Inferring cancer common and specific gene networks via multi-layer joint graphical model

Yuanxiao Chen [a], Xiao-Fei Zhang [b], Le Ou-Yang [a],*

[a] Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen Key Laboratory of Media Security, and Guangdong Laboratory of Artificial Intelligence and Digital Economy(SZ), Shenzhen University, Shenzhen, China
[b] School of Mathematics and Statistics & Hubei Key Laboratory of Mathematical Sciences, Central China Normal University, Wuhan, China

A B S T R A C T

Cancer is a complex disease caused primarily by genetic variants. Reconstructing gene networks within tumors is essential for understanding the functional regulatory mechanisms of carcinogenesis. Advances in high-throughput sequencing technologies have provided tremendous opportunities for inferring gene networks via computational approaches. However, due to the heterogeneity of the same cancer type and the similarities between different cancer types, it remains a challenge to systematically investigate the commonalities and specificities between gene networks of different cancer types, which is a crucial step towards precision cancer diagnosis and treatment. In this study, we propose a new sparse regularized multi-layer decomposition graphical model to jointly estimate the gene networks of multiple cancer types. Our model can handle various types of gene expression data and decomposes each cancer-type-specific network into three components, i.e., globally shared, partially shared and cancer-type-unique components. By identifying the globally and partially shared gene network components, our model can explore the heterogeneous similarities between different cancer types, and our identified cancer-type-unique components can help to reveal the regulatory mechanisms unique to each cancer type. Extensive experiments on synthetic data illustrate the effectiveness of our model in joint estimation of multiple gene networks. We also apply our model to two real data sets to infer the gene networks of multiple cancer subtypes or cell lines. By analyzing our estimated globally shared, partially shared, and cancer-type-unique components, we identified a number of important genes associated with common and specific regulatory mechanisms across different cancer types.

1. Introduction

Biological processes within cells generally involve the regulations between genes. Cancer is a complex disease that is mainly caused by genetic variants. Reconstructing the gene regulatory networks within tumor tissues or cancer cell lines and detecting abnormal gene regulations can help to reveal the underlying molecular mechanisms, which can provide critical insights into the diagnosis and treatment of different cancer types [1]. Due to the highly dynamic nature of gene network (it usually changes at any time with the change of environment or cell states) [1,2], it is difficult to capture gene networks through biological experiments. Recently, with the rapid development of bulk tissue and single-cell RNA sequencing (scRNA-seq) technologies, we can quantify the expression profiles of individual tissues or cells, which pave the way for inferring gene networks at tissue or single cell resolution via computational approaches [3].

Over the past decades, various computational approaches have been proposed for inferring gene networks from microarrays or bulk RNA-seq data, such as boolean models [4], information theory-based models [5], differential equation-based models [6] and Gaussian graphical models [7]. Among these methods, Gaussian graphical models (GGMs) are popular due to their ability in predicting the direct interactions between genes. Due to technical inefficiencies, scRNA-seq data usually have excess number of false zero values (also known as dropout events) and higher levels of noise [8–10]. Thus, network inference methods developed for bulk RNA-seq data or microarrays may not be applicable to scRNA-seq data. Recently, several computational models have been developed for inferring gene networks from scRNA-seq data. For example, Woodhouse *et al.*

[11] proposed a boolean-based network inference model to estimate gene networks from scRNA-seq data. Li *et al.* [12] adapted the Gaussian graphical model to infer cell-type-specific gene co-expression networks from scRNA-seq data. But most of these methods are designed for single network estimation, which focus on inferring a single network for a specific cell type and ignore the similarities between the gene networks of different cell types.

To estimate multiple gene networks jointly, several multi-network estimation methods have been proposed. These methods mainly rely on Gaussian graphical models, as they are well suitable for generalization to multi-network inference frameworks. For example, Danaher *et al.* [13] proposed two joint graphical models to simultaneously estimate multiple networks that share certain characteristics. By considering the similarities or differences between multiple networks are driven by common or specific hub nodes, Mohan *et al.* [14] developed a node-based joint graphical model. To identify the common and unique components of different networks when inferring multiple networks, Zhang *et al.* [15] proposed a joint graphical model, and estimated multiple gene networks across multiple subpopulations and data types simultaneously. Although these methods have been successfully used in multi-network estimation, they are all designed for bulk RNA-seq data. Unlike bulk RNA-seq data, scRNA-seq data are non-Gaussian distributed [16,17], and they usually contain a great number of missing values. To handle non-Gaussian data with missing values and identify the common and unique network structures of different cell types, Wu *et al.* [18] proposed a joint Gaussian copula graphical model to jointly estimate the gene networks of different cell types.

Although the methods proposed by Wu *et al.* [18] can jointly estimate multiple networks, they only explored the common and unique components of different networks from one layer. Cancer, however, is a heterogeneous disease. A single cancer type may have a wide variety of subtypes or cancer cell lines, while distinct cancer types may share certain characteristics [19,20]. Thus, the similarities between different cancer types are heterogeneous and multi-layered. For example, there may be gene regulatory relationships shared across all cancer types that act as the backbone of living cells and perform essential functions necessary for all cell types. But for subtypes or cell lines belonging to the same cancer type, there may be some partially shared network structure, shared only within them, that maintains the essential activity of that cancer type. Thus, new network inference methods that can capture the multi-layer similarities between different cancer types are needed.

To address the above problems, in this study, we propose a novel sparse regularized multi-layer decomposition graphical (SRMDG) model to estimate the gene networks of multiple cancer subtypes or cell lines jointly. To deal with various types of gene expression data, our model first employs a Gaussian copula graphical model with modified Kendall's tau correlation to relax the normality assumption of GGMs and deal with data with missing values. Then our model decomposes the gene network of each cancer subtype (or cell line) into three parts: 1) a globally shared component which captures the edges shared across all cancer types, 2) a partially shared component which captures the edges that only shared by subtypes or cell lines belonging to the same cancer type and 3) a cancer-type-unique component that captures the edges unique to this subtype or cell line (as shown in Fig. 1). Our model also imposes sparse penalties on all these components to encourage the inference of sparse gene

networks. We first conduct simulation studies to evaluate the performance of our proposed SRMDG model. By considering various scenarios and comparing the performance of our model with various state-of-the-art network inference methods, we were able to validate the effectiveness of our model in jointly estimating gene networks of multiple cancer types and identifying common and specific components of these networks. We also apply our model to two real datasets and reconstruct the gene networks of multiple cancer subtypes or cell lines. By analyzing the function of the hub genes in the components of the network that we identified as partially shared and cancer-type unique, we have revealed a number of mechanisms associated with the differentiation of cancer cell lines and the development of cancer.

2. Methods

2.1. Sparse Gaussian graphical models

Gaussian graphical models have been widely used in gene network inference since they can measure the conditional dependency relationships between genes, and distinguish between direct and indirect interactions [21]. Let $x_1, x_2, ..., x_n \in \mathbb{R}^{p \times 1}$ denote n different samples collected from a p-dimensional multivariate Gaussian distribution $N_p(0, \Sigma)$ (the samples are assumed to be centralized), where $\Sigma \in \mathbb{R}^{p \times p}$ is the covariance matrix. Here, a pair of random variables is conditional independent if and only if the corresponding entry in the precision matrix $\Omega = \Sigma^{-1}$ is zero. Therefore, by treating each gene as a random variable and the gene expression data of a subject on p genes as a p-dimensional random vector which follows a p-dimensional multivariate Gaussian distribution, inferring the conditional dependency relationships between p genes can be transformed into the problem of estimating the corresponding precision matrix.

To estimate the precision matrix Ω, especially when the dimension of random vector p is much larger than the number of samples n, the following graphical lasso model [21,22] has been proposed:

$$min_{\Omega \in S_{++}^p} - logdet(\Omega) + tr(S\Omega) + \lambda \sum_{i \neq j} |\Omega_{ij}|. \tag{1}$$

where $\Omega \in S_{++}^p$ is the positive-definite restriction, $S = (1/n) \sum_{m=1}^{n} x_m x_m^T$ is the sample covariance matrix, $tr(\cdot)$ and $det(\cdot)$ denote the trace and determinant of a matrix, and λ is the tuning parameter which controls the sparsity of the estimation of Ω.

When random vector $X = (X_1, ..., X_p)$ does not follow a Gaussian distribution, if we can find a group of monotone function $\{f_d\}_{d=1,...,p}$ such that $f(X) = (f_1(X_1), ..., f_p(X_p))$ follows a Gaussian distribution $N_p(0, \Sigma)$, then X is said to follow a nonparanormal distribution [23–26], i.e., $X \sim NPN_p(f, \Sigma)$. Here, the precision matrix $\Omega = \Sigma^{-1}$ still captures the conditional dependency relationships between $X_1, ..., X_p$.

2.2. Problem statement

Suppose there are $K = K_1 + K_2 + \cdots + K_T$ different types of cancer subtypes or cell lines, and according to the cancer types they belong to, suppose there are T different cancer types and each cancer type includes K_t subtypes or cell lines. For l-th subtype (or cell line) which belongs to t-th cancer type, we can collect the gene expression data of n_{tl} samples on p common genes. Let $X^{(tl)} = [x_1^{(tl)}, x_2^{(tl)}, ..., x_{n_{tl}}^{(tl)}]^T \in \mathbb{R}^{n_{tl} \times p}$ denote the gene expression data for
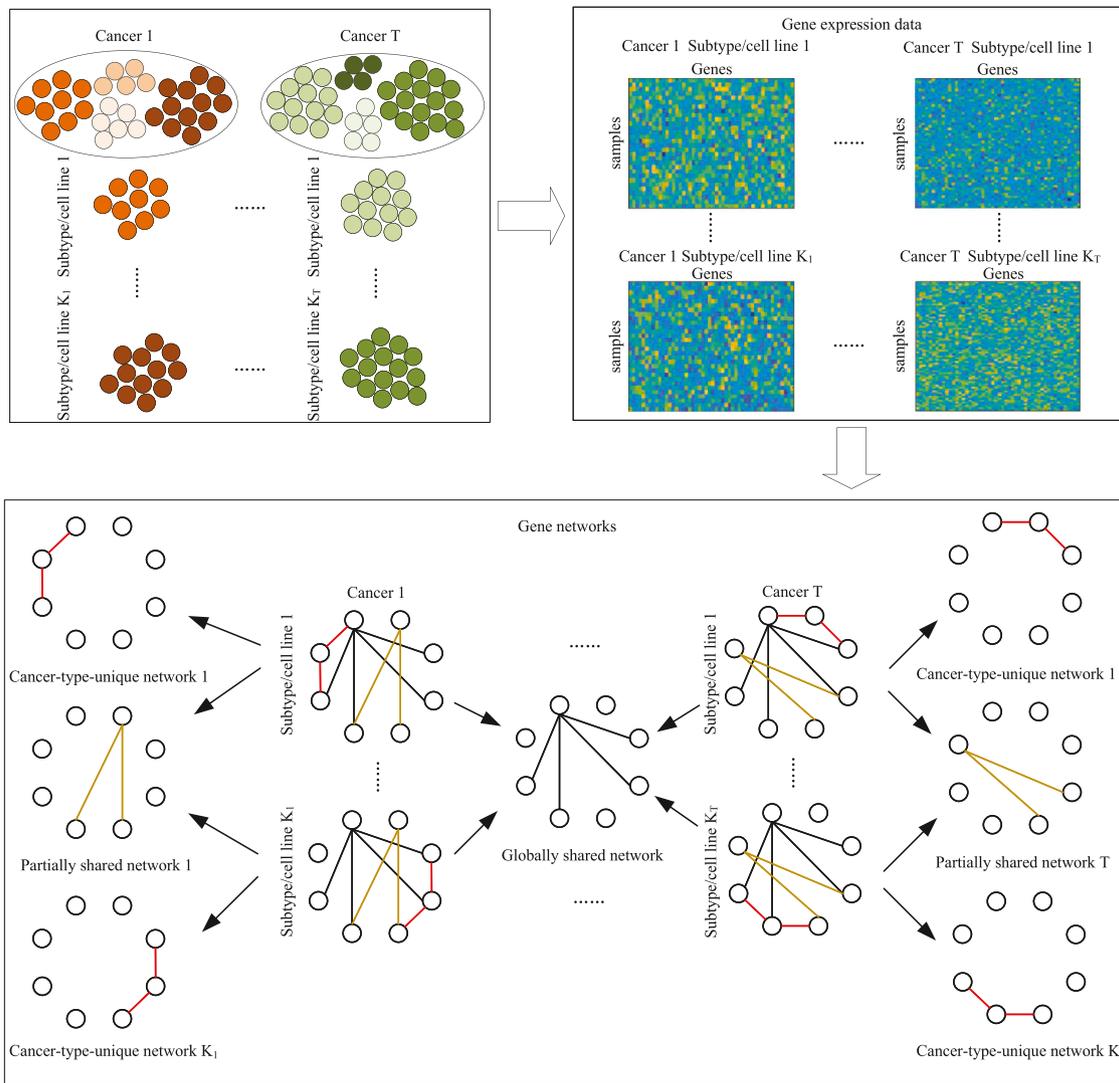
Fig. 1. Flowchart of our SRMDG model. Gene expression data collected from different cancer types are used as input to SRMDG to jointly infer multiple gene networks. Each network can be decomposed into three parts: a globally shared subnetwork, a partially shared subnetwork, and a cancer-type-unique subnetwork.

l-th subtype (or cell line) of t-th cancer type, where p is the number of genes that are common to all samples. Assume the gene expression data $X^{(tl)}$ of each subtype (or cell line) follows a nonparanormal distribution, i.e., $x_m^{(tl)} \sim NPN(f^{(tl)}, \Sigma^{(tl)})$ for $m = 1, \ldots, n_{tl}$. Note that for each subtype (or cell line), the conditional dependency relationships between p genes can be captured from the nonzero elements of the corresponding precision matrix $\Omega^{(tl)} = (\Sigma^{(tl)})^{-1}$. Thus, our goal is to jointly estimate these precision matrices $\{\Omega^{(tl)}\}_{t=1,\ldots,T}^{l=1,\ldots,K_t}$. For the sake of convenience, we denote $\{X^{(tl)}\}_{t=1,\ldots,T}^{l=1,\ldots,K_t}$, $\{\Sigma^{(tl)}\}_{t=1,\ldots,T}^{l=1,\ldots,K_t}$ and $\{\Omega^{(tl)}\}_{t=1,\ldots,T}^{l=1,\ldots,K_t}$ as $\{X^{(tl)}\}$, $\{\Sigma^{(tl)}\}$ and $\{\Omega^{(tl)}\}$.

