R}_{ij}^{kg}=\begin{cases}\sin\Big(\frac{\pi}{2}\hat{\tau}_{ij}^{kg}\Big), & i\neq j\\ 1, & i=j.\end{cases} \tag{11}$$

To make sure that the sample correlation matrices are positive semidefinite, we project the estimated correlation matrices into the cone of positive semidefinite matrices by solving the following optimization problem:

$$\hat{R}_p^{kg}=\arg\min_{R\succeq 0}\Big\|\hat{R}^{kg}-R\Big\|_\infty. \tag{12}$$

Finally, $\hat{R}_p^{kg}$ is used to replace sample covariance matrices S^{kg} in (4). The remaining computations are the same as those used for fitting Gaussian data (Section S3.4 in the supplementary material).

F. Selection of Tuning Parameters

To apply JEGN, we need to determine the tuning parameters λ and α. We use λ to control the level of sparsity of the resulting gene networks. Large λ tends to yield sparse networks, and small λ tends to yield dense networks. We use α to control the difference of sparsity between the common and unique subnetworks. Large α yields sparse common subnetworks and dense subpopulation-unique subnetworks and, therefore, the estimated gene networks across different subpopulations tend to be quite different from each other. In most applications,

gene network reconstruction is a part of exploratory data analysis. Tuning parameter selection should be guided by practical considerations, such as network interpretability, stability, and sparsity [26]. To produce interpretable results, we set α to be a fixed value (α_{opt}), which helps to explore the subpopulation-specific network variations. Then, the tuning parameter λ is selected using stability selection [50], which simultaneously considers the sparsity and stability of the resulting networks.

For a given value of α_{opt}, we choose λ so as to use the least amount of penalty that simultaneously makes the resulting gene network sparse and stable. We draw S random sample sets $\mathcal{D}_1, \ldots, \mathcal{D}_S$ from the $n = \sum_{g=1}^{G} n_g$ subjects, each with size $0.8n$. For now, we choose λ from a given vector of regularization parameter, Λ. We estimate KG gene networks $\{\hat{\Omega}_s^{kg}(\lambda, \alpha_{\text{opt}})\}$ for each $\mathcal{D}_s$ and each λ from Λ. Stability score is computed as

$$\text{Stab}(\lambda) = \sum_{k=1}^{K}\sum_{g=1}^{G}\left(\sum_{i<j}\bar{B}_{ij}^{kg}(\lambda)\left(1-\bar{B}_{ij}^{kg}(\lambda)\right)\Big/\binom{p}{2}\right)\Big/KG \tag{13}$$

where $\bar{B}_{ij}^{kg}(\lambda) = (1/S)\sum_{s=1}^{S}\mathbf{1}(\hat{\Omega}_{s,ij}^{kg}(\lambda, \alpha_{\text{opt}}) \neq 0)$ is the frequency of a pair of genes (i, j) being connected in the reconstructed networks. The optimal value for λ is chosen according to the following criterion:

$$\lambda_{\text{opt}} = \arg\min_{\gamma\in\Lambda}\left\{\max_{\lambda\geq\gamma}\text{Stab}(\lambda) \leq \beta\right\}. \tag{14}$$

In this article, we set the number of random sample sets $S = 20$ and the stability parameter $\beta = 0.01$.

III. Experiments and Results

In this section, we first evaluate the performance of the proposed method using simulated data, and then apply it to the real-world data to identify gene network variations across different subtypes of breast cancer.

A. Simulation Study

1) Data Generation: We consider $G = 4$ subpopulations and $K = 3$ data types. We generate $4 \times 3 = 12$ networks (either Erdős-Rényi or scale-free) with $p = 100$ genes for the 4 subpopulations and 3 data types. To model the similarities and differences among subpopulations, Ω^{kg} is generated as $\Omega^{kg} = R^k + M^{kg} + \delta^k I$, where R^k represents the subnetwork common to all subpopulations, M^{kg} represents the subnetwork unique to the gth subpopulation, and δ^k is selected large enough to guarantee the positive definiteness of Ω^{kg}. For each subpopulation, the edges are preserved across all data types but the edge values are different. That is, $\Omega^{1g}, \ldots, \Omega^{Kg}$ have an identical pattern of nonzero elements but different values. To make a fair comparison with the methods which have a Gaussian assumption, gene expression measurements are simulated using a multivariate Gaussian distribution with precision matrix Ω^{kg}. Details are presented in Section S3.5 in the supplementary material. Let τ be the ratio of the number of edges unique to individual subpopulations to the number of edges common to all subpopulations. A larger value of τ means that the generated subpopulation-specific networks are more different from each other. Let n_g be the sample size. We set $\tau = 0.25, 0.5, 1$, and $n_g = 25, 50, 100$. The performance is evaluated with every combination of τ and n_g values.

2) Simulation Results: We compare JEGN with two popular joint graphical models: 1) fused graphical lasso (FGL) and 2) group graphical lasso (GGL) [26]. These two methods simultaneously estimate multiple networks which share certain characteristics. FGL uses a fused lasso penalty to encourage similar edge values across the estimated networks, and GGL uses a group lasso penalty to encourage a shared pattern of network structures. When applying FGL, networks are estimated for each data type separately. That is, for a data type, we jointly fit G networks for the G subpopulations using FGL such that the networks have similar edge values across different subpopulations. When applying GGL, networks are estimated for each subpopulation separately. For a subpopulation, K networks corresponding to different data types are estimated jointly using GGL such that the networks have identical structures. The JGL function with ''penalty = fused'' and ''penalty = group'' from the R-package JGL is used to implement FGL and GGL, respectively. The tuning parameters of GGL are reparameterized as suggested by [26], $\omega_1 = \lambda_1 + (1/\sqrt{2})\lambda_2$ and $\omega_2 = (1/\sqrt{2})\lambda_2/(\lambda_1 + (1/\sqrt{2})\lambda_2)$. More details of the compared methods are presented in Section S3.1 in the supplementary material. We run the three methods with a variety of values of tuning parameters. For JEGN, the tuning parameter α that controls the level of similarity between subpopulation-specific networks is set to $\alpha = 0.15,\ 0.25,\ 0.35,\ 0.45$. For FGL and GGL, tuning parameters (λ_2 for FGL, and ω_2 for GGL) that control network similarity are set to the values used by Danaher *et al.* [26], $\lambda_2 = 0.025,\ 0.05,\ 0.075,\ 0.1$ and $\omega_2 = 1,\ 0.7,\ 0.5,\ 0.15$. For all the three methods, the parameters (λ for JEGN, λ_1 for FGL, and ω_1 for GGL) that control the level of sparsity of the estimated networks are set to a total of 10 possible values equally spaced in log scale between 0.1 and 2.

