



HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2022 January 01.

Published in final edited form as:

J Am Stat Assoc. 2021 ; 116(534): 605–618. doi:10.1080/01621459.2020.1775611.

Bayesian Structure Learning in Multi-layered Genomic Networks

Min Jin Ha¹, Francesco Claudio Stingo², Veerabhadran Baladandayuthapani³

¹Department of Biostatistics, The University of Texas MD Anderson Cancer Center

²Department of Statistics, Computer Science, Applications “G. Parenti”, The University of Florence

³Department of Biostatistics, University of Michigan

Abstract

Integrative network modeling of data arising from multiple genomic platforms provides insight into the holistic picture of the interactive system, as well as the flow of information across many disease domains including cancer. The basic data structure consists of a sequence of hierarchically ordered datasets for each individual subject, which facilitates integration of diverse inputs, such as genomic, transcriptomic, and proteomic data. A primary analytical task in such contexts is to model the layered architecture of networks where the vertices can be naturally partitioned into ordered layers, dictated by multiple platforms, and exhibit both undirected and directed relationships. We propose a multi-layered Gaussian graphical model (mlGGM) to investigate conditional independence structures in such multi-level genomic networks in human cancers. We implement a Bayesian node-wise selection (BANS) approach based on variable selection techniques that coherently accounts for the multiple types of dependencies in mlGGM; this flexible strategy exploits edge-specific prior knowledge and selects sparse and interpretable models. Through simulated data generated under various scenarios, we demonstrate that BANS outperforms other existing multivariate regression-based methodologies. Our integrative genomic network analysis for key signaling pathways across multiple cancer types highlights commonalities and differences of p53 integrative networks and epigenetic effects of BRCA2 on p53 and its interaction with T68 phosphorylated CHK2, that may have translational utilities of finding biomarkers and therapeutic targets.

Keywords

Multi-level data integration; Multi-layered Gaussian graphical models; Bayesian variable selection

1 Introduction

Cancer is a complex disease that is caused by deregulation of several molecular processes and cellular pathways, usually triggered by genetic alterations in specific sets of genes (Hanahan and Weinberg, 2000, 2011; Creixell et al., 2012). Pathway and network analysis aim to gain insight into these underlying interactive mechanisms, and to reduce data

Corresponding author Min Jin Ha MJHa@mdanderson.org.

involving thousands of altered genes and proteins to a smaller, and more interpretable set of functionally altered processes (Pe'er and Hachohen, 2011). The Cancer Genome Atlas (TCGA) has generated a rich source of multi-dimensional genomics (omics for short) data for patients across multiple tumor types and their subtypes (<http://cancergenome.nih.gov>). More recently, The Cancer Proteome Atlas (TCPA) (<http://bioinformatics.mdanderson.org/main/TCPA:Overview>) has generated a complementary set of reverse phase protein array (RPPA)-based proteomic data across most of these patients' samples, covering major oncogenic signaling pathways (Li et al., 2013; Akbani et al., 2014). Multi-level integration and processing of network information across these modalities is emerging, based on the principle that any biological mechanism is a systematic conflation of multiple molecular events and their interactions (Kristensen et al., 2014).

The basic data structure constitutes of a finite sequence of hierarchically-ordered datasets for each individual subject, which facilitates integration of diverse inputs such as genomic, transcriptomic, and proteomic data to investigate the unified regulatory mechanisms underlying the biology of various cancers. Figure 1 shows a conceptual data structure that allows for the characterization of dependencies within and between the datasets from the ordered layers. Two types of edges characterize the network structure: undirected edges within each layer, and directed edges to distinguish variables in a layer from variables in all the previous layers. Most integrative analyses rely on directed relationships between different data platforms based on the biological mechanisms (Wang et al., 2013), and the variables observed from each platform constitute a layer. For example, following the central dogma of biology, the first layer could constitute DNA-level data (copy number, methylation), the second layer the transcriptomic (mRNA expression), and the third layer proteomic data. In this context, the directed edges capture cross-platform (e.g., transcriptional, translational) dependencies, and the undirected edges capture the within-platform dependencies.

Modeling each of the layers independently does not account for the hierarchical multi-layered structure of the data, and can only provide information on how networks operate at a static point in time or under a static condition. A critical next step is to understand a holistic and dynamic picture of the interactive system and the information flow, which can only be achieved by simultaneously modeling multiple layers of data. Our objective is to investigate the conditional dependencies among the variables from multiple layers, while accounting for the order defined by the underlying layered structure of the data. Statistically, this translates to a structural estimation of graphs with a mixture of directed and undirected edges. This objective poses significant methodological and computational challenges, however, to building a single graphical model that includes hierarchical multiple graphs with directed edges for dependencies, constrained by orders between layers, as well as with undirected edges for unconstrained dependencies within layers.

Chain graphs have been used to model the layered architecture of networks, where the vertices can be naturally partitioned into ordered sets that exhibit undirected and directed acyclic relations within and between the sets. The conditional independencies that correspond to missing edges (i.e., Markov properties in the chain graphs) have been studied by Lauritzen and Wermuth (1989); Frydenberg (1990); Andersson et al. (2001). Drton and

Eichler (2006) introduced a point-estimation approach to maximum likelihood estimation given the graph structure, and Drton and Perlman (2008) proposed a constraint-based method to estimate the structure. However, all of the above-mentioned methods for chain graphs are restricted to low-dimensional data with sample sizes larger than the number of vertices.

In its simplest form (two layers), a chain graph is equivalent to a multiple multivariate regression model. Multiple predictors affect multiple responses that exhibit a correlation structure, and both regression coefficients (for directed edges) and the error precision matrix (for undirected edges) are assumed to be typically sparse. Approaches based on the doubly-penalized joint likelihood have been proposed by Rothman et al. (2010); Yin and Li (2011). A joint L_1 penalty was imposed on both the regression coefficients and the precision matrix, resulting in a penalized likelihood that is bi-convex (but not convex), which implies that the optimization algorithm may be unstable and fail to converge (Lee and Liu, 2012). Cai et al. (2012); Chen et al. (2016) proposed a two-step approach, in which only regression coefficients are estimated in the first step, and then the error inverse covariance matrix is obtained in the second step, given the estimated regression parameters. To select good initial parameters for this bi-convex problem, Lin et al. (2016) proposed an L_1 penalized maximum likelihood estimation with prescreening of variables, and extended this methodology to multi-layered graphs with an arbitrary number of layers. In a Bayesian framework, Bhadra and Mallick (2013); Consonni et al. (2017) proposed a Bayesian model based on the hyper-inverse Wishart prior on the covariance matrix, which assumes that the UGs within each layer are decomposable. The decomposability assumption is generally restrictive, explores a smaller model space, and potentially provides misspecification of the structure for networks in many real applications.

In this article, we propose a novel approach for a multi-layered Gaussian graphical model (mlGGM) that accommodates data from an arbitrary number of layers. The mlGGM is a special case of the chain graph model, where the random variables that correspond to the vertices are assumed to follow a multivariate Gaussian distribution, and the zero structures of the mean parameter and the precision parameter directly link to the absence of the directed and undirected edges, respectively. We construct a regression-based formulation that converts the mlGGM into a more tractable node-wise multiple regression model. We generalize the node-wise multiple regression approach that coherently accounts for conditional independencies in mlGGMs. We jointly select both undirected and directed edges that point to a vertex via Bayesian variable selection priors for each of the regressions, allowing for a much larger graph space than that of decomposable graphs. Moreover, the prior formulation allows for the incorporation of relevant prior knowledge through the edge-specific informative prior, and provides a computationally more efficient procedure to estimate mlGGMs.

Through simulation studies under various settings of the mlGGMs, we demonstrate the utility of our node-wise regression framework in the structural recovery of the mlGGM, and compare its performance to those of related multivariate regression-based methods. We also numerically evaluate sign consistency (i.e., whether our univariate regression-based formulation correctly finds the signs of the undirected and directed dependencies when we

have the known structure). We illustrate the applicability and versatility of mlGGMs to infer integrated genomic networks for multi-omic data, using biological hierarchies among the platforms across multiple cancers. The signed topological structure obtained from our method allows for more refined inference of both cross- and within-platform dependencies, including the inhibition/activation between platforms and positive/negative correlations within platforms. Furthermore, we investigate the commonalities and differences in their multi-layered network structures across cancer types, and show translational utilities of these integrative networks.

The rest of this article is organized as follows. Section 2 provides data structure and backgrounds on mlGGMs. Section 3 presents the joint model and the corresponding prior construction. We introduce our Bayesian node-wise selection (BANS) framework in Section 4. We discuss posterior inference on graphical structure estimation in Section 5. Simulation studies are carried out in Section 6. We demonstrate the utility of our method in multi-layered genomic network studies across cancer types in Section 7. Section 8 concludes the article.

2 Background

2.1 Data structure

We consider a graphical model over p biological units across multi-omic data, $\mathbf{Y} = (Y_1, \dots, Y_p)^T \in \mathbb{R}^p$. Here p could constitute different platform-specific observations (e.g., genes, proteins, etc.), but for ease of conceptual and technical exposition, we use genes throughout. A graph of \mathbf{Y} can be denoted by $G = (V, E)$, where $V = \{1, \dots, p\}$ includes all genes across platforms, and E may contain both directed (\rightarrow) and undirected edges ($-$) between the genes. We assume to know *a priori* that a partitioning $T = \{\tau_k \mid 1 \leq k \leq q\}$, $q \leq p$, is a family of pairwise disjoint *ordered* layers of p genes. The ordered partitioning T implies that any edges between the layers are directed. In other words, each layer τ_k constitutes genes from a platform, and the layers are ordered according to a biological hierarchy. Formally, the partitioning T is called a *dependence chain* if $k < l \Rightarrow \forall u \in \tau_k, v \in \tau_l$. In other words, when $k < l$, any edges between τ_k and τ_l point from a vertex in τ_k to a vertex in τ_l . For each $v \in V$, let $1 \leq \ell(v) \leq q$ be the index, such that $v \in \tau_{\ell(v)}$. We assume without loss of generality that the vertices are labeled, such that, $\ell(u) < \ell(v) \Rightarrow u < v$. Factorization based on biological hierarchies. The integrative genomic analyses for multi-platform genomics data can be performed based on coherent biological justifications motivated by their hierarchical dependencies. Following the central dogma of biology, where the epigenetic and DNA level, such as methylation and copy number variation, potentially regulate mRNA expression (transcription regulation), which in turn is known to regulate protein expressions (translation regulation) (Morris and Baladandayuthapani, 2017). For our case study, we integrate four datasets from copy number aberration (CNA), methylation, mRNA expression, and protein expression, and the dependence chain (T) constitutes four (q) disjoint sets that have their own unique order: (CNA, methylation) < mRNA < protein (see Figure S14)