2.3. Sparse regularized multi-layer decomposition graphical model

To estimate $\{\Omega^{(tl)}\}$ from their gene expression data $\{X^{(tl)}\}$, based on nonparanormal distribution, we can first formulate the following loss function:

$$L(\{\Omega^{(tl)}\}) = \sum_{t=1}^{T} \sum_{l=1}^{K_t} n_{tl}(tr(S^{(tl)}\Omega^{(tl)}) - logdet(\Omega^{(tl)})). \tag{2}$$

where $S^{(tl)}$ denotes the empirical nonparanormal covariance matrix of l-th subtype in t-th cancer type. Since gene expression data may contain a certain proportion of missing values, similar to previous studies [18,27–29], we adopt the modified Kendall's tau estimator to estimate $S^{(tl)}$. In particular, the modified Kendall's tau correlation is defined as follows:

$$\hat{\tau}_{jk}^{(tl)} = \frac{1}{n_{jk}^{(tl)}(n_{jk}^{(tl)} - 1)} \sum_{i=1,i'=1,i\neq i'}^{n_{tl}} b_{ij}^{(tl)} b_{ik}^{(tl)} b_{i'j}^{(tl)} b_{i'k}^{(tl)} sign(B_{ii'jk}^{(tl)}). \tag{3}$$

where $b_{ij}^{(tl)} = 1$ if $X_{ij}^{(tl)} \neq 0$ and $b_{ij}^{(tl)} = 0$ otherwise. $n_{jk}^{(tl)} = \sum_{i=1}^{n_{tl}} b_{ij} b_{ik}$ is the number of valid samples for gene pair (j, k), and $B_{ii'jk}^{(tl)} = (X_{ij}^{(tl)} - X_{i'j}^{(tl)})(X_{ik}^{(tl)} - X_{i'k}^{(tl)})$. Then, we can calculate $S^{(tl)}$ based on the following bridge function:

$$S_{jk}^{(tl)} = \begin{cases} sin\left(\frac{\pi}{2}\hat{\tau}_{jk}^{(tl)}\right), & j \neq k, \\ 1, & j = k. \end{cases} \tag{4}$$

To make sure the empirical nonparanormal covariance matrix is positive semidefinite, following previous studies [15,18], we replace $S^{(tl)}$ with its nearest semidefinite matrix.

Note that the gene networks of different cancer subtypes may have some common and specific network structures, where the common structures can reflect their similarities and the specific structures can reflect their difference. Thus, to identify the common and specific network structures of different cancer types systematically, we decompose $\Omega^{(tl)}$ into $\Omega^{(tl)} = A + Z^{(t)} + D^{(tl)}$. Here, A denotes the globally shared edges that shared across all cancer types, $Z^{(t)}$ denotes the partially shared edges that only shared by subtypes belonging to t-th cancer type ($t = 1, …, T$), and $D^{(tl)}$ denotes the specific edges that unique to l-th subtype in t-th cancer type ($t = 1, …, T, l = 1, …, K_t$). The sparsity of A, $Z^{(t)}$ and $D^{(tl)}$ reflects the degree of similarities between different cancer types.

In addition to the loss function in Eq. (2), we introduce the following penalty function to learn the similarities between different cancer types:

$$P(\{\Omega^{(tl)}\}) = \lambda\alpha(K_1 + K_2 + \cdots + K_T)\sum_{i \neq j}|A_{ij}|$$
$$+ \lambda\beta\sum_{t=1}^{T}\left(K_t\sum_{i \neq j}|Z_{ij}^{(t)}|\right) + \lambda(1 - \alpha - \beta)\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{i \neq j}|D_{ij}^{(tl)}|$$
$$s.t. \quad \Omega^{(tl)} = A + Z^{(t)} + D^{(tl)} \quad for\ t = 1,…,T, \quad l = 1,…,K_t. \tag{5}$$

where $\lambda > 0$ is a tuning parameter that controls the sparsity of the estimated networks. The larger the value of λ, the sparser the estimated networks. $0 < \alpha < 1$ and $0 < \beta < 1$ are used to control the sparse level of A, $Z^{(t)}$ and $D^{(tl)}$. The larger the values of α and β, the sparser the estimated A and $Z^{(t)}$. The smaller the values of α and β, the denser the estimated A and $Z^{(t)}$. The parameter selection strategy is introduced in next section.

By adding the penalty function (5) to loss function (2), we obtain the following sparse regularized multi-layer decomposition graphical model (SRMDG) for joint estimation of multiple gene networks:

$$\min_{\{\Omega^{(tl)},A,Z^{(t)},D^{(tl)}\}}\sum_{t=1}^{T}\sum_{l=1}^{K_t}n_{tl}\{tr(S^{(tl)}\Omega^{(tl)}) - logdet(\Omega^{(tl)})\}$$
$$+ \lambda\alpha(K_1 + K_2 + \cdots + K_T)\sum_{i \neq j}|A_{ij}|$$
$$+ \lambda\beta\sum_{t=1}^{T}\left(K_t\sum_{i \neq j}|Z_{ij}^{(t)}|\right) + \lambda(1 - \alpha - \beta)\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{i \neq j}|D_{ij}^{(tl)}|$$
$$s.t. \quad \Omega^{(tl)} = A + Z^{(t)} + D^{(tl)} \quad for\ t = 1,…,T, \quad l = 1,…,K_t. \tag{6}$$

2.4. Algorithm

Alternating direction method of multipliers (ADMM) algorithm [30] is used to solve problem (6), and the augmented Lagrangian of the objective function (6) is as follows:

$$L_\rho(\{\Omega^{(tl)}\}, \{A\}, \{Z^{(t)}\}, \{D^{(tl)}\}, \{Q^{(tl)}\})$$
$$= \sum_{t=1}^{T}\sum_{l=1}^{K_t}n_{tl}\{tr(S^{(tl)}\Omega^{(tl)}) - logdet(\Omega^{(tl)})\}$$
$$+ \lambda\alpha(K_1 + K_2 + \cdots + K_T)\sum_{i \neq j}|A_{ij}|$$
$$+ \lambda\beta\sum_{t=1}^{T}\left(K_t\sum_{i \neq j}|Z_{ij}^{(t)}|\right) + \lambda(1 - \alpha - \beta)\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{i \neq j}|D_{ij}^{(tl)}|$$
$$+ \sum_{t=1}^{T}\sum_{l=1}^{K_t} < Q^{(tl)}, \Omega^{(tl)} - (A + Z^{(t)} + D^{(tl)})>$$
$$+ \frac{\rho}{2}\sum_{t=1}^{T}\sum_{l=1}^{K_t}\|\Omega^{(tl)} - (A + Z^{(t)} + D^{(tl)})\|_F^2 \tag{7}$$

where $Q^{(tl)}$ is dual variable, $\|A\|_F$ denotes Frobenius norm of matrix A, $<A, B> = tr(AB^T)$, ρ is the penalty parameter. Each parameter can be updated by fixing other parameters in turn and the algorithm can be seen in Algorithm 1, where the operator Expand [31,32] is given as follows:

$$Expand(A, \rho, n) = argmin_{\Omega \in S_{++}^p}\left\{-nlogdet(\Omega) + \frac{\rho}{2}\|\Omega - A\|_F^2\right\}$$
$$= \frac{1}{2}U\left(D + \sqrt{D^2 + \frac{4n}{\rho}I}\right)U^T \tag{8}$$

where UDU^T is the eigenvalue decomposition of a symmetric matrix A. The operator Γ is given by:

$$\Gamma(A, \lambda) = argmin_X\left\{\frac{1}{2}\|X - A\|_F^2 + \lambda\sum_{i \neq j}|X_{ij}|\right\} \tag{9}$$

we set $\rho_{incr} = 1.2$ to speed up the convergence of the algorithm. In the implementation of our algorithm, the stopping criterions for the inner loop are:

$$\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{ij}|(\Omega_{ij}^{(tl)})^{(k)} - (\Omega_{ij}^{(tl)})^{(k-1)}| \leq 10^{-5}\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{ij}|(\Omega_{ij}^{(tl)})^{(k)}| \tag{10}$$

$$\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{ij}|(\Omega_{ij}^{(tl)})^{(k)} - ((A_{ij})^{(k)} + (Z_{ij}^{(t)})^{(k)} + (D_{ij}^{(tl)})^{(k)})|$$
$$\leq 10^{-5}\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{ij}|(\Omega_{ij}^{(tl)})^{(k)}| \tag{11}$$

where $(\cdot)^{(k)}$ denotes the estimated parameters at the k-th iteration.

Algorithm 1. ADMM algorithm for solving SRMDG.

unique subnetworks. The number of edges in the cancer-type-unique subnetwork is the number of edges in the globally shared subnetwork multiply by ϕ. In this study, we set $\phi = 0.3, 0.5, 1$;

Require: Sample covariance matrices $S^{(tl)}$, tuning parameters λ, α and β, samples sizes n_{tl}

Ensure: Estimated precision matrices $\hat{\Omega}^{(tl)}$, estimated globally shared sub-network $\hat{A}$, estimated partially shared subnetworks $\hat{Z}^{(t)}$ and estimated cancer-type-unique networks $\hat{D}^{(tl)}$

1: Initialization: $\Omega^{(tl)} = I$, $A = I$, $Z^{(t)} = I$, $D^{(tl)} = I$, $Q^{(tl)} = 0$, $\rho = 0.1$, $\rho_{incr} = 1.2$, $\rho_{max} = 10^{10}$.

2: **while** not converged **do**

3:     $\Omega^{(tl)} \leftarrow Expand(A + Z^{(t)} + D^{(tl)} - \frac{1}{\rho}(n_{tl}S^{(tl)} + Q^{(tl)}), \rho, n_{tl})$

4:     $A \leftarrow \Gamma(\frac{1}{K_1 + ... + K_T} \sum_{t=1}^{T} \sum_{l=1}^{K_t} \frac{Q^{(tl)}}{\rho} + \Omega^{(tl)} - Z^{(t)} - D^{(tl)}, \frac{\lambda\alpha}{\rho})$

5:     $M^{(t)} \leftarrow \frac{1}{K_t} \sum_{l=1}^{K_t} (\frac{Q^{(tl)}}{\rho} + \Omega^{(tl)} - A - D^{(tl)})$

6:     $\{Z^{(1)}, ..., Z^{(T)}\} \leftarrow \Gamma(\{M^{(1)}, ..., M^{(T)}\}, \frac{\lambda\beta}{\rho})$

7:     $D^{(tl)} \leftarrow \Gamma(\frac{Q^{(tl)}}{\rho} + \Omega^{(tl)} - A - Z^{(t)}, \frac{\lambda(1-\alpha-\beta)}{\rho})$

8:     $Q^{(tl)} \leftarrow Q^{(tl)} + \rho(\Omega^{(tl)} - (A + Z^{(t)} + D^{(tl)}))$

9:     $\rho \leftarrow min(\rho \cdot \rho_{incr}, \rho_{max})$

10: **end while**

11: **return** $\hat{\Omega}^{(tl)} = A + Z^{(t)} + D^{(tl)}$, $\hat{A} = A$, $\hat{Z}^{(t)} = Z^{(t)}$, $\hat{D}^{(tl)} = D^{(tl)}$

---

3. Simulation studies

3.1. Data generation

To simulate conventional bulk sequence data and scRNA-seq data separately, we generate two types of synthetic data using different methods. In this subsection, we discuss the details of the data generation procedure.