We evaluate the performance in terms of precision and recall. Let $\hat{\Omega}_{ij}^{kg}$ be the (i, j)th entry of a given estimator $\hat{\Omega}^{kg}$, and let Ω_{ij}^{kg} be the (i, j)th entry of the true Ω^{kg}, then precision and recall are defined as

$$\text{precision} = \frac{\sum_{k=1}^{K}\sum_{g=1}^{G}\sum_{i<j}1\left\{\hat{\Omega}_{ij}^{kg}\neq 0,\ \Omega_{ij}^{kg}\neq 0\right\}}{\sum_{k=1}^{K}\sum_{g=1}^{G}\sum_{i<j}1\left\{\hat{\Omega}_{ij}^{kg}\neq 0\right\}}$$

$$\text{recall} = \frac{\sum_{k=1}^{K}\sum_{g=1}^{G}\sum_{i<j}1\left\{\hat{\Omega}_{ij}^{kg}\neq 0,\ \Omega_{ij}^{kg}\neq 0\right\}}{\sum_{k=1}^{K}\sum_{g=1}^{G}\sum_{i<j}1\left\{\Omega_{ij}^{kg}\neq 0\right\}}.$$

Here, $1\{\cdot\}$ is an indicator function. The precision–recall curve is used to quantify the performance as a function of the tuning parameter that controls the sparsity level of the estimated networks. Besides the precision–recall curve, we also plot the sum of errors in edge values against the recall values to assess

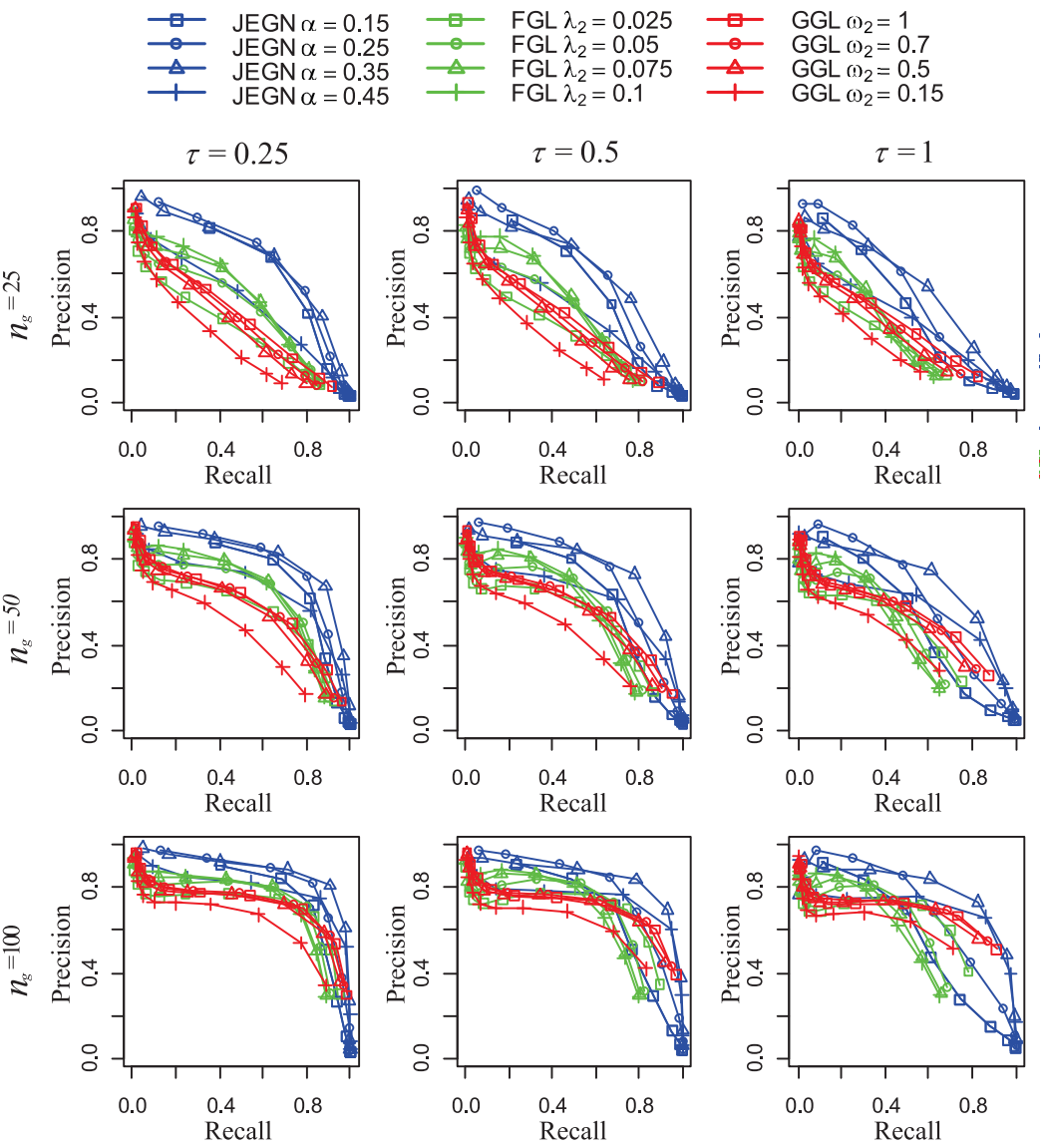


Fig. 2. Simulation results on Erdős–Rényi networks. Each row corresponds to the sample size (n_g), and each column to the ratio of the number of population-unique edges to the number of common edges (τ). For each plot, the precision values are plotted against the recall values. In each plot, each colored curve corresponds to a fixed value of the similarity tuning parameter (α for JEGN, ω_2 for GGL, and λ_2 for FGL), as the sparsity tuning parameter (λ for JEGN, ω_1 for GGL, and λ_1 for FGL) is varied. Results are averaged over 50 random generations of the data.

the performance, where the errors between the estimated edge values and the true edge values are defined as

$$\text{error} = \sum_{k=1}^{K} \sum_{g=1}^{G} \sqrt{\sum_{i<j} \left(\hat{\Omega}_{ij}^{kg} - \Omega_{ij}^{kg} \right)^2}.$$