Notationally, if $u \rightarrow v$, the vertex u is a *parent* of v . Let $\text{pa}_v = \{u \in V : u \rightarrow v \in E\}$ and $\text{pa}_A = \bigcup_{v \in A} \text{pa}_v$, A be the sets of parents of v and a subset $A \subseteq V$, respectively. We assume the dependence chain allows for factorization of DAGs at the layer level (i.e., the graph in

Figure 1 is a DAG of the dependence chain, τ_1, \dots, τ_q). We denote the sub-vector of \mathbf{Y} corresponding to a subset ACV by Y_A . The joint probability distribution of \mathbf{Y} can be factorized as

$$P(\mathbf{Y}) = \prod_{\tau \in T} P(Y_\tau | Y_{\text{pa}_\tau}) \quad (1)$$

(Lauritzen, 1996). Through factorization, the network structure in Figure 1 can be viewed as $q-1$ two-layered models, such as $\{Y_{\tau_m} : m = 1, \dots, k-1\}$ versus Y_{τ_k} for $k > 1$, plus a one-layered UG model for Y_{τ_1} . In the example of using DNA methylation, CNA, mRNA expression and protein expression, the multi-layered networks can be constructed from two UG models for CNA and methylation, and two two-layered models for mRNA and protein expressions.

To complete our chain graph model, we specify the Markov property. For a dependence chain T , we define the *cumulatives* to be the set $C_l = \bigcup_{k=1}^l \tau_k$ for $1 \leq l \leq q$. Andersson et al. (2001) proposed a Markov property for chain graphs and set the pairwise Markov property for G as

$$\begin{aligned} u - v \notin E &\Rightarrow Y_u \perp Y_v \mid Y_{C_{t(v)}, \{u, v\}} \text{ for } t(u) = t(v) \\ u \rightarrow v \notin E &\Rightarrow Y_u \perp Y_v \mid Y_{C_{t(v)-1}, \{u\}} \text{ for } t(u) < t(v). \end{aligned} \quad (2)$$

A missing directed edge between two random variables Y_u and Y_v implies that they are conditionally independent, given all other variables in $\tau_1, \dots, \tau_{k(v)-1}$, while the conditional set of missing undirected edges is all other variables in $\tau_1, \dots, \tau_{k(v)}$.

2.2 Multi-layered Gaussian graphical models

In this section, we specify each factor in (1). We assume that the $p \times 1$ random vector \mathbf{Y} follows the multivariate Gaussian distribution $N(\boldsymbol{\theta}, \boldsymbol{\Omega}^{-1})$, with the positive definite precision matrix $\boldsymbol{\Omega}$. The mlGGM $G = (V, E)$, following the Markov property stated in (2), is

$$\mathbf{Y} = \mathbf{B}\mathbf{Y} + \boldsymbol{\delta}, \boldsymbol{\delta} \sim N(\boldsymbol{\theta}, \mathbf{K}^{-1}), \quad (3)$$

where $\mathbf{B} = (b_{vu})$ is a $p \times p$ matrix with $u \rightarrow v \notin E \Leftrightarrow b_{vu} = 0$, and $\mathbf{K} = (\kappa_{vu})$ is a positive definite $p \times p$ matrix with $v - u \notin E \Leftrightarrow \kappa_{uv} = \kappa_{vu} = 0$. Then the precision matrix of \mathbf{Y} is

$$\boldsymbol{\Omega} = (\mathbf{I} - \mathbf{B})^T \mathbf{K} (\mathbf{I} - \mathbf{B}), \quad (4)$$

where \mathbf{I} is an identity matrix. \mathbf{B} is a coefficient matrix for which the zero structures encode the directed edges between layers. The precision matrix of the error $\boldsymbol{\delta}$, \mathbf{K} is a symmetric matrix, where the nonzero off-diagonal elements represent the undirected edges within a layer after taking out the effects from the directed edges, and the i th diagonal element is the inverse variance of Y_i .

The mlGGM in (3) for a dependence chain T with $|T| = q > 1$ can be expressed as one GGM for τ_1 and $q-1$ two-layered GGMs by the factorization in (1). For a matrix $\mathbf{A} = (a_{ij})$, we

denote sub-matrices $\mathbf{A}_{S_1, S_2} = (a_{ij})_{i \in S_1, j \in S_2}$ and $\mathbf{A}_S = (a_{ij})_{i \in S, j \in S}$. We also denote the transpose of a sub-matrix, $\mathbf{A}_{S_1, S_2}^T = (\mathbf{A}_{S_2, S_1})^T$. Under the factorization in equation (1), the model in (3) can be re-expressed as the component-wise conditional distributions:

$$\mathbf{Y}_\tau \mid \mathbf{Y}_{\text{pa}_\tau} \sim N(\mathbf{B}_{\tau, \text{pa}_\tau} \mathbf{Y}_{\text{pa}_\tau}, \mathbf{K}_\tau^{-1}) \text{ for all } \tau \in \mathcal{T}. \quad (5)$$

Because the first layer τ_1 has an empty parent set, it has zero mean in (5) and is equivalent to the UG. The conditional distribution corresponds to a multivariate multiple regression, where the block of variables \mathbf{Y}_τ is regressed on the parents $\mathbf{Y}_{\text{pa}_\tau}$.

There are other formulations of chain graphs based on the Lauritzen-Wermuth-Frydenberg (LWF) Markov property (Lauritzen and Wermuth, 1989; Frydenberg, 1990). In the case of continuous variables with a joint multivariate Gaussian distribution, the Markov property in (2) is coherent with data generated by the linear system in (5) (Cox and Wermuth, 1993; Andersson et al., 2001; Drton and Eichler, 2006). Further details on implication of Markov properties are in Section S1.

3 Joint model and prior construction

In this section, we discuss the estimation problem of the multi-layered graph structure that includes both directed and undirected edges, which induces network-based integration of multi-omic data. Under known \mathcal{T} , our focus is to estimate the zero-structures of \mathbf{B} and \mathbf{K} that encode the directed edges and undirected edges between and within layers, respectively. For directed edges, the problem boils down to finding the parents, pa_v for $v \in \tau_k \subseteq V$ from the cumulative C_{k-1} , which is the union of all the preceding layers of τ_k . The model corresponding to the first layer, τ_1 , is given by

$$\mathbf{Y}_{\tau_1} \sim N(\mathbf{0}, \mathbf{K}_{\tau_1}^{-1}), \quad (6)$$

where there are no predictors. From the second component to the last component, τ_2, \dots, τ_q with $|\mathcal{T}| = q$, the structure of the model is expressed by the following multivariate multiple regressions:

$$\mathbf{Y}_{\tau_k} = \mathbf{B}_{\tau_k, C_{k-1}} \mathbf{Y}_{C_{k-1}} + \delta_{\tau_k}, \text{ and } \delta_{\tau_k} \sim N(\mathbf{0}, \mathbf{K}_{\tau_k}^{-1}), k = 2, \dots, q, \quad (7)$$

where C_k is the cumulative for the k th layer. The model for the first layer only involves the parameter for the precision matrix \mathbf{K}_{τ_1} , which is a traditional model for UGs (Lauritzen, 1996). The parameters of interest are the coefficient matrices $\{\mathbf{B}_{\tau_k, C_{k-1}} : k = 2, \dots, q\}$, which encode the directed edges across the layers, and $\{\mathbf{K}_{\tau_k} : k = 1, \dots, q\}$, which encode the conditional dependencies within the layers, after adjusting for the effects from the directed edges. Fitting the set of regression models can be performed independently and in parallel, assuming priors over the parameters that are independent across the regressions. Component-wise regressions. In this component-wise regression framework, we can use the

Inverse Wishart prior for $\mathbf{K}_{\tau_k}^{-1}$ and the independent normal priors for elements of $\mathbf{B}_{\tau_k, C_{k-1}}$:
for all $k = 1, \dots, q$,

$$\begin{aligned} \mathbf{K}_{\tau_k}^{-1} &\sim \text{InverseWishart} | \tau_k | (\delta_{\tau_k}, \lambda_{\tau_k} \mathbf{I} | \tau_k |), \\ b_{vw} &\sim N(0, c_{vw}^2 / \kappa_{vw}) \text{ for all } v \in \tau_k \text{ and } w \in C_{k-1}, \end{aligned} \quad (8)$$

where $\lambda_{\tau_k} > 0$ and $\delta_{\tau_k} > 0$ for all $k = 1, \dots, q$, and c_{vw} are constants. The prior of each regression coefficient b_{vw} is dependent on the precision κ_{vw} of the variable Y_v (v th diagonal element of \mathbf{K}).

In essence, we are able to divide the estimation problem of the mlGGM in (3) into smaller multivariate regression problems by specifying the priors that are independent across $\tau \in \mathcal{T}$ in (8). The Inverse-Wishart distribution is the global conjugate prior for covariance matrices. However, the number of parameters greatly increases as the total number of variables from multiple layers increases. For estimating the two-layered GGM for the k th component, τ_k , we have $|\tau_k| |C_{k-1}| + |\tau_k| (|\tau_k| + 1) / 2$ number of parameters for the directed edges, undirected edges, and the variances of the variables in τ_k . The number of parameters is 3775, even for $|\tau_k| = 50$ and $|C_{k-1}| = 50$. Moreover, the number of parameters becomes larger when we handle downstream layers, due to the increasing number of preceding layers. Bhadra and Mallick (2013) and Consonni et al. (2017) proposed a joint selection of the nonzero elements of $\mathbf{B}_{\tau_k, C_{k-1}}$ and \mathbf{K}_{τ_k} , where the undirected relations encoded in \mathbf{K}_{τ_k} correspond to the limited space of decomposable graphs using the hyper-inverse Wishart prior (Dawid and Lauritzen, 1993). Moreover, the method selects entire columns of the coefficient matrix $\mathbf{B}_{\tau_k, C_{k-1}}$, and fails to select single elements of this matrix (i.e., a variable in an upper layer is either connected to all variables in a lower layer or to none), and the same precision matrix \mathbf{K}_{τ_k} informs the selection of the coefficient matrix $\mathbf{B}_{\tau_k, C_{k-1}}$. Assumptions on both $\mathbf{B}_{\tau_k, C_{k-1}}$ and \mathbf{K}_{τ_k} , while improving computational efficiency due to conjugate formulation and allowing for the exact calculation of the marginal likelihood of the graph, impose the artificial restriction on both undirected and directed structures, and may potentially result in misspecification of the networks structures. In particular, the space of decomposable graphs corresponding to \mathbf{K}_{τ_k} is increasingly sparse with the increasing size of τ_k . For example, the percentages of graphs that are decomposable decrease as 95%, 80%, 55%, 29% and 12% for $|\tau_k| = 4, 5, 6, 7$, and 8, respectively (Armstrong, 2005).