3.1.1. Gaussian data generation

In this study, we consider $T = 3$ cancers and 4 subtypes for each cancer, resulting in 12 scale-free networks corresponding to 12 subtypes. Each network contains $p = 100$ common genes. We generate different sample sizes $n = 50, 100, 200$ to evaluate the performance. The details of generating Gaussian distributed data for scale-free networks are described as follows:

1. Utilize the "sample_pa()" function in "igraph" package to generate a scale-free network with p nodes as the globally shared subnetwork that shared by all cancer types;
2. For each cancer type, use the "sample_gnm()" function in "igraph" package to generate a partially shared subnetwork that shared by subtypes belonging to this cancer type. The number of edges in the partially shared subnetwork is the number of edges in the globally shared subnetwork multiply by ε. In this study, we set $\varepsilon = 0.8, 1, 1.2$;
3. For subtypes belonging to same cancers, use the "sample_gnm()" function in "igraph" package to generate their cancer-type-

4. For each network, create a $p \times p$ symmetric matrix to store the weights assigned to edges in the network. In this study, the edge weights are generated from $Unif([-1, -0.5] \cup [0.5, 1])$;
5. $\Omega^{(tl)}$ is generated as $\Omega^{(tl)} = A + Z^{(t)} + D^{(tl)} + \delta^{(t)}I$, where A denotes the globally shared subnetwork, $Z^{(t)}$ denotes the partially shared subnetwork, $D^{(tl)}$ denotes the cancer-type-unique subnetwork and $\delta^{(t)}$ is used to guarantee the positive definiteness of $\Omega^{(tl)}$, I is a $p \times p$ identity matrix;
6. Generate n_{tl} independent observations from $N(0, (\Omega^{(tl)})^{-1})$ and use them as gene expression data sets. Then the generated data can be used to estimate the sample covariance matrices $S^{(tl)}$.

3.1.2. Non-Gaussian data generation

We employ the data generation procedure described in [33] to generate the non-Gaussian data. The networks are generated in the same way as for Gaussian data (Steps 1–3). When the adjacency matrices are generated, we use the "create_network_from_adjacency_matrix()" function and the "gen_partial_correlations()" function in "SeqNet" package to generate the corresponding networks and precision matrices respectively. Then we use the "gen_rnaseq()" function in "SeqNet" package to generate the gene expression data with the default settings in the breast cancer dataset.

3.2. Compared methods and evaluation metrics

In SRMDG, the values of tuning parameters α and β that control the level of similarity among networks are set to $\alpha = 0.3, 0.2, 0.1$,

β = 0.2, 0.1 and parameter λ that controls the level of sparsity is set to 10 possible values equally spaced in log scale between 0.1 and 2.5. To evaluate the performance of SRMDG, we compare it with three state-of-the-art network inference methods:

- Graphical Lasso (glasso) [34], a network inference method designed for single network estimation. When applying glasso, networks are estimated for each subtype separately and the tuning parameter that controls the sparsity of the estimated network is set to λ = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6.
- Group graphical Lasso (GGL) [13], a method for jointly estimate multiple Gaussian graphical models with a group Lasso penalty to control the similarities between individual networks. To easy interpret "sparsity" and "similarity", GGL reparameterize the tuning parameters as: $\omega_1 = \lambda_1 + (1/\sqrt{2})\lambda_2$, $\omega_2 = (1/\sqrt{2})\lambda_2 / (\lambda_1 + (1/\sqrt{2})\lambda_2)$. We set ω_2 = 0.7, 0.5, 0.3, 0.15 and ω_1 are set to the total of 10 possible values equally spaced in log scale between 0.01 and 1.5. When applying GGL, the networks of subtypes belonging to the same cancer type are estimated jointly.
- Joint estimation of gene networks across multiple subpopulations and data types (JEGN) [15] uses a group Lasso penalty to encourage a similar pattern of sparsity among data types and to encourage a shared network structure among different subpopulations. The tuning parameter of JEGN is set to α = 0.2, 0.3, 0.4, 0.5 for Gaussian data and α = 0.1, 0.2, 0.3, 0.4 for non-Gaussian data and β are set to the total of 10 possible values equally spaced in log scale between 0.1 and 2. For JEGN, the number of data types is set to one, and the networks of subtypes belonging to the same cancer type are estimated jointly.

All of these methods are implemented in R language. Since glasso and GGL are designed for Gaussian data, we use Kendall's tau coefficient to estimate the sample covariance matrices for non-Gaussian data.

We evaluate the performance of various methods in terms of precision and recall, as real networks are typically sparse [35]. Let $\hat{\Omega}_{ij}^{(tl)}$ and $\Omega_{ij}^{(tl)}$ denote the (i, j)-th entry of the estimated precision matrix $\hat{\Omega}^{(tl)}$ and the true precision matrix $\Omega^{(tl)}$, then the precision and recall can be computed as follows:

$$precision = \frac{\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{i<j} 1\{\hat{\Omega}_{ij}^{(tl)} \neq 0, \Omega_{ij}^{(tl)} \neq 0\}}{\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{i<j} 1\{\hat{\Omega}_{ij}^{(tl)} \neq 0\}} \qquad (12)$$

$$recall = \frac{\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{i<j} 1\{\hat{\Omega}_{ij}^{(tl)} \neq 0, \Omega_{ij}^{(tl)} \neq 0\}}{\sum_{t=1}^{T}\sum_{l=1}^{K_t}\sum_{i<j} 1\{\Omega_{ij}^{(tl)} \neq 0\}} \qquad (13)$$

where $1\{\cdot\}$ is an indicator function and we set $\hat{\Omega}_{ij} = 0$ if $|\hat{\Omega}_{ij}| < 10^{-5}$ and $\hat{\Omega}_{ij} = 1$ otherwise.

3.3. Simulation results

To study the effect of the sample size on network inference, we generate synthetic data with different sample sizes, i.e., n = 50, 100, 200. The number of genes is set to p = 100.

Figs. 2–4 illustrate the results of various methods on Gaussian data with ε = 0.8, 1, 1.2 respectively. Here, the results of each method are obtained by averaging over 5 random repetitions. Different columns correspond to different sample sizes, and different rows

correspond to different values of ϕ. Within each plot, each colored line corresponds to the results of a method, e.g., with fixed values of α and β for SRMDG (α for JEGN and ω_2 for GGL), and varying values of λ for SRMDG and JEGN (ω_1 for GGL). We can find from these figures that JEGN outperforms GGL and glasso, demonstrating the benefit of considering the similarities between different networks. Our SRMDG can achieve the best performance in all cases, indicating the effectiveness of our model in capturing the multi-layer similarities between different networks.

To simulate dropout events in scRNA-seq data, we also generate non-Gaussian data with missing values and consider different missing rates. For 50% of the samples, the missing rate δ is set to be 0.5 and for the rest of the samples, δ is set to be 0.1. The results on these datasets are shown in Figs. 5, 6 and 7. We can see from these figures that SRMDG still outperforms the other compared methods in most cases.

4. Real data analysis

4.1. Parameter selection

When analyzing real datasets, there are three tuning parameters (α, β and λ) in our model that need to be predefined. λ is used to control the overall sparsity of the estimated networks, while α and β are used to control the sparsity of the estimated globally shared, partially shared and cancer-type-unique subnetworks. To get an optimal model and balance the complexity of the model simultaneously, the values of α, β and λ are determined via Akaike information criterion (AIC):

$$AIC = \sum_{t=1}^{T}\sum_{l=1}^{K_t} \{n_{tl}tr(\hat{S}^{(tl)}\hat{\Omega}^{(tl)}) - n_{tl}logdet(\hat{\Omega}^{(tl)}) + 2n_e^{(tl)}\} \qquad (14)$$

where $2n_e^{(tl)}$ is the number of non-zero entities in the estimated precision matrix $\hat{\Omega}^{(tl)}$, $\hat{S}^{(tl)}$ is the estimated convariance matrix. To avoid the condition that one of the estimated networks in A, $Z^{(t)}$, $D^{(tl)}$ is empty, the selection of α, β and λ are restricted such that the number of edges in each network is larger than 150.

4.2. Joint analysis of breast and ovarian cancer

To evaluate the performance of SRMDG on real datasets, we first apply it to the bulk tissue gene expression data of breast cancers and ovarian serous cystadenocarcinoma (OV) cancers collected from The Cancer Genome Atlas (TCGA: https://tcga.xenahubs.net) [36]. Here, we focus on 389 genes covered by mTOR, Notch, P53, and WNT pathways, which are critical cancer-related pathways [37]. In particular, for breast cancer, the microarrays of four subtypes are collected from Agilent G450 platform, including 94 Basal-like tumors, 58 HER2-enriched tumors, 231 Luminal A tumors and 127 Luminal B tumors. For ovarian cancer, the microarrays of four gene expression subtypes are collected from Agilent G450 platform, including 67 differentiated samples, 82 immunoreactive samples, 68 mesenchymal samples, and 75 proliferative samples.

We first analyze our estimated globally shared subnetwork shared by all cancer types. Fig. 8 presents the globally shared subnetwork predicted by our model. Nodes marked with yellow indicate the top-10 genes with the highest degree. Among these genes, ARAF, RPS6KB2, AKT1 and AKT1S1 play important roles in cell proliferation, differentiation, migration and survival [38–40]. Overexpression of MAPKAP1 has been shown to be associated with many types of
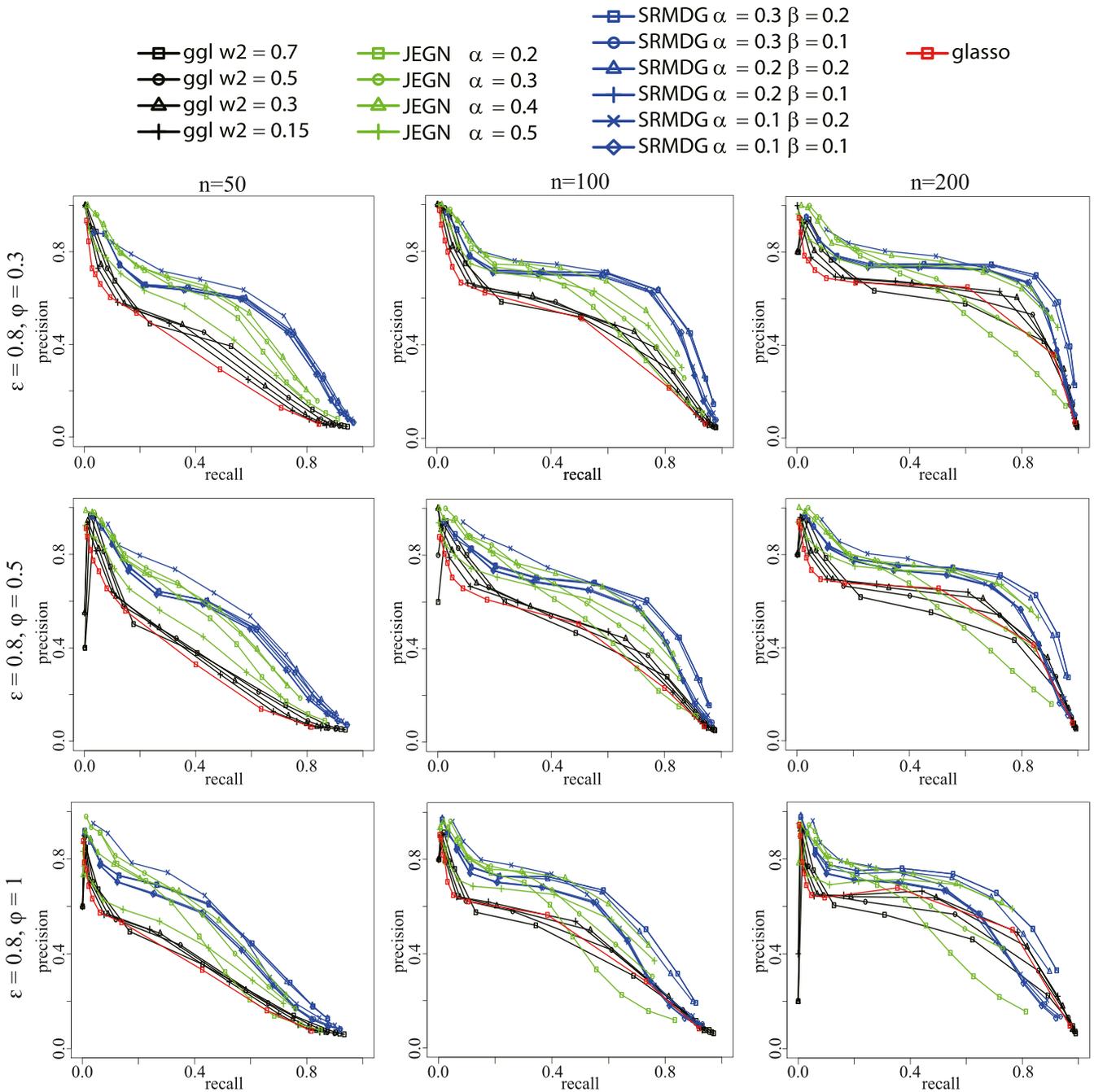
Fig. 2. The results on Gaussian data with $\varepsilon = 0.8$. Different columns correspond to different sample sizes, and different rows correspond to different values of ϕ. Within each plot, each colored line corresponds to the results of a method, e.g., with fixed values of α and β for SRMDG (α for JEGN and ω_2 for GGL), and varying values of λ for SRMDG and JEGN (ω_1 for GGL). The results are obtained by averaging over 5 random repetitions.

cancer [41]. SERPINF1, which acts as a multifunctional secreted protein, has anti-tumor function [42]. The expression of PLCB3 is higher in tumor tissue than in normal tissue [43]. As a novel tumor suppressor, CHUK/IKKα appears in a large number of human organs [44]. FAS and its ligand (FasL) are important in apoptosis and carcinogenesis and mutations in genes encoding FAS have been found to increase the risk of developing multiple types of cancer [45].