Fig. 2 shows the results on Erdős-Rényi networks averaged over 50 random generations of the data. Each row represents the sample size (n_g), and each column represents the ratio of the number of population-unique edges to the number of common edges (τ). Within each plot, each colored curve illustrates the results obtained using a fixed value of the tuning parameter that controls the level of similarity (α for JEGN, λ_2 for FGL, and ω_2 for GGL), as the tuning parameter that controls the level of sparsity (λ for JEGN, λ_1 for FGL, and ω_1 for GGL) is varied. For a suitable range of the similarity tuning parameters, JEGN performs better than FGL and GGL. Since FGL only borrows strength across different subpopulations and does not consider the shared network structures across different data types, a comparison with FGL can show the gain of JEGN due to the consideration of common information provided by different data types. The advantage of JEGN over GGL is partially due to the fact that GGL only borrows strength across different data types while JEGN can exploit the similarities among data types as well as the similarities among subpopulations. The results on scale-free networks (Fig. 3) also show that JEGN outperforms FGL and GGL. Figs. S1 and S2 in the supplementary material plot the sum of errors in edge values against the recall values on the Erdős–Rényi networks and scale-free networks, respectively. The results indicate that JEGN obtain a lower error than FGL and GGL at the same recall levels.

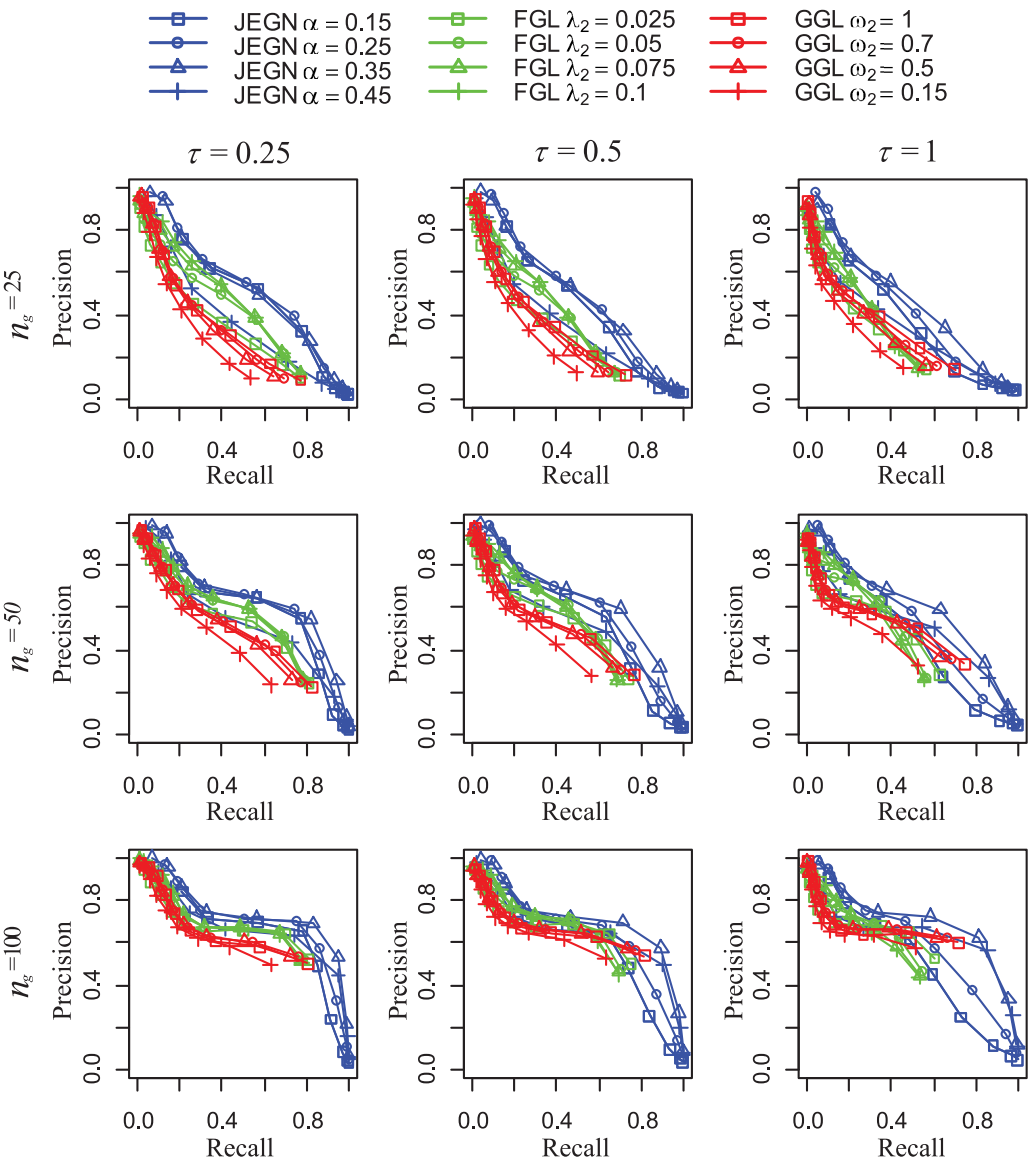


Fig. 3. Simulation results on scale-free networks. Each row corresponds to the sample size (n_g), and each column to the ratio of the number of population-unique edges to the number of common edges (τ). For each plot, the precision values are plotted against the recall values. In each plot, each colored curve corresponds to a fixed value of the similarity tuning parameter (α for JEGN, ω_2 for GGL, and λ_2 for FGL), as the sparsity tuning parameter (λ for JEGN, ω_1 for GGL, and λ_1 for FGL) is varied. Results are averaged over 50 random generations of the data.

B. TCGA Breast Cancer Application

1) Datasets: Breast cancer is a heterogeneous disease with multiple distinct molecular subtypes. Four main breast cancer subtypes have been identified based on gene expression data: 1) luminal A; 2) luminal B; 3) HER2-enriched; and 4) P basal-like [25]. The associations between breast cancer subtypes and clinical, the genomic and proteomic features have been comprehensively analyzed [25]. However, there are few works studying the association between the breast cancer subtype classification and gene network rewiring. To understand the characterization of breast cancer subtypes from the perspective of network dynamics, we would like to infer the cancer subtype-specific gene networks and learn the network variations across different cancer subtypes. We consider two gene expression datasets that measure gene activities using different data types: 1) microarray (Agilent G450) and 2) RNA sequencing. We download the two datasets (level 3) from TCGA (version: May 6 2017). Gene activity measurements of 489 primary breast cancers on 15 718 genes are available for the two data types. We obtain the mRNA-expression cancer subtype classification information of breast cancers from [25]. Out of the 489 breast cancers, there are 214 luminal A cancers, 117 luminal B cancers, 55 HER2-enriched cancers, and 90 basal-like cancers. We perform a pathway-based analysis to explore the pathway variations across different cancer subtypes. Only the 139 genes that belong to both the breast cancer pathway (hsa05224) from the kyoto encyclopedia of genes and genomes (KEGGs) database [51] and the considered gene expression datasets are analyzed.

