Title
Detecting differential patterns of interaction in molecular pathways

Journal
Biostatistics, 16(2)

ISSN
1465-4644

Authors
Yajima, Masanao
Telesca, Donatello
Ji, Yuan
et al.

Publication Date
2015-04-01

DOI
10.1093/biostatistics/kxu054

Peer reviewed

Biostatistics (2014), **0**, 0, *pp.* 1–19
doi:10.1093/biostatistics/TwoGraphs˙Biostatistics

Detecting Differential Patterns of Interaction in Molecular Pathways

MASANAO YAJIMA[1,*] , DONATELLO TELESCA[2], YUAN JI[3], PETER MÜLLER[4]

[1] *Fred Hutchinson Cancer Research Center.* [2] *UCLA, Department of Biostatistics.* [3] *Center for Clinical and Research Informatics, NorthShore University HealthSystem.* [4] *The University of Texas at Austin, Department of Mathematics.*

Summary

We consider statistical inference for potentially heterogeneous patterns of association characterizing the expression of bio-molecular pathways across different biologic conditions. We discuss a modeling approach based on Gaussian directed acyclic graphs and provide computational and methodological details needed for posterior inference. Our application finds motivation in reverse phase protein array data from a study on acute myeloid leukemia, where interest centers on contrasting refractory vs. relapsed patients. We illustrate the proposed method through both synthetic and case study data.

Key words: Conditional Independence, Directed Acyclic Graphs, Gaussian Markov Models, Reversible Jumps MCMC.

[*5]For Correspondence: Masanao Yajima, e-mail: myajima@fredhutch.org.

1. INTRODUCTION

In "omics" studies, data often manifests as an $n \times p$ matrix Y, which forms the empirical basis for the exploration of scientific hypotheses relating a set of p genes, metabolites, proteins, etc. to an underlying biological process. When inference focuses on molecular interactions, biomelecular pathways are often interpreted as describing marginal or conditional independence restrictions, compatible with a data generating mechanism for Y. See for example: Dobra *and others* 2004 or Telesca *and others* 2012. More precisely, the joint sampling distribution of Y is assumed to factor according to stochastic restrictions encoded in a graphical model $\mathcal{G}$ (Lauritzen, 1996) and inference about molecular pathways is interpreted as inference about a graph $\mathcal{G}$ supporting the observed data Y.

These biomedical studies are often designed to contrast characteristics of population subgroups. In this manuscript we consider the problem of contrasting differential association structures characterizing labeled subsets of a data matrix Y. Our application finds motivation in a proteomics study of Acute Myeloid Leukemia (AML) patients (Tibes *and others* 2006). This study targets a specific proteomic pathway, thought to be associated with disease progression. Interest centers on understanding what components of this pathway behave differently when one compares refractory patients to patients experiencing relapse after standard treatment. Figure 1 shows the targeted protein expression level for AML patients, measured using a high-throughput proteomic technology called reverse phase protein array (RPPA) (Tibes *and others*, 2006), along with the sample partial correlation matrix of the expression levels for both refectory and relapsed patients in the upper and lower triangle respectively. The two sample partial correlation mostly agree with each other, yet there are clear discrepancies that may signify differing interactions mechanism.

There is a large body of literature focused on the estimation of a single graphical model. However, limited attention has been given to the estimation of a collection of graphs ac-

counting for data heterogeneity. Flexible models for heterogeneous association have been introduced by Ickstadt *and others* (2010) and Rodriguez *and others* (2011). These models focus on relaxing the assumption of multivariate Gaussianity, rather then testing specific contrasts in biological pathways. More in line with our inferential goals are the procedures introduced by Guo *and others* (2011) and Danaher *and others* (2011). Both manuscripts consider regularized estimation of inverse covariance matrices, using hierarchical extensions of the graphical lasso, originally proposed by Friedman *and others* (2008). The proposed procedures are shown to be scalable to large undirected graphs, with estimators that enjoy asymptotic consistency.