For partially shared subnetworks that are shared by subtypes belonging to the same cancer types, Table 1 shows the top-5 genes with the highest degrees in the partially shared subnetworks of different cancer types (genes that are presented in the globally shared subnetwork are removed). For genes detected in breast cancer, Chen *et al.* [46] showed that PIK3CD is associated with breast cancer. Evron *et al.* [47] found that CCND2 may be involved
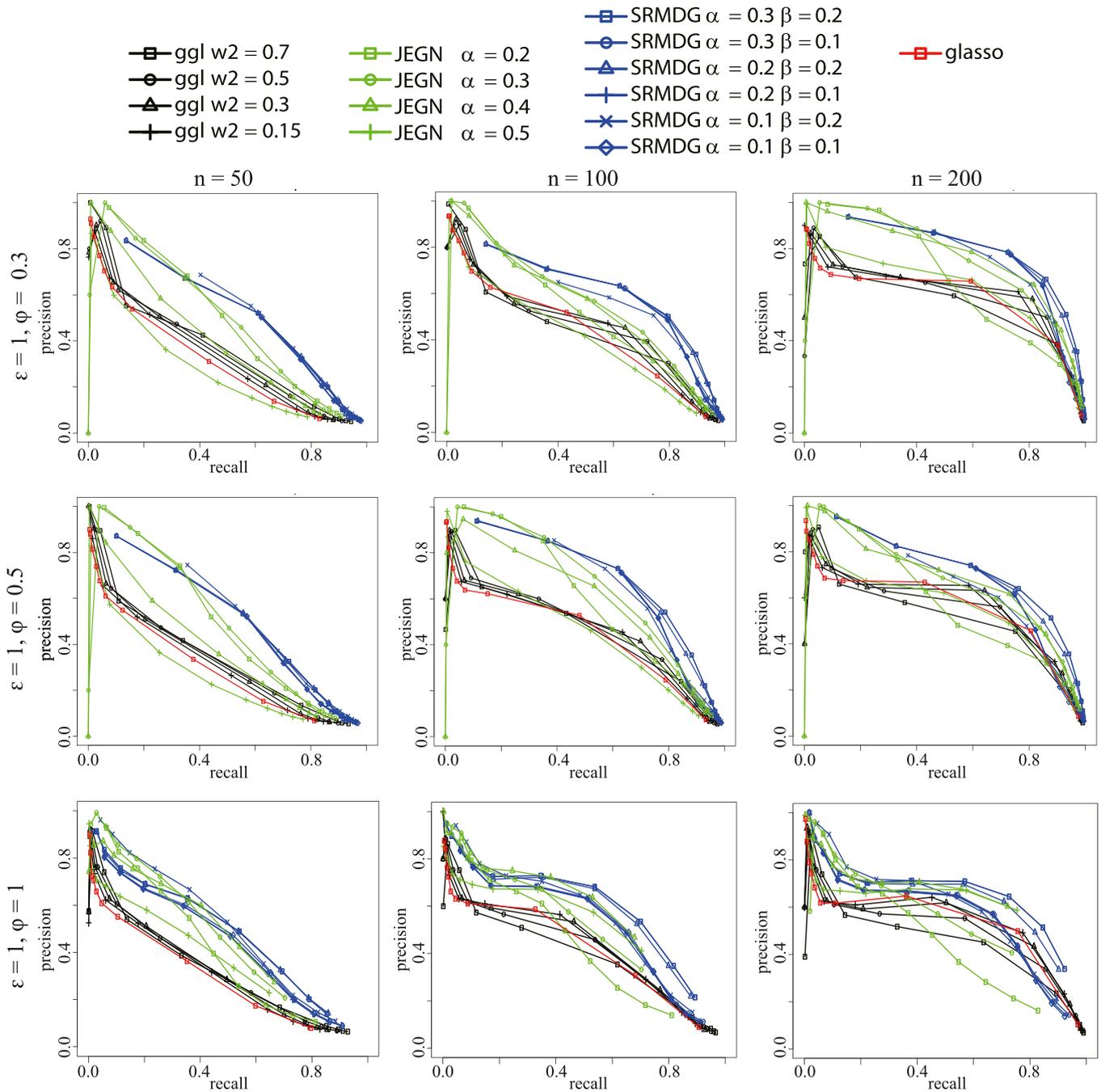
Fig. 3. The results on Gaussian data with $\varepsilon = 1$. Different columns correspond to different sample sizes, and different rows correspond to different values of ϕ. Within each plot, each colored line corresponds to the results of a method, e.g., with fixed values of α and β for SRMDG (α for JEGN and ω_2 for GGL), and varying values of λ for SRMDG and JEGN (ω_1 for GGL). The results are obtained by averaging over 5 random repetitions.

in the development of breast cancer. SFRP2 acts as a target for epigenetic inactivation [48], and PRR5 is frequently deleted during breast carcinogenesis [49]. Zhang *et al.* [50] found that MFNG modulates Notch activation in human and mouse Claudin-low breast cancer cell lines. For genes detected in ovarian cancer, Hoffmann *et al.* [51] found mutations of FZD9 in High-grade serous ovarian cancer (HGSOC) could impair signal transduction. Hao *et al.*

[52] discovered that knocking out GTSE1 significantly reduced the proliferation of ovarian cancer OVCA420 cells. Increasing evidence suggests that E2F1 is associated with ovarian carcinoma [53,54]. Aird *et al.* [55] showed that RRM2 affects the growth of human epithelial ovarian cancer cells. The expression of APC2 is significantly reduced in ovarian cancer cells compared to other cell lines [56].
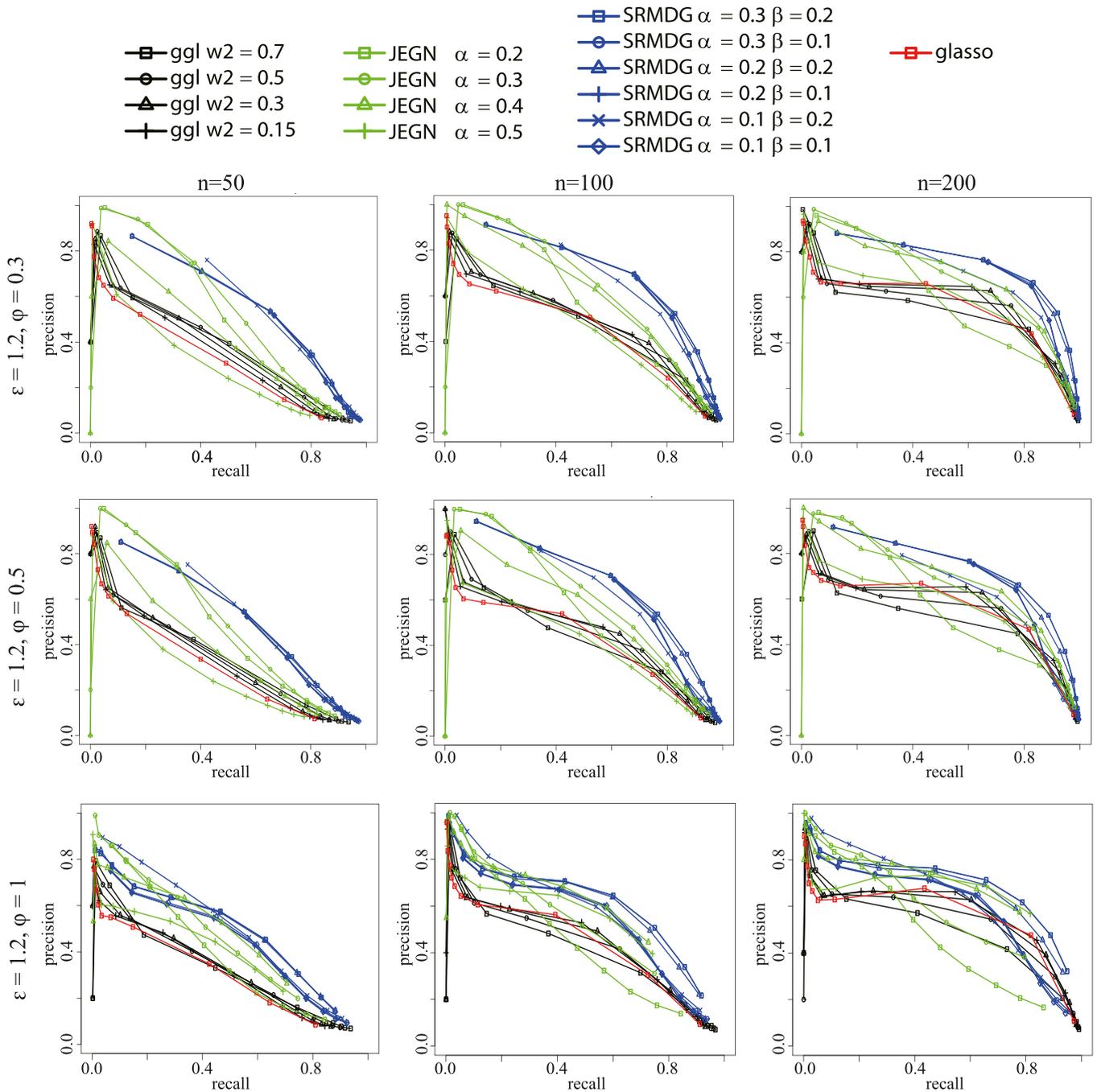
Fig. 4. The results on Gaussian data with $\varepsilon = 1.2$. Different columns correspond to different sample sizes, and different rows correspond to different values of ϕ. Within each plot, each colored line corresponds to the results of a method, e.g., with fixed values of α and β for SRMDG (α for JEGN and ω_2 for GGL), and varying values of λ for SRMDG and JEGN (ω_1 for GGL). The results are obtained by averaging over 5 random repetitions.

For cancer-type-unique subnetworks, Table 2 shows the top-5 genes with the highest degree in each cancer-type-unique subnetwork (genes presented in the globally shared subnetwork and partially shared subnetworks are removed). In Basal-like tumors, it is common to see the functional loss of RB1 [57] and FZD4 is up-regulated in the Basal-like and mesenchymal subtypes [58,59]. In HER2-enriched tumors, TLE3 is overexpressed in the largest number of HER2-positive breast cancer patients [60]. Extensive data suggest an intimate relationship between ESR1 and HER2-enriched tumors [61–63], with high FGF2 levels promoting HER2 expression in breast cancer cells [64–66]. Monaco *et al.* [67] found that LPIN1 was lower in the HER2 subtype. In both Luminal A and Luminal B tumors, Darbeheshti *et al.* [68] showed a significant positive correlation between EGFR expression and luminal tumor
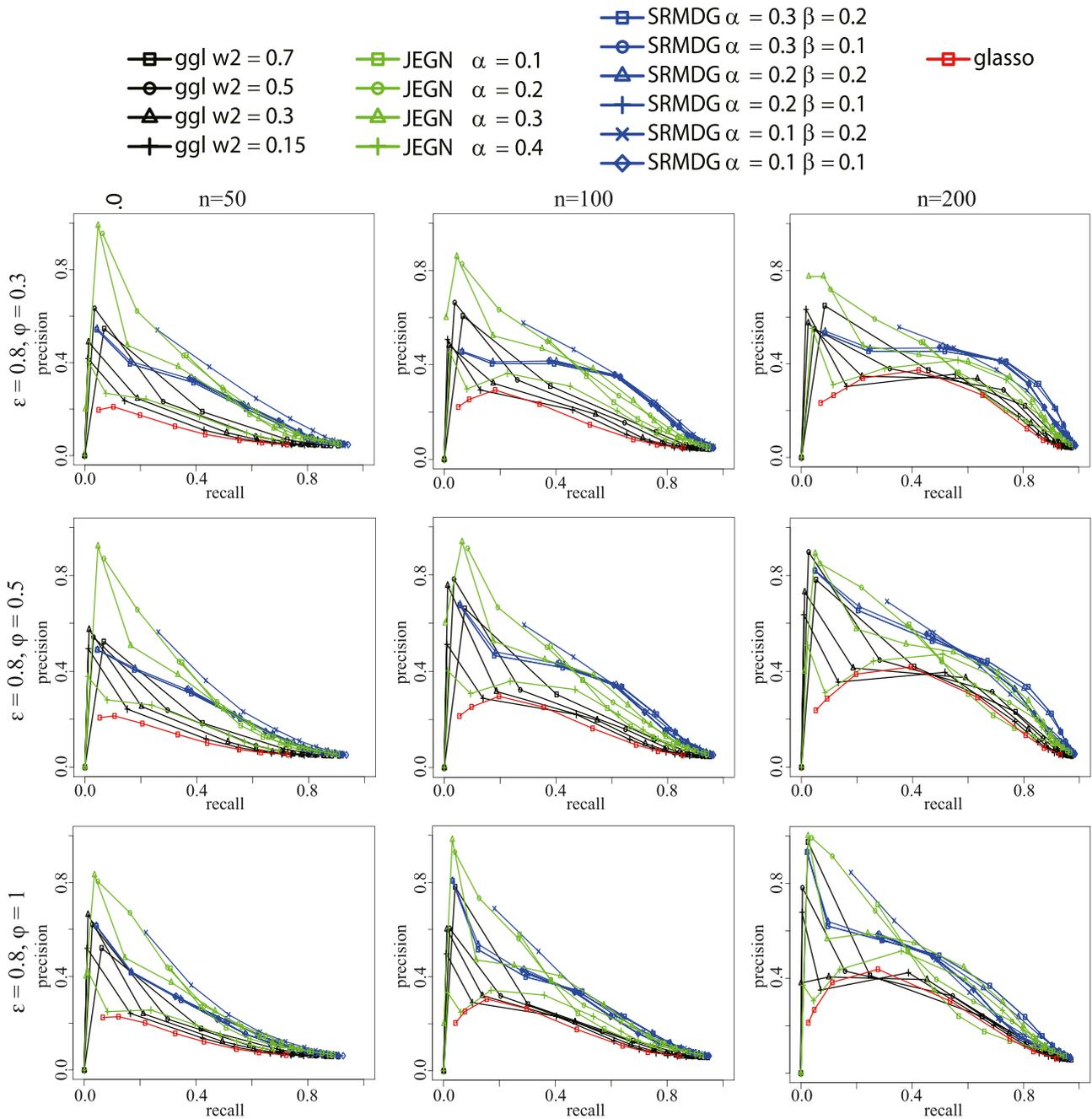
Fig. 5. The results on non-Gaussian data with missing values with $\varepsilon = 0.8$. Different columns correspond to different sample sizes, and different rows correspond to different values of ϕ. Within each plot, each colored line corresponds to the results of a method, e.g., with fixed values of α and β for SRMDG (α for JEGN and ω_2 for GGL), and varying values of λ for SRMDG and JEGN (ω_1 for GGL). The results are obtained by averaging over 5 random repetitions.