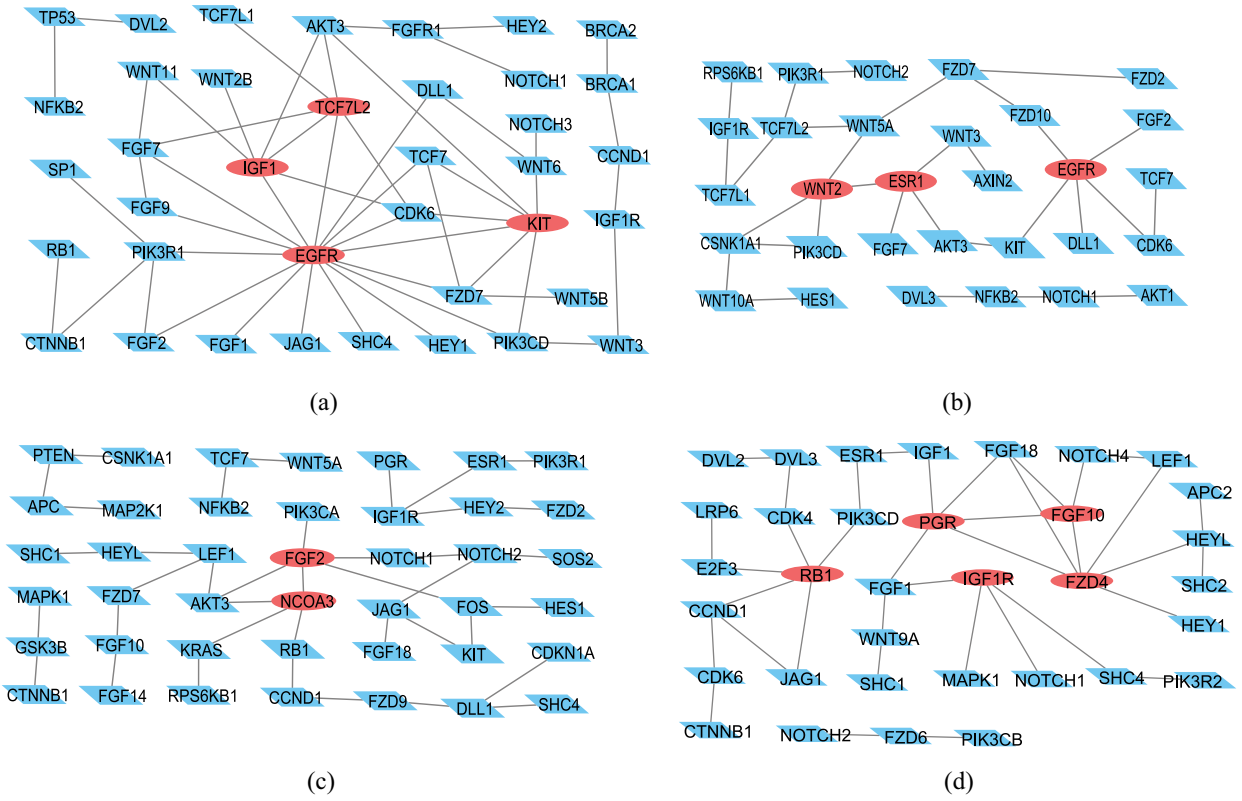


Fig. 4. Cancer subtype-unique networks estimated by JEGN. Each network consists of edges that are unique to individual subtype. In each network, the hub nodes which are analyzed in the text are highlight in red. Only connected components with more than three nodes are displayed. (a) Luminal A. (b) Luminal B. (c) HER2-enriched. (d) Basal-like.

2) Cancer Subtype-Specific Gene Network Variations: We apply JEGN to the eight datasets corresponding to the two data types and the four cancer subtypes. Since the expression values of genes typically do not completely follow the Gaussian distributions, we use nonparanormal JEGN (Section II-E). In other words, we first use the adjusted Kendall's tau rank correlation to obtain a nonparametric sample estimate of correlation matrices, and then compute an estimator of gene networks from the rank-based sample estimate of correlation matrices using JEGN (Section S3.4 in the supplementary material). To avoid disparate levels of sparsity between the cancer subtypes, each cancer subtype is equally weighted instead of by the sample size in (4) following the method of [26]. Since the goal of this analysis is to understand the gene network variations across different cancer subtypes, we aim to identify certain edges common to all cancer subtypes and certain edges unique to individual cancer subtypes. To produce interpretable results, we fix the tuning parameter that controls the similarities among cancer subtype-specific networks at $\alpha = 0.4$. We select a value of λ ranging from 0.1 to 2 using the stability selection criterion presented in Section II-F. We run JEGN with the selected tuning parameters $\alpha = 0.4$ and $\lambda = 0.95$.

The estimated cancer subtype-specific gene networks are presented in Figs. S3–S6 in the supplementary material. Since JEGN uses a group lasso penalty across different data types, the networks estimated from different data types have the identical structures for all cancer subtypes. It is noticeable that most edges are common to all cancer subtypes, and only a small number of edges are unique to individual cancer subtypes (Table S1 in the supplementary material). This confirms that the cancer subtype-specific gene networks are similar to each other since they belong to the same cancer type and also have important differences stemming from the intertumor heterogeneity.

We gain more insights into the cancer subtype-unique subnetworks (Fig. 4). Here, a cancer subtype-unique network is defined as a network that consists of edges only connected in the considered cancer subtype (but not connected in the other three cancer subtypes). Since the hub nodes often play key roles in the underlying network [52], we are interested in the biological significance of hub genes in the cancer subtype-unique subnetworks. A node is considered as a hub if it involves at least 5% edges in the total network. For the luminal A subtype, we consider genes that have at least 5 connected partners as hubs. For the other three cancer subtypes, genes with degree not less than 4 are considered as hubs. We find that most hub genes are the characteristic signatures of breast cancer subtypes (Fig. 4 and Table S2 in the supplementary material). For example, the epidermal growth factor receptor (EGFR) is identified as a hub in two cancer subtype-unique networks (luminal A and luminal B). EGFR overexpression is associated with the development of a number of cancers including breast cancer, considered in this article. EGFR is also a well-known signature of molecular subtypes and a useful therapeutic target [25], [53]. IGF1, KIT, and TCF7L2 are also identified as the hub nodes in the luminal A subtype.