4 Bayesian node-wise selection (BANS) framework

To circumvent the above-mentioned challenges, we develop a model selection procedure that allows for more general graph space than the procedures proposed by Bhadra and Mallick (2013); Consonni et al. (2017), as well as greater computational efficiency than the joint model formulation. We propose a Bayesian node-wise selection (BANS) method to jointly estimate undirected edges and directed edges of mlGGMs, using a node-wise regression framework. For a node, the Bayesian variable selection approach simultaneously finds its neighbors and parents, which are connected by undirected and directed edges, respectively.

4.1 Working model

For all $v \in V$, C_v and P_v are defined by the set of all other vertices in the same layer as v , $\tau_{(v)}$, $\{v\}$ and all the preceding vertices, $C_{(v)-1}$, respectively. We consider multivariate regression between $\tau_{(v)}$ as responses and P_v as predictors. Note that the sets in $\{P_v : v \in \tau\}$ are all the same, and we consider $P_v = P_\tau$ for $v \in \tau$. We re-express models (6) and (7) to node-wise regressions:

$$Y_v = \mathbf{Y}_{P_v}^T \mathbf{b}_v + \delta_v \text{ for all } v \in V,$$

where $\delta = (\delta_1, \dots, \delta_p)^T \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{K}^{-1})$ and \mathbf{b}_v is a $|P_v| \times 1$ vector of $\{b_{vu} : u \in P_v\}$. For the well-known conclusion that gives the relation between the concentration matrix of the multivariate normal distribution and the regression coefficients (Anderson, 1984), we have the following lemma.

Lemma 1. *For all $v \in V$, $\delta_v = \delta_{C_v}^T \boldsymbol{\alpha}_v + e_v$, where $\boldsymbol{\alpha}_v = -\mathbf{K}_{C_v, v} / \kappa_{vv}$ and $e_v \sim \mathcal{N}(0, 1/\kappa_{vv})$ is independent of $\delta_{V, \{v\}}$.*

The undirected edges encoded in the zero structures of the precision matrix \mathbf{K} can be found by the zero structures of the regression coefficients obtained from regressing each residual ϵ_v to the residuals corresponding to the other vertices in the same layer.

Proposition 1. *For all $v \in V$, $Y_v = \mathbf{Y}_{P_v}^T \mathbf{b}_v + \mathbf{Y}_{C_v}^T \boldsymbol{\alpha}_v - \mathbf{Y}_{P_v}^T \mathbf{B}_{C_v, P_v}^T \boldsymbol{\alpha}_v + e_v$, where $\boldsymbol{\alpha}_v = -\mathbf{K}_{C_v, v} / \kappa_{vv}$ and $e_v \sim \mathcal{N}(0, 1/\kappa_{vv})$ is independent of $\delta_{V, \{v\}}$.*

The proof of Proposition 1 is provided in Section S2. From Proposition 1, we coherently convert the estimation problem of mlGGM in (3) to that of node-wise regressions. For a given node, both undirected edges between vertices in the same layer and directed edges toward the vertex can be jointly uncovered by model selection for a regression model. Our working model can be re-expressed as follows: for all $v \in V$,

$$Y_v = \sum_{i \in P_v} b_{vi} Y_i + \sum_{j \in C_v} \alpha_{vj} Y_j + \sum_{j \in C_v} \alpha_{vj} \sum_{i \in P_v} -b_{ij} Y_i + e_v, \quad (9)$$

where $e_v \sim \mathcal{N}(0, 1/\kappa_{vv})$. Our model (9) for a vertex $v \in V$ involves regression coefficients for the directed edges from the vertices in the preceding layers ($P_v \rightarrow v$), undirected edges from the vertices in the same layer ($C_v - v$), and directed edges from the vertices in the preceding layers to the vertices in the same layer other than $v(P_v \rightarrow C_v)$.

A simple illustrative example.—Consider the linear model corresponding to the chain graph with two layers for genes $\{Y_1, Y_2\}$ and proteins $\{Y_3, Y_4\}$ described with solid lines in Figure 2 (a): $Y_1 = \delta_1$, $Y_2 = \delta_2$, $Y_3 = b_{31} Y_1 + \delta_3$, and $Y_4 = b_{42} Y_2 + \delta_4$, where ϵ_1 and ϵ_2 and (δ_3, δ_4) are mutually independent, and ϵ_3 and ϵ_4 (the residuals of proteins after taking out effects from genes) have bivariate normal distribution with an arbitrary covariance matrix. The undirected edge between proteins Y_3 and Y_4 is estimated by the regression coefficients, α_{43} and α_{34} , from the two regressions for responses Y_3 and Y_4 in our working model with

true nonzero regression coefficients, as shown in Figure 2 (b). For example, for the protein Y_3 , our working model includes effects for the parent gene Y_1 , the neighbor protein Y_4 , and the indirect effect of the neighbor's parent gene Y_2 . Therefore, when we select directed and undirected edges for a biological unit v , the indirect effects of its neighbor's parents are adjusted.

The next step is to deduce the prior distribution of $\{\mathbf{a}_v | v \in V\}$ from the joint prior on the precision matrix \mathbf{K} . The priors in (8) are re-expressed as follows: for all $\tau \in T$,

$$\begin{aligned} \alpha_{vw} | \kappa_{vv} &\sim N(0, 1 / (\lambda_\tau \kappa_{vv})) \text{ for all } v \in \tau \text{ and } w \in C_v, \\ \kappa_{vv} &\sim \text{Gamma}\left(\frac{\delta_\tau + |\tau| - 1}{2}, \frac{\lambda_\tau}{2}\right) \text{ for all } v \in \tau, \\ b_{vw} &\sim N(0, c_{vw}^2 / \kappa_{vv}) \text{ for all } v \in \tau \text{ and } w \in P_v, \end{aligned} \quad (10)$$

where the Wishart prior on \mathbf{K}_τ translates to normal and gamma priors on \mathbf{a}_v and K_{vv} for $v \in \tau$.

4.2 Model selection priors

Our goal is to infer the dependence structure of the underlying chain graph of the data \mathbf{Y} . Using our framework, the undirected and directed edges of the chain graph can be selected using zero restrictions on the regression parameters, \mathbf{B} and \mathbf{a} . Based on the priors in (10), model selection is achieved through a mixture prior on the regression coefficients: for all $v \in \tau$, $w \in P_v$, $u \in C_v$, and $\tau \in T$,

$$\begin{aligned} \text{Directed edges : } b_{vw} | \gamma_{vw}, \kappa_{vv} &\sim \gamma_{vw} N(0, c_{vw}^2 / \kappa_{vv}) + (1 - \gamma_{vw}) \delta_0, \\ \text{Undirected edges : } \alpha_{vu} | \eta_{vu}, \kappa_{vv} &\sim \eta_{vu} N(0, 1 / (\lambda_\tau \kappa_{vv})) + (1 - \eta_{vu}) \delta_0, \\ \text{Variance parameters : } \kappa_{vv} &\sim \text{Gamma}\left(\frac{\delta_\tau + |\tau| - 1}{2}, \frac{\lambda_\tau}{2}\right), \end{aligned} \quad (11)$$

where δ_0 is the Dirac delta function and c_{vw} , and λ_τ and δ_τ are fixed hyperparameters. We complete the formulation of our model by specifying the prior on γ_{vw} and η_{vw} :

$$P(\gamma_{vw} = 1) = p_{vw} \text{ and } P(\eta_{vw} = 1) = q_{vw},$$

where p_{vw} and q_{vw} are fixed hyperparameters. The binary indicators, γ_{vw} and η_{vw} , are latent variables that encode the directed structure between layers and the undirected structure within a layer. If $\gamma_{vw} = 1$, the arrow from w to v ($w \rightarrow v$) is included in the graph, and $\gamma_{vw} = 0$ otherwise. If $\eta_{vw} = 1$, the undirected edge between v and w ($v - w$) is present in the graph.

5 Likelihood and posterior inference

For each vertex, we implement the Bayesian procedure to select the parents from the preceding layers and neighbors from the same layers as the vertex. Let $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_p)$ and \mathbf{Y}_A for a set $A \subset V$ be a $n \times p$ data matrix and $n \times |A|$ sub-matrix of \mathbf{Y} for which the columns correspond to the set of vertices A . From Proposition 1, we have the following node-wise likelihood function:

$$\begin{aligned}
L &= \prod_{\tau} \prod_{v \in \tau} p(y_v \mid \mathbf{Y}_{C_v \cup P_v}, \boldsymbol{\alpha}_v, \kappa_{vv}, \mathbf{B}_{\tau}, P_v) \\
&= \prod_{\tau} \prod_{v \in \tau} \left(\frac{\kappa_{vv}}{2\pi} \right)^{n/2} \exp \left\{ -\frac{\kappa_{vv}}{2} \left(\mathbf{y}_v - \mathbf{Y}_{P_v} \mathbf{b}_v - \mathbf{Y}_{C_v} \boldsymbol{\alpha}_v + \mathbf{Y}_{P_v} \mathbf{B}_{C_v, P_v}^T \right)^T \right. \\
&\quad \left. \left(\mathbf{y}_v - \mathbf{Y}_{P_v} \mathbf{b}_v - \mathbf{Y}_{C_v} \boldsymbol{\alpha}_v + \mathbf{Y}_{P_v} \mathbf{B}_{C_v, P_v}^T \right) \right\},
\end{aligned}$$

where $\boldsymbol{\alpha}_v = -\mathbf{K}_{C_v, v} / \kappa_{vv}$. We construct the likelihood L by pooling conditional densities within each of the two-layer models. The inference using the likelihood L is not equivalent to the joint likelihood from our original model in (6) and (7), because the resulting local distributions are likely to be *inconsistent* in that there is no joint distribution $p(\mathbf{Y}_{\tau} \mid \mathbf{Y}_{P_{\tau}})$ from each of the local distributions $p(\mathbf{Y}_v \mid \mathbf{Y}_{C_v \cup P_v})$ (Heckerman et al., 2000). The results are asymmetry of the structure, magnitudes, and signs inferred from $\{\boldsymbol{\alpha}_v : v \in V\}$ between α_{43} and α_{34} in Figure 2. The neighborhood selection approach (Meinshausen and Bühlmann, 2006) for UGs requires a symmetrization step after estimating the regression coefficients for all node-wise regressions. Similarly, in our model framework, the set of undirected edges can be defined by $\{v-w : \eta_{vw} = 1 \text{ and } \eta_{wv} = 1\}$ or $\{v-w : \eta_{vw} = 1 \text{ or } \eta_{wv} = 1\}$. For more accurate inference on the graphical structure, instead of using the post-hoc symmetrization, we implement MCMC sampling (BANS) to estimate the posterior distributions with the symmetric constraint $\eta_{vw} = \eta_{wv}$ and showed better accuracy in the structural learning than post-hoc symmetrization obtained from BANS-parallel (Section S7.5 in the Supplementary Materials). However, considering the gain in computation efficiency by using the node-wise parallelizable procedure, BANS-parallel is scalable and useful for high-dimensional problems.