size. Roarty *et al.* [69] found that mRNA expression of ROR1 decreased in luminal A and luminal B subtypes. Finn *et al.* [70] found that advanced ER+ /luminal subtypes of breast cancer can be significantly suppressed by combining a CDK4/6 inhibitor with an aromatase inhibitor. Mamoor *et al.* [71] found that the expression of SFRP1 in breast cancer was associated with overall survival in patients with luminal A subtype. The Notch pathway plays a key part in estrogen resistance in luminal B breast cancer [72]. Lubecka *et al.*

[73] suggested an increase in methylation within the MAML2 enhancer region in breast cancer cells.

In differentiated subtypes, Wei *et al.* [74] found that NKD2 plays a role in inhibiting ovarian cancer cell proliferation, colony formation, and cell migration, and can suppress tumor progression. PPARD target genes upregulated in ovarian cancer TAMs (tumor-associated macrophages) [75]. Some genes, including WNT10B, WNT9B, and WIF-1, exhibit tumor-specific expression or down-
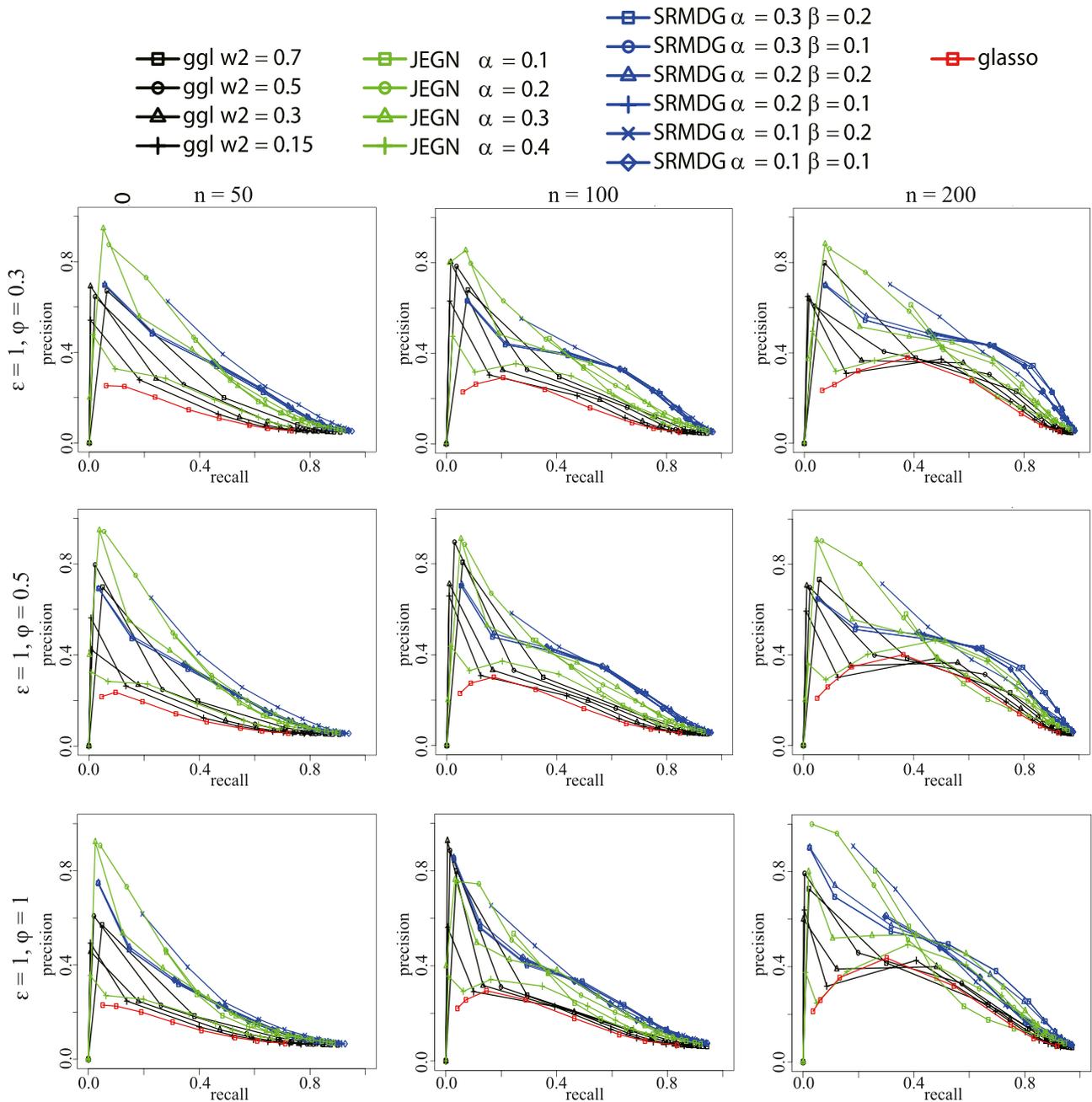
Fig. 6. The results on non-Gaussian data with missing values with $\varepsilon = 1$. Different columns correspond to different sample sizes, and different rows correspond to different values of ϕ. Within each plot, each colored line corresponds to the results of a method, e.g., with fixed values of α and β for SRMDG (α for JEGN and ω_2 for GGL), and varying values of λ for SRMDG and JEGN (ω_1 for GGL). The results are obtained by averaging over 5 random repetitions.

regulated expression compared to normal ovarian surface epithelium [76]. Davidson *et al.* [77] reported that CCNE1 expression levels were significantly higher in differentiated ovarian cancer than in diffuse malignant peritoneal mesothelioma. In the immunoreactive subtype, FOSL1 expression levels are correlated with tumor grade, stage, aggressiveness, and response to chemotherapy, and MAPK9 is up-regulated. [78]. Activation levels of EIF4E are higher in ovarian cancer cells than in normal ovarian epithelial cells [79]. In the mesenchymal subtype, Cecco *et al.* [80] suggested that FGF2 signaling plays a central role in maintaining the plasticity of

ovary-derived cells throughout the carcinogenesis process. Wang *et al.* [81] found that miR-130a, which is over-expressed in HGSOC, is a negative regulator of TSC1, and inhibition of miR-130a depresses levels of mesenchymal markers. IGF1 may inhibit the expression of E-cadherin and facilitate the epithelial-to-mesenchymal transition in ovarian cancer cells [82,83]. In the proliferative subtype, ESR1 regulates ovarian cancer cell proliferation and apoptosis by mediating E2 signaling [84]. HEY1 and HEYL have been shown to be associated with ovarian cancer [85]. Ma *et al.* [86] suggested that SIVA1 is stably overexpressed in ovarian
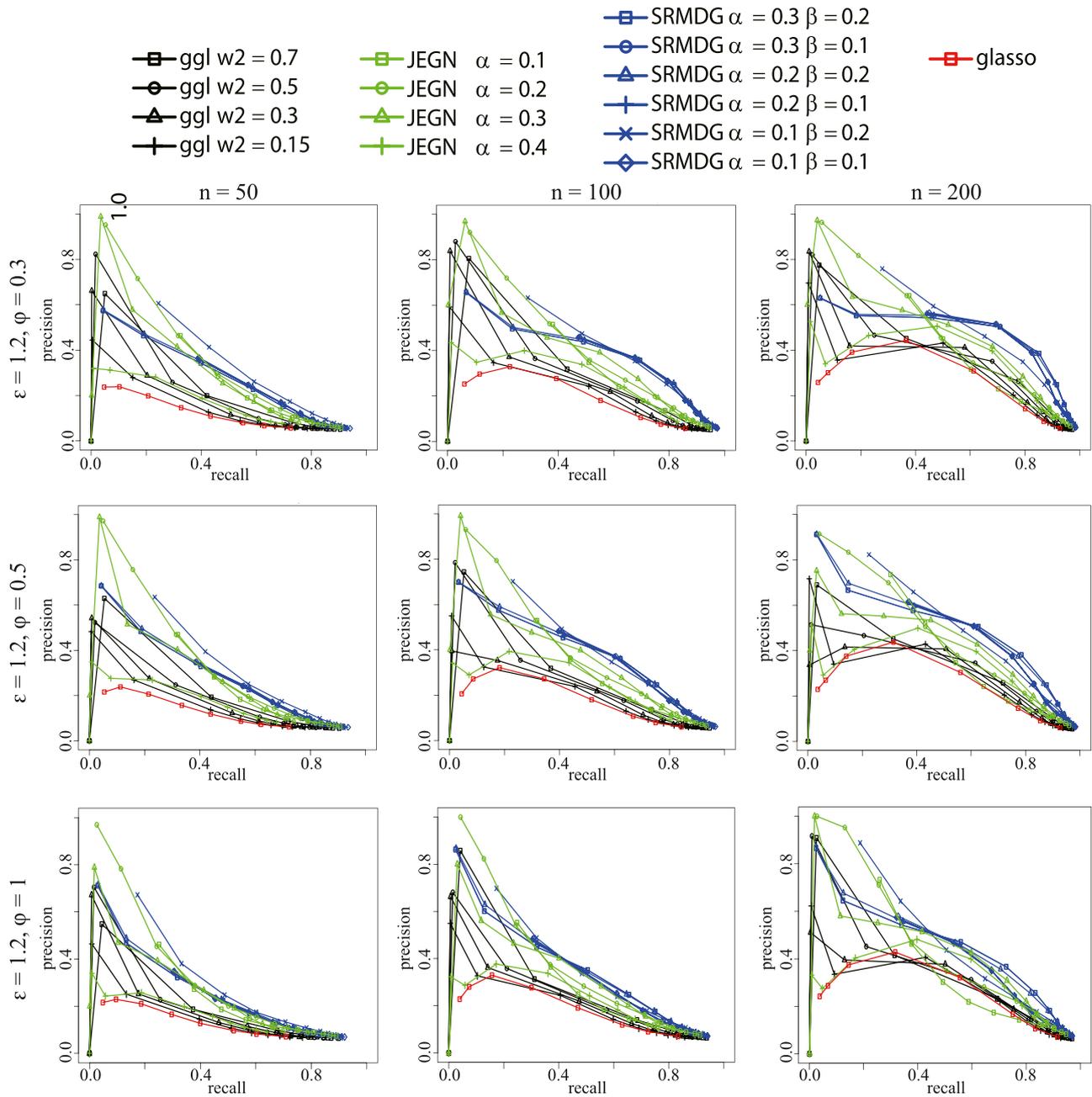
Fig. 7. The results on non-Gaussian data with missing values with ε = 1.2. Different columns correspond to different sample sizes, and different rows correspond to different values of ϕ. Within each plot, each colored line corresponds to the results of a method, e.g., with fixed values of α and β for SRMDG (α for JEGN and ω_2 for GGL), and varying values of λ for SRMDG and JEGN (ω_1 for GGL). The results are obtained by averaging over 5 random repetitions.

cancer cell lines, and that overexpression of SIVA1 inhibits cancer cell proliferation.