Insulin-like growth factor 1 (IGF1) is a ligand of the type I insulin-like growth factor receptor (IGF1R) that is also identified as a hub in the basal-like subtype. IGF1–IGF1R pathway plays an important role in regulating the breast cancer initiation and progression [54] and is associated with cancer phenotypes and subtypes [55]. IGF1R has been considered as a breast cancer therapeutic target [56], and IGF1R expression involves with response to chemotherapy [57]. KIT (CD117) is a well-known proto-oncogene. Overexpression or mutations of KIT can lead to a variety of cancers. The expression level of KIT is varied across breast cancer subtypes [58]. TCF7L2 is a transcription factor that regulates the transcription of several genes to carry out multiple functions. Genetic alterations in TCF7L2 has found to be associated with breast cancer risk [59]. ESR1 and WNT2 are also identified as hub nodes in the luminal B cancers. ESR1 is one form of the estrogen receptor (ER) and is differentially expressed in breast cancer. Tian *et al.* [29] has shown that ESR1 is a key hub gene driving network rewiring between different breast cancer subtypes. WNT2 is a member of WNT gene family. WNT2 is implicated in various cancers and has been reported to be highly expressed in human breast cancer [60]. FGF2 and NCOA3 are the hub nodes in the subnetwork unique to the HER2-enriched subtype. FGF2 and FGF10, which is identified as a hub in the basal-like subtype, are two members of the fibroblast growth factor family. Giulianelli *et al.* [61] have found that FGF2 can promote breast tumor growth. FGF10 contributes to the progression of multiple human cancers including breast cancer [62]. Nuclear receptor coactivator 3 (NCOA3), which is also called amplified in breast 1 (AIB1), is a co-activator of ER. High expression of NCOA3 is found in most human breast tumors and is associated with drug resistance, poor prognosis, and survival rate [63]. Besides FGF10 and IGF1R mentioned above, FZD4, PGR, and RB1 are also identified as the hub genes in the basal-like subtype. FZD4 has been reported to be associated breast cancer subtypes [64]. Progesterone receptor (PGR) is a well-known ER-regulated gene that is differentially expressed in different breast cancer subtypes. The expression level of PGR is associated with tumor grade, ER expression, and Nottingham prognostic index [65]. Retinoblastoma tumor suppressor (RB1) is an important regulator of processes involved in cancer development. RB1 loss is believed to occur in breast cancer. The mutation rate, copy number status, mRNA, and protein expression of RB1 are significantly changed across different breast subtypes [25].

A limma analysis [66] is performed to calculate whether these hub genes are differentially expressed across different subtypes. For both microarray and RNA-seq datasets, all of the 13 hub genes are differentially expressed (adjusted p-values < 0.05, Table S2 in the supplementary material). These results suggest that JEGN can identify certain differentially expressed genes even though we only consider the changes of network structures.

We also run JEGN on each data type separately to show the effect of integrating different data types. Experimental results are presented in Section S3.6 in the supplementary material. We also provide additional comparisons of JEGN with FGL and GGL in Section S3.7 in the supplementary material.

IV. Discussion and Conclusion

In this article, we have developed a new graphical model for JEGN. The novelty of our method lies in three aspects.

1) Both the similarities among subpopulations and the similarities among data types are exploited. Since the two types of similarities have different properties, two different approaches are used to analyze them. First, each subpopulation-specific network is decomposed as a sum of common and unique components. The similarities among subpopulations can be captured by the common components. Second, a group lasso penalty is applied to gene networks corresponding to different data types. The similarities among data types can be captured by the common structures preserved across data types. Previous joint graphical models can only learn one type of network similarities. If these joint graphical models are simply applied to gene expression data across multiple subpopulations and data types, the heterogeneity in the two types of similarities cannot be exploited.
2) Our model is able to reveal the subpopulation-specific network variations automatically. Due to the representation of each subpopulation-specific network as a sum of common and unique components, the edges unique to individual subpopulation can be quantified by the unique components. The changes of the gene networks across different subpopulations can be learned explicitly by comparing these subpopulation-unique subnetworks. Most previous methods do not use this decomposition, therefore, they cannot exploit the gene network rewiring directly [27], [39].
3) Since most real data do not follow Gaussian distribution, we have extended JEGN to nonparanormal JEGN following the method of the nonparanormal graphical models. According to this extension, the non-Gaussian data that follow the nonparanormal distribution can be fitted well. Most existing joint graphical models do not consider this useful extension.

We have applied our method to infer gene networks for different subtypes of breast cancer. Gene network variations across subtypes are analyzed. Hub nodes in the estimated subtype-unique subnetworks rediscover well-known genes associated with subtype classification and provide interesting predictions. The R-package of our method is available online and free of charge. Interested readers can use our method to analyze gene network rewiring across subtypes of other cancers.v One can also consider the subpopulations as different cancer types, then a pan-cancer analysis can be performed [37]. Gene interactions common to multiple cancers and gene interactions unique to individual cancers can be identified. The similarities and differences among different cancer types can be exploited in terms of the changes of gene networks as well.

The decomposition of Ω^{kg} as a sum of $R^k + M^{kg}$ might not be unique if we do not impose constraints on R^k and M^{kg}. Our model might be unidentifiable due to this decomposition. In order to obtain unique decomposition, we can make constraints following the method of [27], for example, $\sum_{g=1}^{G} M^{kg} = 0$ for

$k = 1, \ldots, K$. In this article, we do not consider these constraints for the following reasons. First, more dual variables will be introduced and the ADMM algorithm will become more complex if these constraints are incorporated. Second, Wang *et al.* [39] have used a similar decomposition and they have shown that their penalty term ensures that there is a unique solution under certain conditions. Similarly, our penalty term (3) ensures that unique estimation of R^k and M^{kg} can be obtained when $\alpha \neq 0.5$.

Besides graphical models, (conditional) mutual information-based methods have also been used widely to reconstruct gene networks since they can quantify the nonlinear relationships among genes [6], [8], [67]. However, these mutual information-based methods are developed to reconstruct a single-gene network from a single dataset. They may be suboptimal since they do not integrate information across different subpopulations and data types. Compared with the mutual information-based methods, our graphical model-based method can exploit both the similarities among subpopulations and the similarities among data types via a regularization approach. However, the graphical model-based methods can only measure linear relationships and will miss nonlinear relationships. It would be beneficial to make full use of the advantages of mutual information and graphical models to develop a new method that can simultaneously learn the nonlinear relationships and exploit the similarities among multiple networks.