5.1 MCMC Sampling

Our neighborhood selection approach for the chain graph model enables us to estimate the chain graph via a node-wise variable selection framework. Now we consider estimating the undirected edges and directed edges toward a vertex v . Since the parameter spaces for the binary indicators $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ are enormous, computing the explicit posterior probabilities for all possible subsets poses computational challenges. Instead, we use a stochastic search variable selection (SSVS) to generate a Gibbs sequence for each vertex v (George and McCulloch, 1993). Our sampling scheme consists of two parts for updating undirected edges and directed edges. The symmetric constraints of the UG structure can be incorporated when sampling $\boldsymbol{\eta}$ by assuming $\eta_{vw} = \eta_{wv}$ for $v-w$. The MCMC algorithm, which is described in detail in Section S3 can be summarized as follows. For a vertex $v \in \tau \subseteq V$ at iteration t :

1. Undirected edges

1.1 Set $\mathbf{y}_v = \mathbf{y}_v - \mathbf{Y}_{P_v} \mathbf{b}_v$ and $\mathbf{X}_v = \mathbf{Y}_{C_v} - \mathbf{Y}_{P_v} \mathbf{B}_{C_v, P_v}^T$.

1.2 Update $\{\boldsymbol{\eta}_k : k \in \tau\}$ and set $ne_v^{(t)} = \{w \in \tau : \eta_{vw} \neq 0\}$.

1.3 Update $\{\boldsymbol{\alpha}_k : k \in \{v\} \cup ne_v^{(t)}\}$, and $\{\boldsymbol{\kappa}_{kk} : k \in \{v\} \cup ne_v^{(t)}\}$.

2. Directed edges

2.1. Set $\mathbf{y}_v = \mathbf{y}_v - \mathbf{Y}_{C_v} \boldsymbol{\alpha}_v$ and $\mathbf{X}_\tau = (\mathbf{Y}_{P_v - a_{vu1}} \mathbf{Y}_{P_v - a_{vu2}} \mathbf{Y}_{P_v \dots})$, where u_1, u_2, \dots are vertices in $ne_v^{(t)}$.

2.2. Update $\{\boldsymbol{\gamma}_v : v \in \tau\}$, \mathbf{B}_{τ, P_v} and κ_{vv} .

Sampling parameters that correspond to undirected edges are not independent among the vertices in the same layer, due to the imposed symmetric constraints. Thus, our MCMC sampling is performed for each layer. Within an MCMC sampling, steps 1 and 2 are repeated for all vertices in a layer. The mlGGM estimation using this MCMC sampling method is called BANS. We also implemented node-wise sampling scheme, called BANS-parallel in Section S6.

5.2 Graphical structure estimation

The posterior samples of the model parameters for undirected and directed edges obtained from our MCMC methods are used to perform Bayesian inference. The MCMC samples explore the distribution of possible graphs, with each graph leading to a different topology based on the model parameters. The maximum a posteriori (MAP) estimate represents the mode of the posterior distribution of possible graphs. This approach is not feasible, however, since the space of possible graphs is large and the most likely graph may still appear only in a very small proportion of MCMC samples. An alternative and practical solution is to select the edges marginally by using all of the MCMC samples and averaging the presence/absence of each edge over the MCMC samples (Hoeting et al., 1999).

We propose a false discovery rate (FDR)-based determination of significant networks. Our MCMC methods are applied to each of the layers as responses, with all the preceding layers as predictors. For each layer, suppose we have M posterior samples of the corresponding parameter set. From M MCMC iterations for all layers, we estimate the posterior marginal probability of edge inclusion for each edge g_{vw} as the proportion of MCMC iterations after the burn-in in which the edge $v - w$ for $t(v) = t(w)$ or $w \rightarrow v$ for $t(w) < t(v)$ was selected in the graph. The values $1 - g_{vw}$ can be considered as Bayesian q -values, or estimates of the local FDR (Storey and Tibshirani, 2003; Newton et al., 2004), if the vw th edge is called a discovery. Given a desired FDR level $\alpha \in (0, 1)$, we call the set of edges $X_{\phi_\alpha} = \{(v, w) : g_{vw} > \phi_\alpha\}$ discoveries. The significance threshold ϕ_α can be determined based on the approach of Baladandayuthapani et al. (2014). We first sort all $\{g_{vw}\}$ in descending order to yield $\{g_{(j)}\}$. Then we set $\phi_\alpha = g_{(\xi)}$, where $\xi = \max\{k \mid k^{-1} \sum_{j=1}^k (1 - g_{(j)}) < \alpha\}$. We expect that only 100 α % of the discovered X_{ϕ_α} are false positives. An alternative approach is to select the set of edges that appear with marginal posterior probability of inclusion greater than $\phi_\alpha = 0.5$ (Barbieri and Berger, 2004). This rule results in an expected FDR for

$$\phi_\alpha \text{ FDR} = \frac{\sum_{(v,w)} (1 - g_{vw}) I(g_{vw} > \phi_\alpha)}{\sum_{(v,w)} I(g_{vw} > \phi_\alpha)}, \text{ where } I \text{ is the indicator function.}$$

6 Simulations

The aim of our simulation study is to compare BANS with other joint estimation approaches under various simulation settings that generate high- and low-dimensional data under the mlGGM in (3) (Section 6.2). We perform sensitivity analysis of our model to

hyperparameters and check the convergence of the proposed algorithm in Section S4 and S5 of the supplementary materials, respectively. Section S7 includes extensive simulation studies evaluating sign consistency of the estimated partial correlations for edges using structured sampling scheme and the performance of BANS and BANS-parallel with different types of priors under various simulation settings of non-Gaussian synthetic data.

6.1 Data generation with random chain graphs

To generate random chain graphs, we assume that the UGs within layers follow the Erdős and Rényi (ER) model (Erdős and Rényi, 1960). A graph that follows the ER model is constructed by randomly connecting the vertices. We assume that each undirected edge within a layer is included in the chain graph with probability p_E independent from all other edges. We also assume that each directed edge between two consecutive layers is independently linked with probability $p_E/2$. Thus, the vertex i in the graph is almost surely connected to $(|\tau_{(i)}|-1)p_E$ undirected edges and $|\tau_{(i)}-1|p_E/2$ directed edges. We generate an adjacency matrix $A = (A_{ij})_{p \times p}$ that represents a random chain graph on $V = \{1, \dots, p\}$ with $A_{ij} = A_{ji} = 1$ for $i = j$, and $A_{ij} = 1$ and $A_{ji} = 0$ for $i \rightarrow j$. We use the random chain graph generation procedure as follows: (1) assign p vertices to the dependence chain $T = \{\tau_k | k = 1, \dots, q\}$ so that the sizes of the layers are mostly the same; (2) set independent realizations of Bernoulli(p_E) in the lower triangular elements of the sub-matrix corresponding to each layer, then symmetrize it; and (3) set independent realizations of Bernoulli($p_E/2$) in the off-block diagonal elements between two consecutive layers. Given a randomly generated chain graph G with the dependence chain T , the observed data are simulated by the mlGGM in equation (3), after setting the intensities of the nonzero elements of \mathbf{B} and \mathbf{K} . The nonzero elements of \mathbf{B} and \mathbf{K} are randomly sampled from $(-1.5, -0.5) \cup (0.5, 1.5)$. To guarantee the positive definiteness of \mathbf{K} , their diagonal elements are filled by column-wise sums of absolute values, plus a small constant. Then we draw a random sample \mathbf{Y} of size n from the distribution $N(\mathbf{0}, \mathbf{\Omega}^{-1})$ with $\mathbf{\Omega}$ in equation (4). In the additional simulation setting that emulates the DNA Damage response network estimated for 309 TCGA lung squamous cell carcinoma (LUSC) samples (Section S7.1), we used the parameter estimates \mathbf{B} and \mathbf{K} obtained from the real data application to generate simulation datasets.

6.2 Performance evaluation

We compare the performance of our BANS method against those of multivariate regression-based methods in learning the topologies of chain graphs under various simulation settings of mlGGMs according to the simulation parameters, the number of variables (p), sample size (n), number of layers (q), and degree of sparsity (p_E). We consider the following multivariate regression-based methods for two-layer models: multivariate regression with covariance estimation (MRCE) (Rothman et al., 2010), and covariate-adjusted precision matrix estimation (CAPME) (Cai et al., 2012), which are available in R packages `MRCE` and `capme`, respectively. We apply `glasso` (Friedman et al., 2008) for MRCE and `CLIME` (Cai et al., 2011) for CAPME to estimate the UG for the first layer, as MRCE and CAPME are not applicable to UGs.