4.3. Joint analysis of CML and HCC cancer cell lines

We then apply SRMDG to scRNA-seq data to assess if SRMDG can help to reveal the common and specific gene regulatory mechanisms of different cancer cell lines. We collect the gene expression data of K562 and KBM7 cell lines in chronic myelogenous leukemia (CML) as well as HuH-7 and HuH-1 cell lines in hepatocellular carcinoma

(HCC). The K562 cell line includes 91 cells [87], which can be downloaded from the GEO database (GSE76312). The KBM7 cell line includes 95 cells [88] and can be downloaded from the GEO database (GSE68596). HuH-1 and HuH-7 cell lines are collected from the same dataset [89] which includes 55 and 63 cells respectively and can be downloaded from the GEO database (GSE103866). In this study, 411 genes belonging to the chronic myelogenous leukemia, hepatocellular carcinoma, Notch, WNT and PI3K-AKT pathways were analyzed.

Fig. 9 shows our predicted globally shared subnetwork. Nodes marked with yellow denote the top-10 genes with the highest
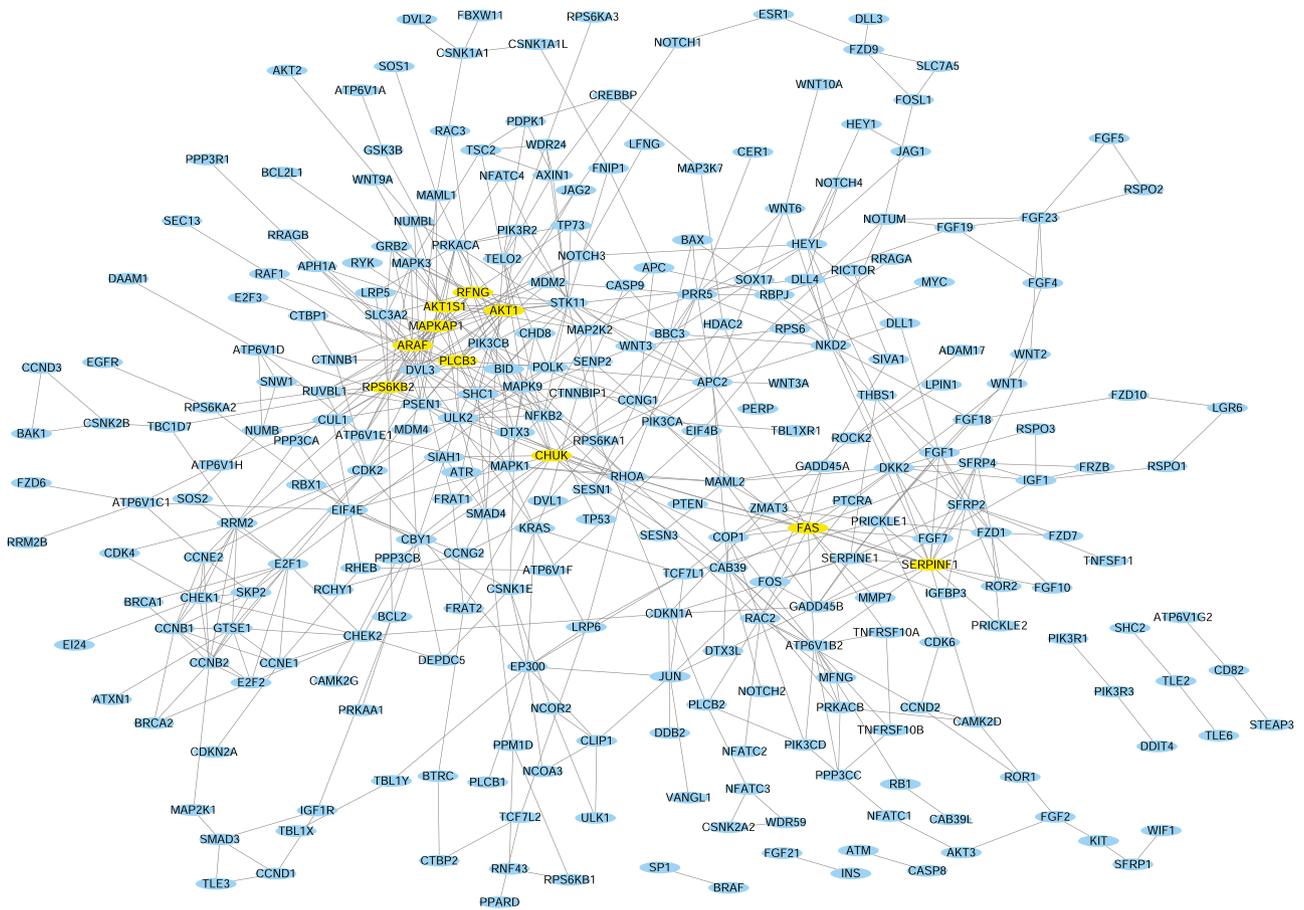
Fig. 8. The globally shared subnetwork shared by all subtypes of breast and ovarian cancers. Nodes marked with yellow denote the top-10 genes with the highest degrees.

Table 1
The top-5 genes with the highest degrees in the partially shared subnetworks of different cancers.

Rank	Breast cancer	Ovarian cancer
1	PIK3CD	FZD9
2	CCND2	GTSE1
3	SFRP2	E2F1
4	PRR5	RRM2
5	MFNG	APC2

degrees. Among these genes, YWHAZ, ACTB, ACTG1, EIF4E, RHOA, and GNG5 have been shown to be closely related to a wide range of cancers and to be involved in a variety of cellular activities such as cell growth, apoptosis and migration/invasion [90–93–96,97]. HDAC2 mutant cells can upregulate tumor-promoting genes such as

mediators of cell cycle progression [98]. ITGB1 is overexpressed in tumor cells and is associated with angiogenesis, tumor progression and metastasis process [99,100]. Heterozygous missense mutations of PPP2R1A is commonly occur in human cancers [101]. Methylated HSP90AB1 promotes the proliferation of cancer cells [102].

For the partially shared subnetwork of cell lines belonging to the same cancer, Table 3 shows the top-5 genes with the highest degrees. In hepatocellular carcinoma (HCC), ATF2 is up-regulated in HCC patients, and by directly targeting miR-548p and controlling its expression, ATF2 can promote HCC cell proliferation, migration, and invasion and inhibited cell apoptosis [103]. Wang *et al.* [104] found that PPP2R5E is up-regulated in HCC patients who exhibit early recurrent disease. Liu *et al.* [105] suggested that PBRM1 is frequently mutated in HCC. GSK3B interacts with MYH9 to dysregulate the β-catenin destruction complex and induces tumor-related signal transduction in the downstream of HCC [106]. In chronic

Table 2
The top-5 genes with the highest degrees in different subtypes.

Rank	Breast cancer				Ovarian cancer			
	Basal-like	HER2-enriched	Luminal A	Luminal B	Differentiated	Immunoreactive	Mesenchymal	Proliferative
1	RB1	TLE3	EGFR	NOTCH1	NKD2	NCOA1	FGF2	HEY1
2	FZD4	ESR1	TLE4	MAML2	PPARD	NOTUM	ULK2	ESR1
3	CAB39	PRKAA1	ROR1	TLE4	WNT9B	FOSL1	TSC1	WDR59
4	RAC2	FGF2	CDK6	EGFR	FZD1	MAPK9	PPP3CC	HEYL
5	COP1	LPIN1	SFRP1	ROR1	CCNE1	EIF4E	IGF1	SIVA1

Fig. 9. The globally shared subnetwork that shared across all cell lines of CML and HCC cancers. Nodes marked with yellow denote the top-10 genes with the highest degrees.

Table 3
The top-5 genes with the highest degrees in the partially shared subnetworks of two cancers.

Rank	Hepatocellular carcinoma	Chronic myelogenous leukemia
1	ATF2	PPP3CA
2	PPP2R5E	GSTP1
3	IFNAR2	ADAM17
4	PBRM1	LAMC2
5	GSK3B	NQO1

Table 4
Top 5 nodes with the highest degrees in different cell lines.

	Hepatocellular carcinoma		Chronic myelogenous leukemia	
Rank	HuH-1	HuH-7	K562	KBM7
1	TBL1Y	JAK3	ARID2	RPS6KB1
2	FGF7	ITGA2B	INSR	E2F2
3	PDGFRA	PPP2R2B	CTBP1	ZNRF3
4	EFNA5	WNT5B	CSNK2B	AKT2
5	SYK	PLCG2	IGF1	RPS6

myelogenous leukemia (CML), Si *et al.* [107] provided a prospective regulatory pathway linking downregulated targets of miR-99a, such as PPP3CA, to the dysregulated vital genes related to leukemic cell expansion. Elhoseiny *et al.* [108] suggested that the GSTP1 gene polymorphism was associated with an increased risk of CML. Sopper *et al.* [109] found that ADAM17 enzymatic activity was significantly elevated in pretreatment of CML plasma samples. Xiao *et al.* [110] demonstrated that up-regulation of NQO1 reduces CML cellular proliferation.

For subnetworks that are specific to a particular cell line, the top-5 genes with the highest degree are analyzed. The results are shown

in Table 4. For the HuH-1 cell line in HCC, the down-regulation of TBL1Y produces a dramatic increase in apoptosis in HCC cells [111]. The expression of FGF7 was found to be increased in human HCC tissues [112]. PDGFRA overexpression in HCC is significantly associated with tumor size and weight [113]. EFNA5 is consistent with poorer survival in patients with HCC [114], and SYK has been proposed as a novel biomarker for HCC [115]. In the HuH-7 cells, expression of ITGA2B is significantly associated with overall survival or relapse-free survival in patients with HBV-related HCC [116]. EZH2 occupancy is associated with reduced expression of NKD1 and PPP2R2B compared to normal liver tissue [117]. WNT5B is up-

regulated on the CS and CSHA membranes in HuH-7 cells [118] and genes associated with cell death and transcriptional regulation such as PLCG2 tend to be up-regulated as well [119].

For the K562 cell line in chronic myelogenous leukemia, knockout of ARID2 can result in lower cell counts compared to normal cells [120]. High expression of CTBP1 in the CML K562 and lymphoblastic leukemia MOLT-4 cell lines is a common expression difference [121]. Aro *et al.* [122] found that autocrine production of IGF1 may play an important role in the growth of K562 cells via IGF-IR. In the KBM7 cell line, the mTORC1-S6K1 signaling pathway mediates the induction of downstream glycolysis [123]. AKT2 inhibits leukemic cell proliferation and induces apoptosis [124]. Clapper *et al.* [125] suggested that the protein levels of BCR-ABL, which translocation occurs in CML, are suppressed by mTOR, particularly via RPS6 during hypoxia.

The above results demonstrate that our model can efficiently identify the globally shared, partially shared and cancer-type-unique subnetworks of multiple cancer subtypes or cell lines collected from multiple cancer types. Our predicted networks can provide important insights into the common and specific regulatory mechanisms for different cancer types.

5. Conclusion

The development of high-throughput sequencing techniques has facilitated the inference of gene networks via computational approaches. Cancer is a heterogeneous disease and the same cancer type may contain numerous different subtypes or cancer cell lines, while different cancer types may have certain commonalities. Thus, the similarities between different cancer types may be heterogeneous and multi-layered. However, existing network inference methods typically ignore the multi-layer heterogeneity between different cancer types, which makes them unsuitable for joint estimation of multiple cancer-type-specific gene networks. In this study, we proposed a novel sparse regularized multi-layer decomposition graphical (SRMDG) model for estimating gene networks of multiple cancer types jointly. By decomposing each network into a combination of globally shared, partially shared, and cancer-type-unique components, our model can effectively identify the common and distinct components of multiple cancer-type-specific gene networks. Moreover, our model is able to handle non-Gaussian data with missing values, which makes it suitable for analyzing various types of gene expression data. Both simulation studies and real data analyses demonstrate the effectiveness of our model in jointly inferring multiple gene networks.

Although our model achieves good performance in multi-network inference, it is based on the assumption that all nodes have the same linking probability and all edges are independent, ignoring other prior information about gene regulatory interactions. However, a real biological network usually has a number of hub nodes whose degrees are much higher than the remaining nodes. Moreover, an growing amount of prior knowledge about gene regulatory interactions is accumulated. In the future, we will try to extend our model to deal with hub-driven gene dynamics and integrate various types of prior knowledge.

CRediT authorship contribution statement

YC conceived and designed the study, performed the statistical analysis and drafted the manuscript. LOY conceived of the study, and participated in its design and coordination and helped to draft the manuscript. XFZ participated in the design of the study and helped to revise the manuscript. All authors read and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.01.017.

References

[1] Fang L, Li Y, Ma L, Xu Q, Tan F, Chen G. Grndb: decoding the gene regulatory networks in diverse human and mouse conditions. Nucleic Acids Res 2021;49(D1):D97–103.

[2] Tian D, et al. Identifying gene regulatory network rewiring using latent differential graphical models. Nucleic Acids Res 2016;44(17). e140-e140.

[3] Van Der Wijst MG, de Vries DH, Brugge H, Westra H-J, Franke L. An integrative approach for building personalized gene regulatory networks for precision medicine. Genome Med 2018;10(1):1–15.