The convergence rate and model selection consistency have been analyzed in two previous studies that also decomposed the networks into common and unique parts [27], [39]. Our model is different from these two works in the following aspects: 1) our method is based on the penalized Gaussian likelihood framework, while the two previous works are based on the constrained ℓ_1 minimization and 2) besides the common-unique decomposition used by all the three models to reveal the network variations across subpopulations, our model also uses the group lasso penalty to exploit the common network structures shared across different data types. It is not easy to conduct the theoretical analysis of our estimator in a similar way to [27] and [39]. We leave the theoretical analysis of our estimator as future work. When solving the optimization problem (4) using the ADMM algorithm, we divide the primal variables into three groups ($\{\Omega\}$, $\{R\}$, $\{M\}$) which are easy to solve separately. However, the convergence of the ADMM algorithm with three groups of primal variables cannot be guaranteed in all situations. Note that our optimization problem (4) can be rewritten as a consensus problem with two groups of primal variables following the method of [68], so that the global convergence can be guaranteed. In this article, we do not consider this alternative approach since previous studies have observed that the ADMM algorithms with more than two groups of variables often converge faster than other variants [69]. We will investigate the performance of two types of ADMM algorithms in the future.

In this article, we mainly study the identification of the functional dependencies among genes. Apart from genes, there are many other biological measurements (e.g., mutation, DNA methylation, microRNA, and long noncoding RNAs) which are important for understanding the mechanisms of cancers. The changes of functional relationships among these measurements may be associated with the classification of cancers. These different types of measurements often follow different probability distributions. For example, gene mutation is usually modeled using a Bernoulli distribution, which does not belong to the nonparanormal family. Since our method is based on the nonparanormal model, it cannot analyze these different types of measurements directly [70]. How to extend our method to fit these different types of measurements is an interesting problem. In addition, our method assumes subpopulation memberships of subjects are known. It cannot be applied in settings with complex but unknown population structures. In the future, we will use our method in the Gaussian mixture-model-based clustering to simultaneously reveal the population memberships and learn the subpopulation-specific network structures [71]–[73].

REFERENCES

[1] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: A network-based approach to human disease," *Nat. Rev. Genet.*, vol. 12, no. 1, pp. 56–68, 2011.

[2] D. Marbach *et al.*, "Wisdom of crowds for robust gene network inference," *Nat. Methods*, vol. 9, no. 8, pp. 796–804, 2012.

[3] S. M. Hill *et al.*, "Inferring causal molecular networks: Empirical assessment through a community-based effort," *Nat. Methods*, vol. 13, no. 4, pp. 310–318, 2016.

[4] Z.-H. You, M. Zhou, X. Luo, and S. Li, "Highly efficient framework for predicting interactions between proteins," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 731–743, Mar. 2017.

[5] Y. Wang, T. Joshi, X.-S. Zhang, D. Xu, and L. Chen, "Inferring gene regulatory networks from multiple microarray datasets," *Bioinformatics*, vol. 22, no. 19, pp. 2413–2420, 2006.

[6] X. Zhang *et al.*, "Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information," *Bioinformatics*, vol. 28, no. 1, pp. 98–104, 2012.

[7] Z. Wang, W. Xu, F. A. S. Lucas, and Y. Liu, "Incorporating prior knowledge into gene network study," *Bioinformatics*, vol. 29, no. 20, pp. 2633–2640, 2013.

[8] J. Zhao, Y. Zhou, X. Zhang, and L. Chen, "Part mutual information for quantifying direct associations in networks," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 18, pp. 5130–5135, 2016.

[9] X. Liu, Y. Wang, H. Ji, K. Aihara, and L. Chen, "Personalized characterization of diseases using sample-specific networks," *Nucl. Acids Res.*, vol. 44, no. 22, p. e164, 2016.

[10] Y. Şenbabaoğlu *et al.*, "A multi-method approach for proteomic network inference in 11 human cancers," *PLoS Comput. Biol.*, vol. 12, no. 2, 2016, Art. no. e1004765.

[11] Y. Zhang, Z. Ou-Yang, and H. Zhao, "A statistical framework for data integration through graphical models with application to cancer genomics," *Ann. Appl. Stat.*, vol. 11, no. 1, pp. 161–184, 2017.

[12] L. Ou-Yang, X. F. Zhang, M. Wu, and X.-L. Li, "Node-based learning of differential networks from multi-platform gene expression data," *Methods*, vol. 129, pp. 41–49, Oct. 2017.

[13] L. Ou-Yang, H. Yan, and X.-F. Zhang, "Identifying differential networks based on multi-platform gene expression data," *Mol. Biosyst.*, vol. 13, no. 1, pp. 183–192, 2017.

[14] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[15] M. Yuan and Y. Lin, "Model selection and estimation in the Gaussian graphical model," *Biometrika*, vol. 94, no. 1, pp. 19–35, 2007.

[16] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[17] S. D. Zhao, T. T. Cai, and H. Li, "Direct estimation of differential networks," *Biometrika*, vol. 101, no. 2, pp. 253–268, 2014.

[18] X. F. Zhang, L. Ou-Yang, S. Yang, X. Hu, and H. Yan, "DiffGraph: An R package for identifying gene network rewiring using differential graphical models," *Bioinformatics*, vol. 34, no. 9, pp. 1571–1573, Dec. 2018.

[19] X.-F. Zhang, L. Ou-Yang, S. Yang, X. Hu, and H. Yan, "DiffNetFDR: Differential network analysis with false discovery rate control," *Bioinformatics*, vol. 35, no. 17, pp. 3184–3186, Jan. 2019.

[20] O. Banerjee, L. E. Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data," *J. Mach. Learn. Res.*, vol. 9, pp. 485–516, Mar. 2008.

[21] O. Dalal and B. Rajaratnam, "Sparse Gaussian graphical model estimation via alternating minimization," *Biometrika*, vol. 104, no. 2, pp. 379–395, 2017.

[22] H. Liu, J. Lafferty, and L. Wasserman, "The nonparanormal: Semiparametric estimation of high dimensional undirected graphs," *J. Mach. Learn. Res.*, vol. 10, no. 1, pp. 2295–2328, 2009.

[23] H. Liu, F. Han, M. Yuan, J. Lafferty, and L. Wasserman, "High-dimensional semiparametric Gaussian copula graphical models," *Ann. Stat.*, vol. 40, no. 4, pp. 2293–2326, 2012.