We investigate the performance of our method under different simulation settings by varying (p, n, q, p_E) . Different measures of structural difference can be used to assess the performance

of our method. We define TP, TN, FP, and FN as the total number of true positive, true negative, false positive, and false negative edges, respectively. For goodness of estimation, we use sensitivity $TP/(TP+FN)$, specificity $TN/(TN+FP)$, and Matthew's correlation coefficient (MCC): $\{(TP \times TN) - (FP \times FN)\} / \{(TP + FP)(TP+FN)(TN + FP)(TN + FN)\}^{1/2}$ that ranges from -1 (total disagreement) to 1 (perfect classification), with a larger value corresponding to a better fit.

To evaluate the sensitivity, specificity, MCC, and number of discoveries, we determine the tuning parameters of glasso, MRCE, CLIME and CAPME using five-fold cross-validation, and control the FDR at 0.1 for BANS. To further compare the performance on graph structure recovery, we obtain the receiver operating characteristic (ROC) curve and the MCC curve as the function of the number of discoveries for each simulation dataset by varying the tuning parameters for MRCE and CAPME, as well as the cutoff for the posterior marginal probability of edge inclusion for BANS. The sensitivity, specificity, MCC, and number of discoveries for the selected tuning parameters, and pAUC and AUC obtained from the ROC curves, are shown in Table S1. In terms of graph structure recovery, our method yields better performance for all settings. The number of discoveries is very close to the number of edges in the true graph, while other methods tend to provide a much greater number of edges. The small standard errors for the number of discoveries suggest that our procedure is stable across simulation replicates. Figure 3 and Figure S11-S13 display the ROC curves and MCC curves averaged over 50 replications, which demonstrate that our method performs better than MRCE and CAPME. Our method shows better ROC curves and MCC curves for all four simulation settings across different tuning parameters. In the high-dimensional case, when $p = 200$ and $n = 100$ (Figure 3), the sensitivities of MRCE and CAPME are flattened for some intervals of 1-specificity, and MRCE shows unstable variable selection performance, because the curves show a decreasing pattern as the tuning parameters decrease.

We evaluated the performance of BANS in simulation settings where data were generated from non-Gaussian, and the graphical structures were the same as the estimated DNA damage response network for 309 TCGA lung squamous cell carcinoma. We performed comparisons to (1) BANS-parallel (Section S6), (2) XMRF (Wan et al., 2016), that fits UGs to mixed data types, (3) Neighborhood selection of UGs using Horseshoe prior (Carvalho et al., 2010) and (4) objective Bayes fractional Bayes factor (OBFBF) (Consonni et al., 2017) using model averaging. Given the simulation settings where the data are generated from Poisson distribution, BANS showed better performance than XMRF (Section S7.2). For the evaluation of our prior choice, the point mass prior in equation (11) showed superior performance than Horseshoe prior and OBFBF in estimating UGs (Section S7.3 and Section S7.4). While BANS-parallel without symmetric constraint in the MCMC sampling showed smaller AUC and MCC values than BANS, it still performed better than other methods such as MRCE and CAPME (Figure S8). Considering the gain in computational efficiency (Section S7.5), BANS-parallel is potentially useful alternative in high-dimensional setting. Finally, we confirmed that posterior inference on signs of edges were accurately estimated from the structured MCMC algorithm given η and γ using BANS (Section S7.6).

7 Pan-Cancer Network Analysis of Multi-platform Omic Data

Biological motivation

Crosstalk within signaling pathways and their perturbation by oncogenes limit single gene-based approaches to understanding cancer biology. Approaches have been developed for discovering mutations that perturb signaling networks (so-called network-attacking mutations) to understand how individual genomic variants initiate network perturbation (Creixell et al., 2012). Although genomic variants are critical to understanding functional cancer networks, it is well-established that complex molecular networks and systems are formed by a large number of interactions of genes and their products, which operate in response to different cellular conditions (Bandyopadhyay et al., 2010). Therefore, systematic approaches to unravelling cancer-specific rewiring of molecular networks are key to the successful identification of network-based drug targets for cancer treatment, in the paradigm of network medicine that acknowledges the application of network topology and dynamics towards identification of therapeutic targets (Barabási et al., 2011).

TCGA PanCancer Atlas initiative (Weinstein et al., 2013) built a uniformly-processed dataset and a unified data analysis pipeline to develop an integrated picture of commonalities and differences across tumor types. Recent studies (Hoadley et al., 2018; Sanchez-Vega et al., 2018) reclassified human tumor types based on molecular similarity and investigated co-occurrence of alterations in tumor signaling pathways, which differentiate between individual tumors and tumor types using TCGA pan-cancer data. (Akbari et al., 2014) investigated correlations between protein and other data types, such as mRNA, copy number, and mutation data across cancer. (Gong et al., 2017) performed expression quantitative trait locus (eQTL) analysis that focuses on the links between genotypes and gene expression for pan-cancer TCGA data. These methods only consider links between two platforms; they do not incorporate within platform dependencies. Using our BANS method, graph-based multi-level integration approach, our goal is to understand unified interplay within and between genomic, epigenomic, transcriptomic, and proteomic platforms to elucidate the commonalities and differences in systems across cancer types.

Data structures

We applied our method to multi-omic datasets from the 7 TCGA tumor types, lung adenocarcinoma (LUAD, n=356), lung squamous cell carcinoma (LUSC, n=309), colon adenocarcinoma (COAD, n=338), rectum adenocarcinoma (READ, n=121), uterine corpus endometrial carcinoma (UCEC, n=393), ovarian serous cystadenocarcinoma (OV, n=227), and skin cutaneous melanoma (SKCM, n=333). For each of the tumor types, we included multi-platform data, DNA methylation, copy number alteration, mRNA expression data, and reverse phase protein array (RPPA)-based proteomic data. Genomic features from each data platform constitute a layer; the ordering of the layers follows biological justifications, because of the natural interplay among diverse genomic features: gene encoded by DNA is transcribed to mRNA, mRNA is translated to protein, and DNA methylation helps to regulate transcription (Morris and Baladandayuthapani, 2017) (Figure S14).

Based on the principle that within-platform interactions arise from pathway-based dependencies that are altered across different tumor types, we selected 10 key signaling pathways, based on emerging literature on RPPA-based proteomic profiling of various tumor types (Akbani et al., 2014; Cherniack et al., 2017; Li et al., 2017). The pathways and the gene membership for each pathway are listed in Table S2. We focus on building integrative networks for all combinations of the 10 pathways and the 7 cancer types. Using TCGA-Assembler (Zhu et al., 2014), we downloaded CNA, DNA methylation, mRNA expression, and protein expression data, then matched the samples across platforms. We applied BANS separately to each type of cancer and pathway combination ($10 \times 7 = 70$ analyses). The MCMC sampler was run for 10,000 iterations of burn-in, followed by 20,000 iterations as the basis for inference. For each analysis, we estimated the graph structure at FDR 0.1 and obtained the signs of each nonzero coefficient using the cutoff $\xi = 0.5$ for the posterior probability of the signs from the structured estimation. Figures S15-S24 display the estimated networks for all pathways and cancer types.

7.1 Global commonalities and differences in mIGGMs

We first investigated the number of edges that are shared and differentiated across different cancer types for each of the possible edges at the platform level. For each combination of the cancer types and pathways, the mIGGM contains 4 layers at CNA, DNA methylation, mRNA expression, and protein expression, and allows 9 different types of dependencies, represented by the 4 types of undirected edges (i.e., CNA–CNA, methylation–methylation, mRNA–mRNA, and protein–protein) and the 5 types of directed edges (i.e., CNA \rightarrow miRNA, CNA \rightarrow protein, methylation \rightarrow mRNA, methylation \rightarrow protein, and mRNA \rightarrow protein) (Figure S14). Across all pathways, our BANS method detected 433 (UCEC), 350 (SKCM), 328 (OV), 240 (READ), 391 (COAD), 394 (LUAD), and 361 (LUSC) directed/undirected edges in the estimated mIGGMs. Across all 10 pathways, we decomposed the edges by the aforementioned 9 types of dependences within and between the four layers, as well as the numbers of intersecting edges across the 7 cancer types are depicted using UpSet plots (shown in Figure 4). UpSet plot is an effective visualization of intersections for more than three sets, and a more-scalable approach than Euler diagrams (Lex and Gehlenborg, 2014). For each of the 9 possible relations within and between layers, the UpSet plot contains column and row bar plots. The column bar plot encodes all edge set intersections in the columns of a matrix using a binary pattern, and renders bars above the matrix columns to represent the number of exclusively intersecting edges that are shared by the corresponding cancer types to the column, but not of the others. The row bar plot displays the total number of edges for each cancer type.

Between platform regulatory relations: Transcriptional and translational effects represented by directed edges CNA \rightarrow mRNA (Figure 4-b) and mRNA \rightarrow Protein (Figure 4-h) tend to be shared across cancer types, which is along expected lines of the basic biological mechanisms. For example, in Figure 4-b, 36 edges were shared by all 7 cancer types, and few unique edges to a cancer type were found. We found few regulatory edges from CNA to proteins, and from methylation to mRNA and Protein across cancer types (Figure 4-c,e,f). For example, for Methylation \rightarrow mRNA in Figure 4-e, we found 7 edges unique to OV and COAD.

Within platform conditional dependencies: The undirected edges represent dependencies within platform after taking out the regulatory effects from upstream platforms. The dependence structures within platforms tend to be unique to cancer type. For CNA (Figure 4-a), UCEC, COAD, LUSC, and LUAD had 23, 15, 11, and 7 edges that are not shared by any other cancer types, while 6 edges were shared across all cancer types. For Methylation (Figure 4-d), OV included 14 unique edges among 48 edges in total. For mRNA (Figure 4-g), SKCM had 13 of 45 edges that were unique. For protein (Figure 4-i), UCEC had the largest protein network, with 62 protein-protein interactions, 13 of which were unique to the cancer, while the same number of the edges were shared across all cancer types.