[4] Saeed MT, Ahmad J, Baumbach J, Pauling J, Shafi A, Paracha RZ, et al. Parameter estimation of qualitative biological regulatory networks on high performance computing hardware. BMC Syst Biol 2018;12(1):1–15.

[5] Liang K-C, Wang X. Gene regulatory network reconstruction using conditional mutual information. EURASIP J Bioinforma Syst Biol 2008;2008:1–14.

[6] Kim J. Validation and selection of ode models for gene regulatory networks. Chemom Intell Lab Syst 2016;157:104–10.

[7] Zhao H, Duan Z-H. Cancer genetic network inference using gaussian graphical models. Bioinforma Biol Insights 2019;13:1177932219839402.

[8] Azizi E, Prabhakaran S, Carr A, Pe'er D. Bayesian inference for single-cell clustering and imputing. Genom Comput Biol 2017;3(1). e46-e46.

[9] Hou W, Ji Z, Ji H, Hicks SC. A systematic evaluation of single-cell rna-sequencing imputation methods. Genome Biol 2020;21(1):1–30.

[10] Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 2015;16(3):133–45.

[11] Woodhouse S, Piterman N, Wintersteiger CM, Göttgens B, Fisher J. Scns: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. BMC Syst Biol 2018;12(1):1–7.

[12] W.V. Li, Y. Li, sclink: Inferring sparse gene co-expression networks from single-cell expression data, bioRxiv (2020).

[13] Danaher P, Wang P, Witten DM. The joint graphical lasso for inverse covariance estimation across multiple classes. J R Stat Soc Ser B, Stat Methodol 2014;76(2):373.

[14] Mohan K, London P, Fazel M, Witten D, Lee S-I. Node-based learning of multiple gaussian graphical models. J Mach Learn Res 2014;15(1):445–88.

[15] Zhang X-F, Ou-Yang L, Yan T, Hu XT, Yan H. A joint graphical model for inferring gene networks across multiple subpopulations and data types. IEEE Trans Cybern 2021;51(2):1043–55. https://doi.org/10.1109/TCYB.2019.2952711

[16] Robinson MD, McCarthy DJ, Smyth GK. edger: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26(1):139–40.

[17] Gallopin M, Rau A, Jaffrézic F. A hierarchical poisson log-normal model for network inference from rna sequencing data. PloS One 2013;8(10):e77503.

[18] Wu N, Yin F, Ou-Yang L, Zhu Z, Xie W. Joint learning of multiple gene networks from single-cell gene expression data. Comput Struct Biotechnol J 2020;18:2583–95.

[19] Zhang J, Zhang S. Discovery of cancer common and specific driver gene sets. Nucleic Acids Res 2017;45(10). e86-e86.

[20] Li WV, Li JJ. Modeling and analysis of rna-seq data: a review from a statistical perspective. Quant Biol 2018;6(3):195–209.

[21] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 2008;9(3):432–41.

[22] Hastie T, Tibshirani R, Wainwright M. Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC; 2019.

[23] Liu H, Lafferty J, Wasserman L. The nonparanormal: semiparametric estimation of high dimensional undirected graphs. J Mach Learn Res 2009;10(10).

[24] Xue L, Zou H. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. Ann Stat 2012;40(5):2541–71.

[25] Liu H, Han F, Yuan M, Lafferty J, Wasserman L. High-dimensional semiparametric gaussian copula graphical models. Ann Stat 2012;40(4):2293–326.

[26] Lafferty J, Liu H, Wasserman L. Sparse nonparametric graphical models. Stat Sci 2012;27(4):519–37.

[27] Wang H, Fazayeli F, Chatterjee S, Banerjee A. Gaussian copula precision estimation with missing values. Artificial Intelligence and Statistics. PMLR; 2014. p. 978–86. (pp.).

[28] Kruskal WH. Ordinal measures of association. J Am Stat Assoc 1958;53(284):814–61.

[29] Wang H, Fazayeli F, Chatterjee S, Banerjee A. Gaussian copula precision estimation with missing values. Artificial Intelligence and Statistics. PMLR; 2014. p. 978–86. (pp.).

[30] Boyd S, Parikh N, Chu E. Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc; 2011.

[31] Danaher P, Wang P, Witten DM. The joint graphical lasso for inverse covariance estimation across multiple classes. J R Stat Soc Ser B, Stat Methodol 2014;76(2):373.

[32] Mohan K, London P, Fazel M, Witten D, Lee S-I. Node-based learning of multiple gaussian graphical models. J Mach Learn Res 2014;15(1):445–88.

[33] Grimes T, Datta S. Seqnet: An r package for generating gene-gene networks and simulating rna-seq data. J Stat Softw 2021;98(12).

[34] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 2008;9(3):432–41.

[35] Chan TE, Stumpf MP, Babtie AC. Gene regulatory network inference from single-cell data using multivariate information measures. Cell Syst 2017;5(3):251–67.

[36] Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. Nat Genet 2013;45(10):1113–20.

[37] Kanehisa M, Goto S. Kegg: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30.

[38] W. Lin , C. Tong , W. Zhang , W. Cen , Y. Wang , J. Li et al., Silencing araf suppresses the malignant phenotypes of gallbladder cancer cells, BioMed Research International 2020.

[39] Pérez-Tenorio G, Karlsson E, Waltersson MA, Olsson B, Holmlund B, Nordenskjöld B, et al. Clinical potential of the mtor targets s6k1 and s6k2 in breast cancer. Breast Cancer Res Treat 2011;128(3):713–23.

[40] Qi Z, Zhang T, Song L, Fu H, Luo H, Wu J, et al. Pras40 hyperexpression promotes hepatocarcinogenesis. EBioMedicine 2020;51:102604.

[41] Ruan C, Ouyang X, Liu H, Li S, Jin J, Tang W, et al. Sin1-mediated mtor signaling in cell growth, metabolism and immune response. Natl Sci Rev 2019;6(6):1149–62.

[42] Steele FR, Chader GJ, Johnson LV, Tombran-Tink J. Pigment epithelium-derived factor: neurotrophic activity and identification as a member of the serine protease inhibitor gene family. Proc Natl Acad Sci 1993;90(4):1526–30.

[43] Wang X, Huang K, Zeng X, Liu Z, Liao X, Yang C, et al. Diagnostic and prognostic value of mrna expression of phospholipase c family genes in hepatitis b virus-associated hepatocellular carcinoma. Oncol Rep 2019;41(5):2855–75.

[44] Li X, Hu Y. Attribution of nf-b activity to chuk/ikk-involved carcinogenesis. Cancers 2021;13(6):1411.

[45] Xu Y, He B, Li R, Pan Y, Gao T, Deng Q, et al. Association of the polymorphisms in the fas/fasl promoter regions with cancer susceptibility: a systematic review and meta-analysis of 52 studies. PLoS One 2014;9(3):e90090.

[46] Chen J-S, Huang J-Q, Luo B, Dong S-H, Wang R-C, Jiang Z-k, et al. Pik 3 cd induces cell growth and invasion by activating akt/gsk-3/-catenin signaling in colorectal cancer. Cancer Sci 2019;110(3):997–1011.

[47] Evron E, Umbricht CB, Korz D, Raman V, Loeb DM, Niranjan B, et al. Loss of cyclin d2 expression in the majority of breast cancers is associated with promoter hypermethylation. Cancer Res 2001;61(6):2782–7.

[48] Veeck J, Noetzel E, Bektas N, Jost E, Hartmann A, Knüchel R, et al. Promoter hypermethylation of the sfrp2 gene is a high-frequent alteration and tumor-specific epigenetic marker in human breast cancer. Mol Cancer 2008;7(1):1–19.

[49] Johnstone CN, Castellví-Bel S, Chang LM, Sung RK, Bowser MJ, Piqué JM, et al. Prr5 encodes a conserved proline-rich protein predominant in kidney: analysis of genomic organization, expression, and mutation status in breast and colorectal carcinomas. Genomics 2005;85(3):338–51.

[50] Zhang S, Chung W-C, Wu G, Egan SE, Miele L, Xu K. Manic fringe promotes a claudin-low breast cancer phenotype through notch-mediated pik3cg induction. Cancer Res 2015;75(10):1936–43.

[51] Hoffmann K, Berger H, Kulbe H, Thillainadarasan S, Mollenkopf H-J, Zemojtel T, et al. Stable expansion of high-grade serous ovarian cancer organoids requires a low-wnt environment. EMBO J 2020;39(6):e104013.

[52] Hao Q, Li J, Zhang Q, Xu F, Xie B, Lu H, et al. Single-cell transcriptomes reveal heterogeneity of high-grade serous ovarian carcinoma. Clin Transl Med 2021;11(8):e500.

[53] Zhan L, Zhang Y, Wang W, Song E, Fan Y, Wei B. E2f1: a promising regulator in ovarian carcinoma. Tumor Biol 2016;37(3):2823–31.

[54] Reimer D, Sadr S, Wiedemair A, Stadlmann S, Concin N, Hofstetter G, et al. Clinical relevance of e2f family members in ovarian canceran evaluation in a training set of 77 patients. Clin Cancer Res 2007;13(1):144–51.

[55] Aird KM, Li H, Xin F, Konstantinopoulos PA, Zhang R. Identification of ribonucleotide reductase m2 as a potential target for pro-senescence therapy in epithelial ovarian cancer. Cell Cycle 2014;13(2):199–207.

[56] Jarrett CR, Blancato J, Cao T, Bressette D, Cepeda M, Young PE, et al. Human apc2 localization and allelic imbalance. Cancer Res 2001;61(21):7978–84.

[57] Herschkowitz JI, He X, Fan C, Perou CM. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal b breast carcinomas. Breast Cancer Res 2008;10(5):1–13.

[58] Yang L, Wu X, Wang Y, Zhang K, Wu J, Yuan Y, et al. Fzd7 has a critical role in cell proliferation in triple negative breast cancer. Oncogene 2011;30(43):4437–46.

[59] Horiuchi D, Kusdra L, Huskey NE, Chandriani S, Lenburg ME, Gonzalez-Angulo AM, et al. Myc pathway activation in triple-negative breast cancer is synthetic lethal with cdk inhibition. J Exp Med 2012;209(4):679–96.

[60] Basu GD, Ghazalpour A, Arguello D, Ashfaq R, Gatalica Z, McGinniss MJ, et al. Frequency of tle3 over-expression in breast carcinoma subtypes including a large cohort of triple negative patients. J Clin Oncol 2012.

[61] Urso L, Vernaci G, Carlet J, LoMele M, Fassan M, Zulato E, et al. Esr1 gene mutation in hormone receptor-positive her2-negative metastatic breast cancer patients: concordance between tumor tissue and circulating tumor dna analysis. Front Oncol 2021:403.

[62] Tolaney SM, Toi M, Neven P, Sohn J, Grischke E-M, Llombart-Cussac A, et al. Clinical significance of pik3ca and esr1 mutations in circulating tumor dna: analysis from the monarch 2 study of abemaciclib plus fulvestrantclinical significance of pik3ca and esr1 mutations in ctdna. Clin Cancer Res 2022:OF1–7.

[63] Newie I, Søkilde R, Persson H, Grabau D, Rego N, Kvist A, et al. The her2-encoded mir-4728-3p regulates esr1 through a non-canonical internal seed interaction. PloS One 2014;9(5):e97200.

[64] Rykala J, Przybylowska K, Majsterek I, Pasz-Walczak G, Sygut A, Dziki A, et al. The−553 t/a polymorphism in the promoter region of the fgf2 gene is associated with increased breast cancer risk in polish women. Arch Med Sci: AMS 2015;11(3):619.

[65] Myers E, Hill A, Kelly G, McDermott E, O'Higgins N, Young L. A positive role for pea3 in her2-mediated breast tumour progression. Br J Cancer 2006;95(10):1404–9.

[66] Matsui K, Sugimori K, Motomura H, Ejiri N, Tsukada K, Kitajima I. Pea3 co-operates with c-jun in regulation of her2/neu transcription. Oncol Rep 2006;16(1):153–8.

[67] Monaco ME. Fatty acid metabolism in breast cancer subtypes. Oncotarget 2017;8(17):29487.

[68] Darbeheshti F, Izadi P, Razavi ANE, Kamali F, Yekaninejad MS, Bazzaz JT. Significance of egfr mrna expression in luminal and triple negative breast tumors. Int J Cancer Manag 2018;11(2).

[69] N. Borcherding, M. Ameka, R. Kolb, Q. Xie, W. Zhang, Wnt5a/ror1 axis in triple-negative breast cancer progression and potential therapy (2014).

[70] Finn RS, Crown JP, Lang I, Boer K, Bondarenko IM, Kulyk SO, et al. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, her2-negative, advanced breast cancer (paloma-1/trio-18): a randomised phase 2 study. Lancet Oncol 2015;16(1):25–35.

[71] S. Mamoor, Differential expression of sfrp1 in cancers of the breast. (2021).

[72] Haughian JM, Pinto MP, Harrell JC, Bliesner BS, Joensuu KM, Dye WW, et al. Maintenance of hormone responsiveness in luminal breast cancers by suppression of notch. Proc Natl Acad Sci 2012;109(8):2742–7.