[24] L. Xue and H. Zou, "Regularized rank-based estimation of high-dimensional nonparanormal graphical models," *Ann. Stat.*, vol. 40, no. 5, pp. 2541–2571, 2012.

[25] The Cancer Genome Atlas Research Network, "Comprehensive molecular portraits of human breast tumors," *Nature*, vol. 490, no. 7418, pp. 61–70, 2012.

[26] P. Danaher, P. Wang, and D. M. Witten, "The joint graphical lasso for inverse covariance estimation across multiple classes," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 76, no. 2, pp. 373–397, 2014.

[27] W. Lee and Y. Liu, "Joint estimation of multiple precision matrices with common structures," *J. Mach. Learn. Res.*, vol. 16, no. 31, pp. 1035–1062, 2015.

[28] M. J. Ha, V. Baladandayuthapani, and K.-A. Do, "DINGO: Differential network analysis in genomics," *Bioinformatics*, vol. 31, no. 21, pp. 3413–3420, 2015.

[29] D. Tian, Q. Gu, and J. Ma, "Identifying gene regulatory network rewiring using latent differential graphical models," *Nucl. Acids Res.*, vol. 44, no. 17, p. e140, 2016.

[30] X. F. Zhang, L. Ou-Yang, and H. Yan, "Node-based differential network analysis in genomics," *Comput. Biol. Chem.*, vol. 69, pp. 194–201, Aug. 2017.

[31] T. Xu, L. Ou-Yang, X. Hu, and X.-F. Zhang, "Identifying gene network rewiring by integrating gene expression and gene network data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 6, pp. 2079–2085, Nov./Dec. 2018.

[32] L. Ou-Yang *et al.*, "Joint learning of multiple differential networks with latent variables," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3494–3506, Sep. 2019.

[33] X. F. Zhang, L. Ou-Yang, and H. Yan, "Incorporating prior information into differential network analysis using non-paranormal graphical models," *Bioinformatics*, vol. 33, no. 16, pp. 2436–2445, 2017.

[34] X.-F. Zhang, L. Ou-Yang, X.-M. Zhao, and H. Yan, "Differential network analysis from cross-platform gene expression data," *Sci. Rep.*, vol. 6, Sep. 2016, Art. no. 34112.

[35] J. Guo, E. Levina, G. Michailidis, and J. Zhu, "Joint estimation of multiple graphical models," *Biometrika*, vol. 98, no. 1, pp. 1–15, 2011.

[36] K. Mohan, P. London, M. Fazel, D. Witten, and S.-I. Lee, "Node-based learning of multiple Gaussian graphical models," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 445–488, 2014.

[37] T. Kling, P. Johansson, J. Sanchez, V. D. Marinescu, R. Jörnsten, and S. Nelander, "Efficient exploration of pan-cancer networks by generalized covariance selection and interactive Web content," *Nucl. Acids Res.*, vol. 43, no. 15, p. e98, 2015.

[38] E. Pierson, D. Koller, A. Battle, and S. Mostafavi, "Sharing and specificity of co-expression networks across 35 human tissues," *PLoS Comput. Biol.*, vol. 11, no. 5, 2015, Art. no. e1004220.

[39] B. Wang, R. Singh, and Y. Qi, "A constrained L_1 minimization approach for estimating multiple sparse Gaussian or nonparanormal graphical models," *Mach. Learn.*, vol. 106, no. 9–10, pp. 1381–1417, 2017.

[40] J. Ma and G. Michailidis, "Joint structural estimation of multiple graphical models," *J. Mach. Learn. Res.*, vol. 17, no. 166, pp. 1–48, 2016.

[41] Y. Wang, Q. Liu, and B. Yuan, "Learning latent variable Gaussian graphical model for biomolecular network with low sample complexity," *Comput. Math. Methods Med.*, vol. 2016, pp. 1–13, 2016.

[42] T. T. Cai, H. Li, W. Liu, and J. Xie, "Joint estimation of multiple high-dimensional precision matrices," *Statistica Sinica*, vol. 26, no. 2, pp. 445–464, 2016.

[43] F. Huang and S. Chen, "Joint learning of multiple sparse matrix Gaussian graphical models," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2606–2620, Nov. 2015.

[44] F. Huang, S. Chen, and S.-J. Huang, "Joint estimation of multiple conditional Gaussian graphical models," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 3034–3046, Jul. 2017.

[45] F. Huang and S. Chen, "Learning dynamic conditional Gaussian graphical models," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 4, pp. 703–716, Apr. 2018.

[46] P. Xu and Q. Gu, "Semiparametric differential graph models," in *Proc. 29th Adv. Neural Inf. Process. Syst.*, 2016, pp. 1064–1072.

[47] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[48] D. M. Witten and R. Tibshirani, "Covariance-regularized regression and classification for high dimensional problems," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 71, no. 3, pp. 615–636, 2009.

[49] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," UIUC, Urbana, IL, USA, Rep. UILU-ENG-09-2215, 2009.

[50] H. Liu, K. Roeder, and L. Wasserman, "Stability approach to regularization selection (StARS) for high dimensional graphical models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 1432–1440.

[51] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucl. Acids Res.*, vol. 28, no. 1, pp. 27–30, 2000.

[52] X.-F. Zhang, L. Ou-Yang, Y. Zhu, M.-Y. Wu, and D.-Q. Dai, "Determining minimum set of driver nodes in protein–protein interaction networks," *BMC Bioinformatics*, vol. 16, no. 1, p. 146, 2015.

[53] S. J. Schnitt, "Classification and prognosis of invasive breast cancer: From morphology to molecular taxonomy," *Mod. Pathol.*, vol. 23, no. S2, pp. S60–S64, 2010.

[54] A. M. Nagle *et al.*, "Loss of E-cadherin enhances IGF1-IGF1R pathway activation and sensitizes breast cancers to anti-IGF1R/INSR inhibitors," *Clin. Cancer Res.*, vol. 24, no. 20, pp. 5165–5177, 2018.

[55] S. M. Farabaugh, D. N. Boone, and A. V. Lee, "Role of IGF1R in breast cancer subtypes, stemness, and lineage differentiation," *Front Endocrinol.*, vol. 6, p. 59, Apr. 2015.

[56] R. C. Ekyalongo and D. Yee, "Revisiting the IGF-1R as a breast cancer target," *NPJ Precision Oncol.*, vol. 1, no. 1, p. 14, 2017.

[57] J. Yuan, Z. Yin, K. Tao, G. Wang, and J. Gao, "Function of insulin-like growth factor 1 receptor in cancer resistance to chemotherapy," *Oncol. Lett.*, vol. 15, no. 1, p. 41, 2018.