7.2 Pan-cancer network signaling

We investigate the extent of cross-signaling within- and between- layers for each pathway across tumor types. For a given network, we define the ratio of the observed number of edges to the total number of possible edges as *connectivity score (CS)*: high (low) CS value indicates high (low) cross-signaling of the network. We also compute standard deviation of CS values across cancer types to represent that the levels of connectivity of the network are different across tumor types: the genes in the network are highly connected (high signaling) in some cancers, but have few connection (low signaling) in other cancers. Figure 5 displays the CS values and the variability are displayed: CS ranged from 0 to 0.5, and the variability ranged from 0 to 0.17. The overall pattern of the heat map suggests that the within-layer sub-networks had a higher level of signaling than between-layer sub-networks. The core reactive pathway showed a high level of network signaling within platforms, and the highest level of protein-protein signaling across tumor types, which suggests that our estimation aligns well with the *a priori* functional characteristics of this pathway, which was defined on the basis of all tumor types (Akbari et al., 2014). The CS was markedly different in methylation-methylation networks for most of the pathways, compared to other types of sub-networks: the apoptosis pathway for LUAD and COAD, cell cycle pathway for OV, EMT pathway for UCEC, TSC/mTOR pathway for LUAD, breast reactive pathway for UCEC, and core reactive pathway for READ, UCEC, and SKCM showed the high level of signaling for methylation-methylation. The transcriptional and translational effects that are represented by CNA \rightarrow mRNA and mRNA \rightarrow protein sub-networks showed the high-level of signaling across pathways and cancer types, compared to other between-layer networks, again following along expected lines of the biological hierarchy.

7.3 p53 Integrative Networks

In this section we focus on a gene/protein of particular interest (i.e., p53) across all tumor types. p53 is the most frequently mutated gene in human tumors, and has a central role as a tumor suppressor and novel therapeutic target (Bouchet et al., 2006; Levine, 2019). The transcription factor p53 is activated downstream of the DNA damage response (DDR) in reaction to cell stress, and mediates distinct outcomes of DDR signaling (Reinhardt and Schumacher, 2012). The TP53 gene encodes the p53 protein that targets a large set of genes associated to apoptosis and cell cycle pathways (Bouchet et al., 2006; Reinhardt and Schumacher, 2012). Due to the high level of inter-connectivity of the DDR with other signaling networks, predicting the efficacy of treatment and designing an optimal

combination therapy to target multiple genes will require a detailed understanding of the tumor-specific signals of other molecules.

Let g_{ij} be the estimated posterior marginal probability of the edge ($i-j$ or $i \rightarrow j$) inclusion. We consider the estimated mlGGMs are weighted graphs, whose edges have been assigned the given posterior inclusion probabilities, and the degree of the nodes across all layers are defined as the summation of the weights for the edges that are connected to the node:

$W_i = \sum_{j=1}^p I(i \rightarrow j \text{ or } i \leftarrow j \text{ or } i - j) g_{ij}, \forall i \in V$. The degree of the nodes in a weighted mlGGM measures the strength of nodes in terms of the total weight of their directed and undirected connections. With the goal of studying the underlying mechanism of p53 protein, we focused on the sub-networks of the estimated mlGGMs, which include the nodes connected to the p53 protein by any lengths of paths including both undirected and directed edges. Figure 6 displays the p53 integrative networks, where the edges are colored and weighted by the signs and the posteriors, g_{ij} , and the sizes of nodes are weighted by their degrees, W_i .

7.3.1 Translational findings—Our results confirm that the transcriptional effect from TP53 at CNA to TP53 gene expression, and translational effects from TP53 gene expression to the p53 protein across 6 tumor types (all except for SKCM), all with positive regulations. In contrast, the SKCM network included no upstream regulatory effects (from DNA and RNA) and only included protein-protein interactions. UCEC and SKCM shared the same protein-protein interactions between p53 protein expression and T68 phosphorylated CHK2 (CHK2PT68): CHK2 is a protein kinase that is activated in response to DNA damage and directly phosphorylated p53 on serine 20, which provides a mechanism for increased stability of p53 (Hirao et al., 2000) and has been suggested as an anticancer therapy target given its role as a tumor suppressor (Zannini et al., 2014). These findings prioritize UCEC and SKCM as potential cancer types for CHK2 targeting. We also found epigenetic effects in the p53 sub-networks: only UCEC had a positive regulatory effect from BRCA2 DNA methylation, which was correlated with CHEK1 DNA methylation. The physical and functional interactions between BRCA2 and TP53 have been reported and hypermethylation of the CpG island in the promoters of BRCA2 gene involve their inactivation and therefore a higher risk of developing a tumor including uterine cancer (Rajagopalan et al., 2010; Bosviel et al., 2012).

Epigenetic drugs have been developed and the anticancer effects are often tested using genome-editing technologies such as CRISPR-Cas9 systems, called Epigenome editing, which provides advantages over direct gene knockout based on RNA interference (Kungulovski and Jeltsch, 2016). Moreover, successes of epigenetic drugs have been reported by the important roles in synergy with other anticancer therapies or in reversing acquired therapy resistance (Morel et al., 2019). Our findings through this deeper investigation of underlying biological mechanisms of p53 networks across multiple molecular levels and cancer types have the potential to facilitate the development of novel therapeutic strategies, specifically gene-drug interactions for single and combination agents.

8 Discussion

We propose a unified Bayesian framework to model the layered architecture of networks from multiple omic data. We employ a multi-layered Gaussian graphical model (mlGGM) to investigate the conditional dependencies among the variables from multiple layers, while accounting for the order defined by biological hierarchies. The mlGGM is built upon a multivariate regression framework with mean and residual precision parameters, for which zero structures represent undirected and directed edges in the chain graph. Our fully probabilistic formulation coherently converts the complex mlGGM into more tractable, node-wise multiple regression models, wherein the zero structures of the regression coefficients encompass both the undirected and directed edges. Our edge-specific variable selection priors on node-wise regression models allow for flexible modeling of any type of graph, without restriction to decomposable graphs, as well as the incorporation of relevant prior knowledge, while maintaining computational efficiency. We applied our Bayesian node-wise selection (BANS) method to Pan-cancer integrative network analysis and found structural commonalities and differences across cancer types. We also identified underlying mechanisms of the p53 protein, which is a novel drug target for cancer treatment.

For identifiability, the main assumption in this article is that the layers have natural ordering. Ma et al. (2008) and Ha et al. (2015) proposed methods to estimate a Markov equivalence class of a chain graph and a DAG (as a special case of the chain graph model) by recovering skeletons on its subgraphs with no ordering information. In the absence of natural ordering, a chain graph is not identifiable and only its Markov equivalence class can be identified. In many applications, the vertices are naturally partitioned into multiple ordered layers, along with time points or the intrinsic biological mechanisms.

For chain graph models, there are two main types of conditional independencies implied by the LWF Markov property and the alternative Markov property (AMP). Through zero structures of the mean parameter \mathbf{B} and the residual precision matrix \mathbf{K} in the mlGGM in (3), we describe the AMP on chain graph models. For the LWF Markov property in high-dimensional Gaussian settings, Sohn and Kim (2012); McCarter and Kim (2014) made a convex formulation by using a conditional joint distribution given the predictors. The AMP Markov property is coherent, however, with data generation (as described in Section S1) by the direct link between the zero structure of the parameters in the multivariate linear regressions, and the presence/absence of edges in the graph.

For the resulting asymmetry of the structure, we impose the constraint on \mathbf{K} during MCMC. Although most graphical model selection approaches based on node-wise regressions focus on estimating the structure, we showed the numerical performance of estimating signs of the directed edges and undirected edges, based on the structured estimation from the conditional posterior $P(\boldsymbol{\alpha}, \boldsymbol{\kappa}, \mathbf{B} | \mathbf{Y}, \boldsymbol{\eta}, \boldsymbol{\gamma})$ where $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ are subjected to the structure \mathcal{G} .

Our node-wise regression-based formulation, BANS to jointly estimate the undirected and directed edges to a node provides a flexible modeling framework. In particular, the proposed approach can be extended to allow nonlinear regulatory relations between layers, instead of the linearity assumption on the parameter \mathbf{B} , and to infer multiple mlGGMs where some of

the graphs may be unrelated, while others share common edges. BANS R codes implementing our method are available on https://github.com/***/BANS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Akbani R, Ng PKS, Werner HM, Shahmoradgoli M, Zhang F, Ju Z, Liu W, Yang J-Y, Yoshihara K, Li J et al. (2014), 'A pan-cancer proteomic perspective on the cancer genome atlas', *Nature communications* 5, 3887.
- Anderson T (1984), 'Multivariate statistical analysis', Wiley and Sons, New York, NY.
- Andersson SA, Madigan D and Perlman MD (2001), 'Alternative markov properties for chain graphs', *Scandinavian journal of statistics* 28(1), 33–85.
- Armstrong H (2005), Bayesian estimation of decomposable Gaussian graphical models, PhD thesis, The University of New South Wales.
- Baladandayuthapani V, Talluri R, Ji Y, Coombes KR, Lu Y, Hennessy BT, Davies MA and Mallick BK (2014), 'Bayesian sparse graphical models for classification with application to protein expression data', *The annals of applied statistics* 8(3), 1443. [PubMed: 26246866]
- Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M et al. (2010), 'Rewiring of genetic networks in response to dna damage', *Science* 330(6009), 1385–1389. [PubMed: 21127252]
- Barabási A-L, Gulbahce N and Loscalzo J (2011), 'Network medicine: a network-based approach to human disease', *Nature reviews genetics* 12(1), 56.
- Barbieri MM and Berger JO (2004), 'Optimal predictive model selection', *Annals of Statistics* pp. 870–897.
- Bhadra A and Mallick BK (2013), 'Joint high-dimensional bayesian variable and covariance selection with an application to eqtl analysis', *Biometrics* 69(2), 447–457. [PubMed: 23607608]
- Bosviel R, Durif J, Guo J, Mebrek M, Kwiatkowski F, Bignon Y-J and Bernard-Gallon DJ (2012), 'Brca2 promoter hypermethylation in sporadic breast cancer', *Omics: a journal of integrative biology* 16(12), 707–710. [PubMed: 23215809]
- Bouchet BP, de Fromental CC, Puisieux A and Galmarini CM (2006), 'p53 as a target for anti-cancer drug development', *Critical reviews in oncology/hematology* 58(3), 190–207. [PubMed: 16690321]
- Cai T, Liu W and Luo X (2011), 'A constrained ℓ_1 minimization approach to sparse precision matrix estimation', *Journal of the American Statistical Association* 106(494), 594–607.
- Cai TT, Li H, Liu W and Xie J (2012), 'Covariate-adjusted precision matrix estimation with an application in genetical genomics', *Biometrika* p. ass058.
- Carvalho CM, Polson NG and Scott JG (2010), 'The horseshoe estimator for sparse signals', *Biometrika* 97(2), 465–480.
- Chen M, Ren Z, Zhao H and Zhou H (2016), 'Asymptotically normal and efficient estimation of covariate-adjusted gaussian graphical model', *Journal of the American Statistical Association* 111 (513), 394–406. [PubMed: 27499564]
- Cherniack AD, Shen H, Walter V, Stewart C, Murray BA, Bowlby R, Hu X, Ling S, Soslow RA, Broaddus RR et al. (2017), 'Integrated molecular characterization of uterine carcinosarcoma', *Cancer Cell* 31(3), 411–423. [PubMed: 28292439]
- Consonni G, La Rocca L and Peluso S (2017), 'Objective bayes covariate-adjusted sparse graphical model selection', *Scandinavian Journal of Statistics* 44(3), 741–764.
- Cox DR and Wermuth N (1993), 'Linear dependencies represented by chain graphs', *Statistical science* pp. 204–218.
- Creixell P, Schoof EM, Erler JT and Linding R (2012), 'Navigating cancer network attractors for tumor-specific therapy', *Nature biotechnology* 30(9), 842.