[73] Lubecka K, Kurzava L, Flower K, Buvala H, Zhang H, Teegarden D, et al. Stilbenoids remodel the dna methylation patterns in breast cancer cells and inhibit oncogenic notch signaling through epigenetic regulation of maml2 transcriptional activity. Carcinogenesis 2016;37(7):656–68.

[74] Wei W, Zheng L, Gao Y, He M, Yang F. Expression and prognostic significance of nkd2 in ovarian cancer. Jpn J Clin Oncol 2021;51(3):459–68.

[75] Schumann T, Adhikary T, Wortmann A, Finkernagel F, Lieber S, Schnitzer E, et al. Deregulation of ppar/Ä target genes in tumor-associated macrophages by fatty acid ligands in the ovarian cancer microenvironment. Oncotarget 2015;6(15):13416.

[76] Gatcliffe T, Monk B, Planutis K, Holcombe R. Wnt signaling in ovarian tumorigenesis. Int J Gynecol Cancer 2008;18(5).

[77] Davidson B, Skrede M, Silins I, Shih I-M, Trope CG, Flørenes VA. Low-molecular weight forms of cyclin e differentiate ovarian carcinoma from cells of mesothelial origin and are associated with poor survival in ovarian carcinoma. Cancer 2007;110(6):1264–71.

[78] Huang Y, Liu Y, Zhu K, Ma X, Lu R, Zhang M. Gsg2 promotes development and predicts poor prognosis of ovarian cancer. Cancer Manag Res 2021;13:499.

[79] Wan J, Shi F, Xu Z, Zhao M. Knockdown of eif4e suppresses cell proliferation, invasion and enhances cisplatin cytotoxicity in human ovarian cancer cells. Int J Oncol 2015;47(6):2217–25.

[80] Cecco LD, Marchionni L, Gariboldi M, Reid JF, Lagonigro MS, Caramuta S, et al. Gene expression profiling of advanced ovarian cancer: characterization of a molecular signature involving fibroblast growth factor 2. Oncogene 2004;23(49):8171–83.

[81] Wang Y, Zhang X, Tang W, Lin Z, Xu L, Dong R, et al. mir-130a upregulates mtor pathway by targeting tsc1 and is transactivated by nf-b in high-grade serous ovarian carcinoma. Cell Death Differ 2017;24(12):2089–100.

[82] Lau M-T, Leung PC. The pi3k/akt/mtor signaling pathway mediates insulin-like growth factor 1-induced e-cadherin down-regulation and cell proliferation in ovarian cancer cells. Cancer Lett 2012;326(2):191–8.

[83] Poljicanin A, Filipovic N, Pusic TV, Soljic V, Caric A, Saraga-Babic M, et al. Expression pattern of rage and igf-1 in the human fetal ovary and ovarian serous carcinoma. Acta Histochem 2015;117(4–5):468–76.

[84] Rovani MT, Gasperin BG, Ilha GF, Ferreira R, Bohrer RC, Duggavathi R, et al. Expression and molecular consequences of inhibition of estrogen receptors in granulosa cells of bovine follicles. J Ovarian Res 2014;7(1):1–9.

[85] Lavery DN, Villaronga MA, Walker MM, Patel A, Belandia B, Bevan CL. Repression of androgen receptor activity by heyl, a third member of the hairy/enhancer-of-split-related family of notch effectors. J Biol Chem 2011;286(20):17796–808.

[86] Ma Y, Liu T, Song X, Tian Y, Wei Y, Wang J, et al. Siva 1 inhibits proliferation, migration and invasion by phosphorylating stathmin in ovarian cancer cells. Oncol Lett 2017;14(2):1512–8.

[87] Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CA, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. Nat Med 2017;23(6):692–702.

[88] Kind J, Pagie L, de Vries SS, Nahidiazar L, Dey SS, Bienko M, et al. Genome-wide maps of nuclear lamina interactions in single human cells. Cell 2015;163(1):134–47.

[89] Zheng H, Pomyen Y, Hernandez M, Li C, Livak F, Greten T, et al. Abstract 936: Single cell analysis reveals cancer stem cell heterogeneities in hepatocellular carcinoma. Cancer Res 2017;77. 936-936.

[90] Gan Y, Ye F, He X-X. The role of ywhaz in cancer: a maze of opportunities and challenges. J Cancer 2020;11(8):2252.

[91] Guo C, Liu S, Wang J, Sun M-Z, Greenaway FT. Actb in cancer. Clin Chim Acta 2013;417:39–44.

[92] Dugina V, Khromova N, Rybko V, Blizniukov O, Shagieva G, Chaponnier C, et al. Tumor promotion by γ and suppression by β non-muscle actin isoforms. Oncotarget 2015;6(16):14556.

[93] Topisirovic I, Ruiz-Gutierrez M, Borden KL. Phosphorylation of the eukaryotic translation initiation factor eif4e contributes to its transformation and mrna transport activities. Cancer Res 2004;64(23):8639–42.

[94] Wendel H-G, Silva RL, Malina A, Mills JR, Zhu H, Ueda T, et al. Dissecting eif4e action in tumorigenesis. Genes Dev 2007;21(24):3232–7.

[95] Jeong D, Park S, Kim H, Kim C-J, Ahn TS, Bae SB, et al. Rhoa is associated with invasion and poor prognosis in colorectal cancer. Int J Oncol 2016;48(2):714–22.

[96] Orchel J, Witek L, Kimsa M, Strzalka-Mrozik B, Kimsa M, Olejek A, et al. Expression patterns of kinin-dependent genes in endometrial cancer. Int J Gynecol Cancer 2012;22(6).

[97] Zhang W, Liu Z, Liu B, Jiang M, Yan S, Han X, et al. Gng5 is a novel oncogene associated with cell migration, proliferation, and poor prognosis in glioma. Cancer Cell Int 2021;21(1):1–20.

[98] Ropero S, Ballestar E, Alaminos M, Arango D, Schwartz S, Esteller M. Transforming pathways unleashed by a hdac2 mutation in human cancer. Oncogene 2008;27(28):4008–12.

[99] Guo G, Gong K, Wohlfeld B, Hatanpaa KJ, Zhao D, Habib AA. Ligand-independent egfr signaling. Cancer Res 2015;75(17):3436–41.

[100] Hu C, Ni Z, Li B-s, Yong X, Yang X, Zhang J-w, et al. htert promotes the invasion of gastric cancer cells by enhancing foxo3a ubiquitination and subsequent itgb1 upregulation. Gut 2017;66(1):31–42.

[101] Haesen D, Asbagh LA, Derua R, Hubert A, Schrauwen S, Hoorne Y, et al. Recurrent ppp2r1a mutations in uterine cancer act through a dominant-negative mechanism to promote malignant cell growth. Cancer Res 2016;76(19):5719–31.

[102] Hamamoto R, Toyokawa G, Nakakido M, Ueda K, Nakamura Y. Smyd2-dependent hsp90 methylation promotes cancer cell proliferation by regulating the chaperone complex formation. Cancer Lett 2014;351(1):126–33.

[103] Li Z-J, Zhang J-P, Li D-Y, Yang H-Y, Liu B-R. Atf2 accelerates the invasion and metastasis of hepatocellular carcinoma through targeting the mir-548p/tuft1 axis. Hepatol Res: J Jpn Soc Hepatol 2023.

[104] Wang SM, Ooi LLP, Hui KM. Upregulation of rac gtpase-activating protein 1 is significantly associated with the early recurrence of human hepatocellular carcinoma. Clin Cancer Res 2011;17(18):6040–51.

[105] Liu Z-H, Lian B-F, Dong Q-Z, Sun H, Wei J-W, Sheng Y-Y, et al. Whole-exome mutational and transcriptional landscapes of combined hepatocellular cholangiocarcinoma and intrahepatic cholangiocarcinoma reveal molecular diversity. Biochim Et Biophys Acta (BBA Mol Basis Dis 2018;1864(6):2360–8.

[106] Lin X, Li A-m, Li Y-H, Luo R-C, Zou Y-J, Liu Y-Y, et al. Silencing myh9 blocks hbx-induced gsk3β ubiquitination and degradation to inhibit tumor stemness in hepatocellular carcinoma. Signal Transduct Target Ther 2020;5(1):1–12.

[107] Si X, Zhang X, Hao X, Li Y, Chen Z, Shi H, et al. Upregulation of mir-99a is associated with poor prognosis of acute myeloid leukemia and promotes myeloid leukemia cell expansion. Blood 2016;128(22):5086.

[108] Elhoseiny S, El-Wakil M, Fawzy M, Rahman AA, et al. Gstp1 (ile105val) gene polymorphism: risk and treatment response in chronic myeloid leukemia. J Cancer Ther 2013;5(01):1.

[109] Sopper S, Mustjoki S, White D, Hughes T, Valent P, Burchert A, et al. Reduced cd62l expression on t cells and increased soluble cd62l levels predict molecular response to tyrosine kinase inhibitor therapy in early chronic-phase chronic myelogenous leukemia. J Clin Oncol 2017;35(2):175–84.

[110] Xiao F-Y, Jiang Z-P, Yuan F, Zhou F-J, Kuang W, Zhou G, et al. Down-regulating nqo1 promotes cellular proliferation in k562 cells via elevating dna synthesis. Life Sci 2020;248:117467.

[111] Wei L, Lee D, Law C-T, Zhang MS, Shen J, Chin DW-C, et al. Genome-wide crispr/cas9 library screening identified phgdh as a critical driver for sorafenib resistance in hcc. Nat Commun 2019;10(1):1–13.

[112] Amann T, Bataille F, Spruss T, Dettmer K, Wild P, Liedtke C, et al. Reduced expression of fibroblast growth factor receptor 2iiib in hepatocellular carcinoma induces a more aggressive growth. Am J Pathol 2010;176(3):1433–42.

[113] Wei T, Zhang L-N, Lv Y, Ma X-Y, Zhi L, Liu C, et al. Overexpression of platelet-derived growth factor receptor alpha promotes tumor progression and indicates poor prognosis in hepatocellular carcinoma. Oncotarget 2014;5(21):10307.

[114] Husain A, Chiu Y-T, Sze KM-F, Ho DW-H, Tsui Y-M, Suarez EMS, et al. Ephrin-a3/epha2 axis regulates cellular metabolic plasticity to enhance cancer stemness in hypoxic hepatocellular carcinoma. J Hepatol 2022.

[115] Hong J, Yuan Y, Wang J, Liao Y, Zou R, Zhu C, et al. Expression of variant isoforms of the tyrosine kinase syk determines the prognosis of hepatocellular carcinoma. Cancer Res 2014;74(6):1845–56.

[116] Shang L, Ye X, Zhu G, Su H, Su Z, Chen B, et al. Prognostic value of integrin variants and expression in post-operative patients with hbv-related hepatocellular carcinoma. Oncotarget 2017;8(44):76816.

[117] Cheng AS, Lau SS, Chen Y, Kondo Y, Li MS, Feng H, et al. Ezh2-mediated concordant repression of wnt antagonists promotes β-catenin–dependent hepatocarcinogenesis. Cancer Res 2011;71(11):4028–39.

[118] Chang P-H, Sekine K, Chao H-M, Hsu S-h, Chern E. Chitosan promotes cancer progression and stem cell properties in association with wnt signaling in colon and hepatocellular carcinoma cells. Sci Rep 2017;7(1):1–14.

[119] Cervello M, Bachvarov D, Lampiasi N, Cusimano A, Azzolina A, McCubrey JA, et al. Novel combination of sorafenib and celecoxib provides synergistic antiproliferative and pro-apoptotic effects in human liver cancer cells. PloS One 2013;8(6):e65569.

[120] Sakai H, Hosono N, Nakazawa H, Przychodzen B, Polprasert C, Carraway HE, et al. A novel genetic and morphologic phenotype of arid2-mediated myelodysplasia. Leukemia 2018;32(3):839–43.

[121] Sewalt RG, Gunster MJ, van der Vlag J, Satijn DP, Otte AP. C-terminal binding protein is a transcriptional repressor that interacts with a specific class of vertebrate polycomb proteins. Mol Cell Biol 1999;19(1):777–87.

[122] Aro AL, Savikko J, Pulkkinen V, von Willebrand E. Expression of insulin-like growth factors igf-i and igf-ii, and their receptors during the growth and megakaryocytic differentiation of k562 cells. Leuk Res 2002;26(9):831–7.

[123] Barger JF, Gallo CA, Tandon P, Liu H, Sullivan A, Grimes HL, et al. S6k1 determines the metabolic requirements for bcr-abl survival. Oncogene 2013;32(4):453–61.

[124] Li H, Liu L, Zhuang J, Liu C, Zhou C, Yang J, et al. Identification of key candidate targets and pathways for the targeted treatment of leukemia stem cells of chronic myelogenous leukemia using bioinformatics analysis. Mol Genet Genom Med 2019;7(9):e851.

[125] Clapper E, Di Trapani G, Tonissen KF. The regulation of bcr-abl in hypoxia is through the mtor pathway. Leuk Lymphoma 2021;62(4):967–78.