[58] S. K. K. Pillai, A. Tay, S. Nair, and C. Leong, "Triple-negative breast cancer is associated with EGFR, CK5/6 and C-kit expression in Malaysian women," *BMC Clin. Pathol.*, vol. 12, no. 1, p. 18, 2012.

[59] J. Chen, T. Yuan, M. Liu, and P. Chen, "Association between TCF7L2 gene polymorphism and cancer risk: A meta-analysis," *PLoS ONE*, vol. 8, no. 8, 2013, Art. no. e71730.

[60] T. C. Dale *et al.*, "Compartment switching of WNT-2 expression in human breast tumors," *Cancer Res.*, vol. 56, no. 19, pp. 4320–4323, 1996.

[61] S. Giulianelli *et al.*, "FGF2 induces breast cancer growth through ligand-independent activation and recruitment of ERα and PRBΔ4 isoform to MYC regulatory sequences," *Int. J. Cancer*, vol. 145, no. 7, pp. 1874–1888, 2019.

[62] N. S. Clayton and R. P. Grose, "Emerging roles of fibroblast growth factor 10 in cancer," *Front. Genet.*, vol. 9, p. 499, Oct. 2018.

[63] B. Burwinkel *et al.*, "Association of NCOA3 (AIB1) polymorphisms with breast cancer risk," *Breast Cancer Res.*, vol. 11, no. 6, pp. 2169–2174, 2005.

[64] S.-G. Pohl, N. Brook, M. Agostino, F. Arfuso, A. P. Kumar, and A. Dharmarajan, "WNT signaling in triple-negative breast cancer," *Oncogenesis*, vol. 6, no. 4, p. e310, 2017.

[65] E. Lim, C. Palmieri, and W. D. Tilley, "Renewed interest in the progesterone receptor in breast cancer," *Brit. J. Cancer*, vol. 115, no. 8, pp. 909–911, 2016.

[66] M. E. Ritchie *et al.*, "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucl. Acids Res.*, vol. 43, no. 7, p. e47, 2015.

[67] X. Zhang, J. Zhao, J.-K. Hao, X.-M. Zhao, and L. Chen, "Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks," *Nucl. Acids Res.*, vol. 43, no. 5, p. e31, 2015.

[68] S. Ma, L. Xue, and H. Zou, "Alternating direction methods for latent variable Gaussian graphical model selection," *Neural Comput.*, vol. 25, no. 8, pp. 2172–2198, 2013.

[69] T. Min and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM J. Optim.*, vol. 21, no. 1, pp. 57–81, 2011.

[70] J. Fan, H. Liu, Y. Ning, and H. Zou, "High dimensional semiparametric latent graphical model for mixed data," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 79, no. 2, pp. 405–421, 2017.
[71] M. Y. Wu, D. Q. Dai, X. F. Zhang, and Y. Zhu, "Cancer subtype discovery and biomarker identification via a new robust network clustering algorithm," *PLoS ONE*, vol. 8, no. 6, 2013, Art. no. e66256.
[72] N. Städler *et al.*, "Molecular heterogeneity at the network level: High-dimensional testing, clustering and a TCGA case study," *Bioinformatics*, vol. 33, no. 18, pp. 2890–2896, 2017.
[73] Y. Zhao, A. K. Shrivastava, and K. L. Tsui, "Regularized Gaussian mixture model for high-dimensional clustering," *IEEE Trans. Cybern.*, vol. 49, no. 10, pp. 3677–3688, Oct. 2019.

Xiao-Fei Zhang received the Ph.D. degree from the Department of Mathematics, Sun Yat-sen University, Guangzhou, China, in 2013.

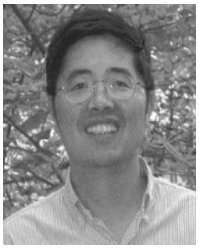
He is currently an Associate Professor with the School of Mathematics and Statistics, Central China Normal University, Wuhan, China. He has coauthored more than 30 journal papers mainly on novel machine learning methodologies applied to biological problems in journals, such as *Bioinformatics*, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and the IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS. His research interests include machine learning, data mining, and bioinformatics.

Le Ou-Yang (M'16) received the Ph.D. degree from Sun Yat-sen University, Guangzhou, China, in 2015.

He is currently an Assistant Professor with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China. His current research interests include machine learning, data mining, and bioinformatics.

Ting Yan received the Ph.D. degree from the Department of Statistics, University of Science and Technology of China, Hefei, China, in 2012.

He is currently a Professor with the School of Mathematics and Statistics, Central China Normal University, Wuhan, China. He has coauthored more than 20 journal papers mainly on network models, random graphs in journals, such as the *Annals of Statistics*, the *Journal of the Acoustical Society of America*, and *Biometrika*. His research interests include network data analysis, high-dimensional data analysis, and variable selection.



Xiaohua Tony Hu received the Ph.D. degree from the University of Regina, Regina, SK, Canada, in 1995.

He joined Drexel University, Philadelphia, PA, USA, in 2002, where he is a Full Professor and the Founding Director of the Data Mining and Bioinformatics Lab, College of Computing and Informatics. He is also serving as the Founding Co-Director of the NSF Center (I/U CRC) on Visual and Decision Informatics, the IEEE Computer Society Bioinformatics and Biomedicine Steering Committee Chair, and the IEEE Computer Society Big Data Steering Committee Chair. He is a Scientist, a Teacher, and an Entrepreneur. He founded the *International Journal of Data Mining* and *Bioinformatics* (SCI indexed) in 2006. Earlier, he worked as a Research Scientist with the World-Leading Research and Development Centers, such as Nortel Research Center, and Verizon Lab (the former GTE Laboratories). In 2001, he founded the DMW Software, Silicon Valley, CA, USA. He has a lot of experience and expertise to convert original ideas into research prototypes, and eventually into commercial products. Many of his research ideas have been integrated into commercial products and applications in data mining fraud detection, database marketing.

Hong Yan (F'06) received the Ph.D. degree from Yale University, New Haven, CT, USA.

He was a Professor of imaging science with the University of Sydney. He is currently a Professor of computer engineering with the City University of Hong Kong, Hong Kong. His research interests include image processing, pattern recognition, and bioinformatics. He has published over 300 journal papers in the above areas.

Prof. Yan received the 2016 Norbert Wiener Award from IEEE SMC Society for contributions to image and biomolecular pattern recognition techniques. He was elected an IAPR fellow for contributions to document image analysis and an IEEE fellow for contributions to image recognition techniques and applications.