- Dawid AP and Lauritzen SL (1993), 'Hyper markov laws in the statistical analysis of decomposable graphical models', *The Annals of Statistics* pp. 1272–1317.
- Drton M and Eichler M (2006), 'Maximum likelihood estimation in gaussian chain graph models under the alternative markov property', *Scandinavian journal of statistics* 33(2), 247–257.
- Drton M and Perlman MD (2008), 'A sinful approach to gaussian graphical model selection', *Journal of Statistical Planning and inference* 138(4), 1179–1200.
- Erdős P and Rényi A (1960), 'On the evolution of random graphs', *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, 17–61.
- Friedman J, Hastie T and Tibshirani R (2008), 'Sparse inverse covariance estimation with the graphical lasso', *Biostatistics* 9(3), 432–441. [PubMed: 18079126]
- Frydenberg M (1990), 'The chain graph markov property', *Scandinavian Journal of Statistics* pp. 333–353.
- George EI and McCulloch RE (1993), 'Variable selection via gibbs sampling', *Journal of the American Statistical Association* 88(423), 881–889.
- Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, Feng J, Liu R, Diao L, Guo A-Y et al. (2017), 'Pancanqtl: systematic identification of cis-eqtls and trans-eqtls in 33 cancer types', *Nucleic acids research* 46(D1), D971–D976.
- Ha MJ, Sun W and Xie J (2015), 'Penpc: A two-step approach to estimate the skeletons of high-dimensional directed acyclic graphs', *Biometrics*.
- Hanahan D and Weinberg RA (2000), 'The hallmarks of cancer', *cell* 100(1), 57–70. [PubMed: 10647931]
- Hanahan D and Weinberg RA (2011), 'Hallmarks of cancer: the next generation', *cell* 144(5), 646–674. [PubMed: 21376230]
- Heckerman D, Chickering DM, Meek C, Rounthwaite R and Kadie C (2000), 'Dependency networks for inference, collaborative filtering, and data visualization', *Journal of Machine Learning Research* 1 (10), 49–75.
- Hirao A, Kong Y-Y, Matsuoka S, Wakeham A, Ruland J, Yoshida H, Liu D, Elledge SJ and Mak TW (2000), 'Dna damage-induced activation of p53 by the checkpoint kinase chk2', *Science* 287(5459), 1824–1827. [PubMed: 10710310]
- Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V et al. (2018), 'Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer', *Cell* 173(2), 291–304. [PubMed: 29625048]
- Hoeting JA, Madigan D, Raftery AE and Volinsky CT (1999), 'Bayesian model averaging: a tutorial', *Statistical science* pp. 382–401.
- Kristensen VN, Lingjærde OC, Russnes HG, Vollan HKM, Frigessi A and Børresen-Dale A-L (2014), 'Principles and methods of integrative genomic analyses in cancer', *Nature Reviews Cancer* 14 (5) 299–313. [PubMed: 24759209]
- Kungulovski G and Jeltsch A (2016), 'Epigenome editing: state of the art, concepts, and perspectives', *Trends in Genetics* 32(2), 101–113. [PubMed: 26732754]
- Lauritzen S (1996), *Graphical models*, Vol. 17, Oxford University Press, USA.
- Lauritzen SL and Wermuth N (1989), 'Graphical models for associations between variables, some of which are qualitative and some quantitative', *The Annals of Statistics* pp. 31–57.
- Lee W and Liu Y (2012), 'Simultaneous multiple response regression and inverse covariance matrix estimation via penalized gaussian maximum likelihood', *Journal of multivariate analysis* 111, 241–255. [PubMed: 22791925]
- Levine AJ (2019), Targeting therapies for the p53 protein in cancer treatments', *Annual Review of Cancer Biology* 3, 21–34.
- Lex A and Gehlenborg N (2014), 'Points of view: Sets and intersections'.
- Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, Yang J-Y, Broom BM, Verhaak RG, Kane DW et al. (2013), 'Tcpc: a resource for cancer functional proteomics data', *Nature methods* 10(11), 1046.
- Li J, Zhao W, Akbani R, Liu W, Ju Z, Ling S, Vellano CP, Roebuck P, Yu Q, Eterovic AK et al. (2017), 'Characterization of human cancer cell lines by reverse-phase protein arrays', *Cancer cell* 31(2), 225–239. [PubMed: 28196595]

- Lin J, Basu S, Banerjee M and Michailidis G (2016), 'Penalized maximum likelihood estimation of multi-layered gaussian graphical models', *Journal of Machine Learning Research* 17, 1–51.
- Ma Z, Xie X and Geng Z (2008), 'Structural learning of chain graphs via decomposition', *Journal of Machine Learning Research* 9(12), 2847–2880. [PubMed: 19759856]
- McCarter C and Kim S (2014), On sparse gaussian chain graph models, in 'Advances in Neural Information Processing Systems', pp. 3212–3220.
- Meinshausen N and Bühlmann P (2006), 'High-dimensional graphs and variable selection with the lasso', *The annals of statistics* pp. 1436–1462.
- Morel D, Jeffery D, Aspeslagh S, Almouzni G and Postel-Vinay S (2019), 'Combining epigenetic drugs with other therapies for solid tumours? past lessons and future promise', *Nature Reviews Clinical Oncology* pp. 1–17.
- Morris JS and Baladandayuthapani V (2017), 'Statistical contributions to bioinformatics: Design, modelling, structure learning and integration', *Statistical modelling* 17(4-5), 245–289. [PubMed: 29129969]
- Newton MA, Noueiry A, Sarkar D and Ahlquist P (2004), 'Detecting differential gene expression with a semiparametric hierarchical mixture method', *Biostatistics* 5(2), 155–176. [PubMed: 15054023]
- Pe'er D and Hachohen N (2011), 'Principles and strategies for developing network models in cancer', *Cell* 144(6), 864–873. [PubMed: 21414479]
- Rajagopalan S, Andreeva A, Rutherford TJ and Fersht AR (2010), 'Mapping the physical and functional interactions between the tumor suppressors p53 and brca2', *Proceedings of the National Academy of Sciences* 107(19), 8587–8592.
- Reinhardt HC and Schumacher B (2012), The p53 network: cellular and systemic dna damage responses in aging and cancer', *Trends in Genetics* 28(3), 128–136. [PubMed: 22265392]
- Rothman AJ, Levina E and Zhu J (2010), 'Sparse multivariate regression with covariance estimation', *Journal of Computational and Graphical Statistics* 19(4), 947–962. [PubMed: 24963268]
- Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadou S, Liu DL, Kantheti HS, Saghafeinia S et al. (2018), 'Oncogenic signaling pathways in the cancer genome atlas', *Cell* 173(2), 321–337. [PubMed: 29625050]
- Sohn K-A and Kim S (2012), Joint estimation of structured sparsity and output structure in multiple-output regression via inverse-covariance regularization., *in 'AISTATS'*, pp. 1081–1089.
- Storey JD and Tibshirani R (2003), 'Statistical significance for genomewide studies', *Proceedings of the National Academy of Sciences* 100(16), 9440–9445.
- Wan Y-W, Allen GI, Baker Y, Yang E, Ravikumar P, Anderson M and Liu Z (2016), 'Xmrf: an r package to fit markov networks to high-throughput genetics data', *BMC systems biology* 10(3), 69. [PubMed: 27586041]
- Wang W, Baladandayuthapani V, Morris JS, Broom BM, Manyam G and Do K-A (2013), 'ibag: integrative bayesian analysis of high-dimensional multiplatform genomics data', *Bioinformatics* 29(2), 149–159. [PubMed: 23142963]
- Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM, Network, C. G. A. R. et al. (2013), 'The cancer genome atlas pan-cancer analysis project', *Nature genetics* 45(10), 1113. [PubMed: 24071849]
- Yin J and Li H (2011), 'A sparse conditional gaussian graphical model for analysis of genetical genomics data', *The annals of applied statistics* 5(4), 2630. [PubMed: 22905077]
- Zannini L, Delia D and Buscemi G (2014), 'Chk2 kinase in the dna damage response and beyond', *Journal of molecular cell biology* 6(6), 442–457. [PubMed: 25404613]
- Zhu Y, Qiu P and Ji Y (2014), 'Tcga-assembler: open-source software for retrieving and processing tcga data', *Nature methods* 11(6), 599–600. [PubMed: 24874569]

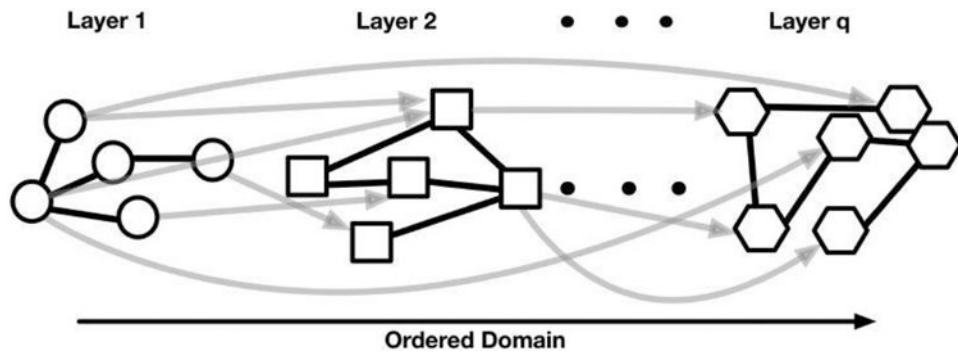


Fig. 1. Example network for q ordered layers. In our application, we consider four layers corresponding to DNA methylation, Copy number aberration, gene expression and protein expression (see Section 7).

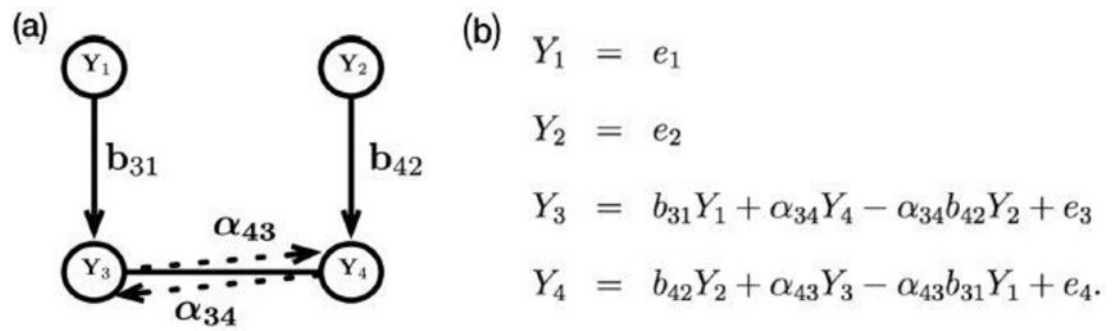


Fig. 2.

An example of a chain graph with layers $\{1, 2\}$, $\{3, 4\}$. (a) Solid lines represent the data generating structure and dotted lines display how we parameterize undirected edges in our neighborhood selection framework. (b) Our working model.

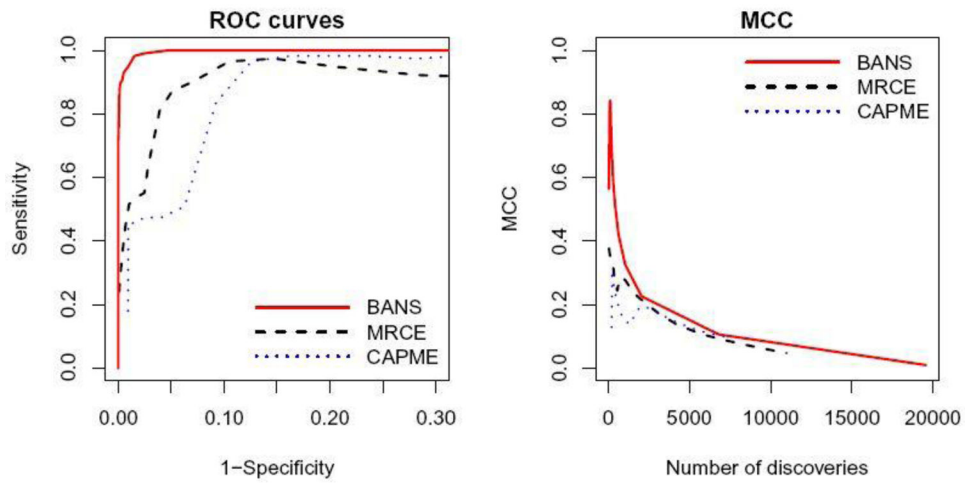


Fig. 3. ROC curves and MCC curves for graph structure learning for $(p, n, q, p_E) = (200, 100, 10, 0.03)$.

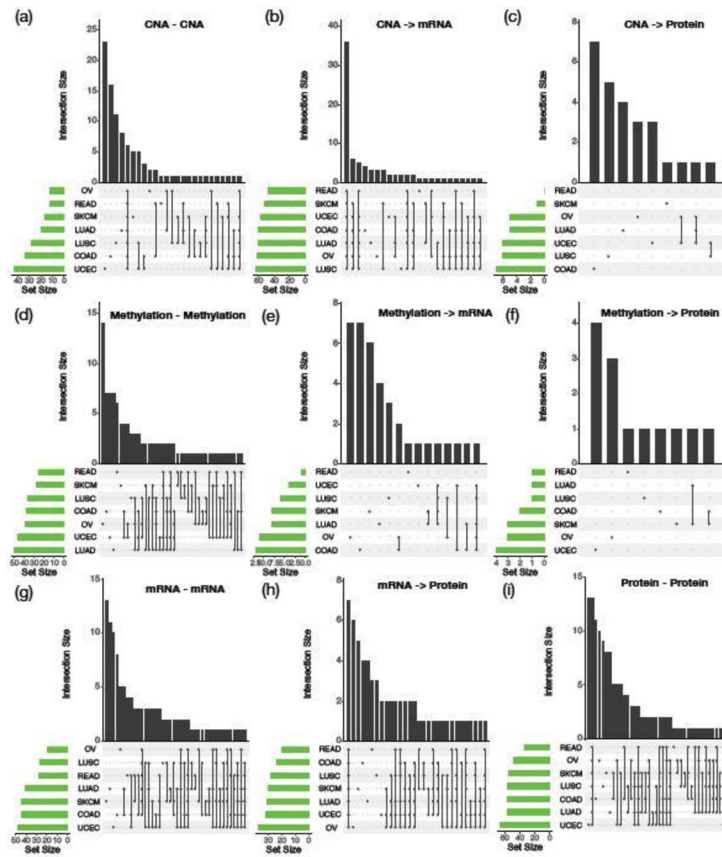


Fig. 4. UpSet plots showing relationships of mIGGMs across all 10 pathways between 7 cancer types. Each column-wise bar corresponds to the number of exclusively intersecting edges that are shared by the cancer types represented by the dark circles, but not of the others, and each row-wise bar displays the total number of edges for the corresponding cancer type.

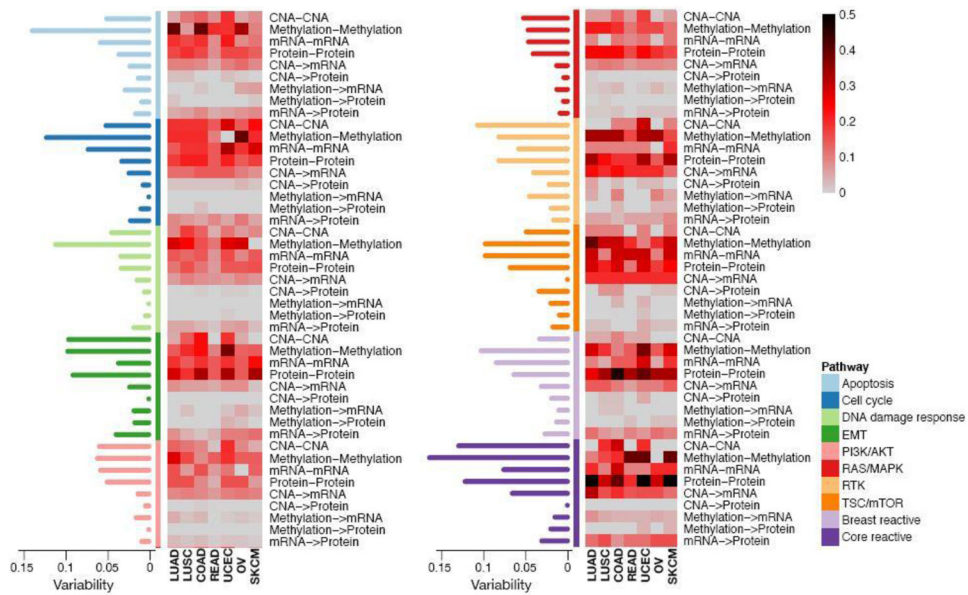


Fig. 5. Heatmap depicting connectivity score (CS) of the within- and between-layer sub-networks of the estimated mGGMs across the 7 cancer types and 10 pathways. The scores are indicated on a low-to-high scale (grey-red-black). The standard deviations of the CS values across cancer types are displayed in the barplots.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

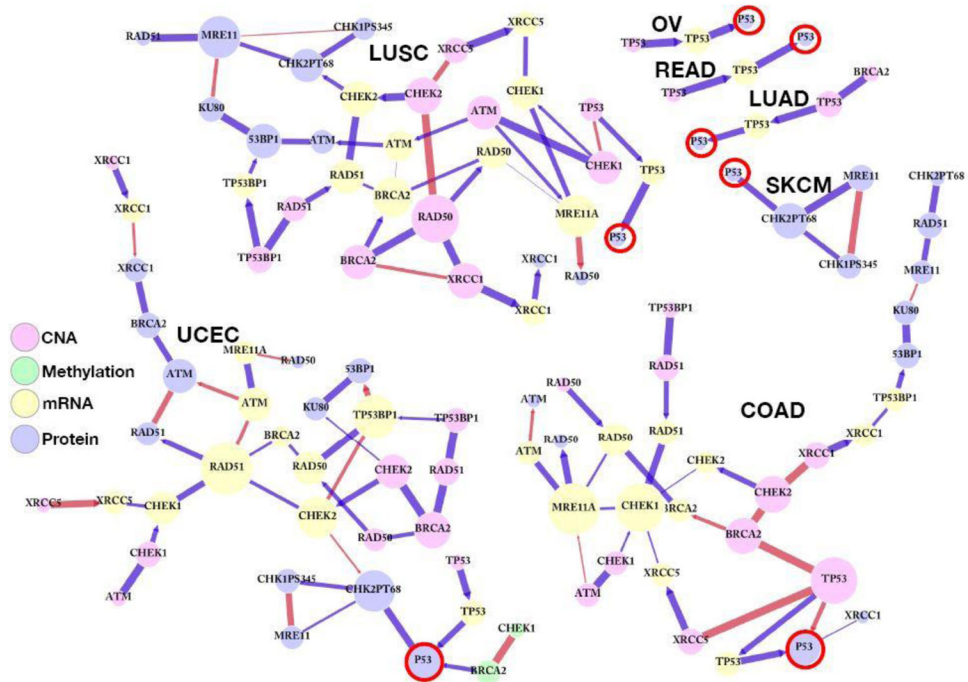


Fig. 6. Integrative sub-networks for the p53 protein (red circle), inferred from BANS for LUAD, LUSC, COAD, READ, UCEC, OV, and SKCM. Each connected component includes all nodes that are connected to the p53 protein by any lengths of paths, including both undirected and directed edges for each cancer. The colors of edges indicate the inferred signs of the edges: negative (red) and positive (blue). The sizes of nodes and the widths of edges are weighted by their degrees and posterior edge inclusion probabilities, respectively.