Beyond statistical exploration of structured dependence, proteomics studies, like the one motivating this manuscript, are inherently designed to test statistical hypotheses that go beyond simpler hypothesis generation exercises. This need finds available statistical techniques to be wanting in several directions. First, substantive understanding of biomolecular interactions is often based on directed graphs. Therefore, statistical inference should consider more general classes of graphical models, that go beyond the estimation of structural zeros in inverse covariances. Second, even in the simplified setting of undirected graphs, the use of lasso-based estimation techniques, beyond exploratory analyses, does raise methodological and theoretical concerns regarding multiplicity correction and error estimation (Kyung *and others*, 2010). Third, gene and protein interactions have been studied for decades, with comprehensive databases providing the opportunity for informed priors on model spaces (The Gene Ontology Consortium, 2000, `http://geneontology.org`; The Kyoto Encyclopedia of Genes and Genomes `http://genome.jp/kegg/`; the BioCarta pathways, `http://biocarta.com/genes/index/asp`; etc.).

This manuscript addresses these issues considering a model of differential interaction based on *Directed Acyclic Graphs* (DAGs). A DAG is a directed graph with no directed cycles,

encoding Markov properties by *d-separation* (Pearl, 1986, 2000). We discuss how scientific knowledge about pathways known a priori can be incorporated in the inference through different classes of prior probabilities. Finally, we propose inference based on Bayesian model determination through decision theoretic principles aimed at controlling false discoveries.

This paper is structured as follows. In Section 2 we propose a Gaussian Differential DAG model followed by computational detail in Section 3. We illustrate the application of the proposed method with a simulated example and an application to the Reverse Phase Protein Array data. We conclude the manuscript with a critical discussion in Section 5.

2. A Model for Differential Interactions

We consider data in the form of an $n \times p$ matrix $Y = [y_{ij}]$, where the rows of Y are assumed to be indexed by two known sample subgroups and are labelled by $\mathbf{s} = (s_1, ..., s_n)'$, with $s_i \in \{0, 1\}$. We refer to the group indexed with $s_i = 0$ as the baseline group and the ones indexed with $s_i = 1$ as the differential group. Within each group, we assume that the joint sampling distribution factors according to a DAG $\mathcal{G}^{(k)} = \{\mathcal{V}, \mathcal{E}^{(k)}\}$, $(k = 0, 1)$; where $\mathcal{V}$ indexes the set of p genes or proteins and $\mathcal{E}^{(k)} \subseteq \{(v_i \to v_j),\ v_\ell \in \mathcal{V}\}$ is the set of edges characterizing $\mathcal{G}^{(k)}$. Finally, the strength of conditional dependence is indexed by two parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Letting $\mathcal{G} = \{\mathcal{G}^{(k)},\ k = 0, 1\}$ denote the collection of graphs associated with Y, the proposed joint probability model is constructed as follows:

$$p(Y, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \mathcal{G} \mid \mathbf{s}) = \underbrace{p(Y \mid \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \mathcal{G}; \mathbf{s})}_{2.1}\ \underbrace{p(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \sigma^2, \mathcal{G}; \mathbf{s})\, p(\sigma^2 \mid \mathcal{G}; \mathbf{s})}_{2.2}\ \underbrace{p(\mathcal{G} \mid \mathbf{s})}_{2.3}. \tag{2.1}$$

The model includes two separate graphs, $\mathcal{G}^{(0)} = \{\mathcal{V}, \mathcal{E}^{(0)}\}$ for the baseline sample $(s_i = 0)$ and $\mathcal{G}^{(1)} = \{\mathcal{V}, \mathcal{E}^{(1)}\}$ for the differential sample $(s_i = 1)$. Our inference will focus on identifying a set of differential interactions partially indexed by the set $\{(\mathcal{E}^{(0)})^c \cap \mathcal{E}^{(1)}\} \cup \{\mathcal{E}^{(0)} \cap (\mathcal{E}^{(1)})^c\}$.

Each sub model is discussed in individual sections, as indicated by the under-braces.

2.1 *Sampling model:* $p(Y \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{G}; \mathbf{s})$

We assume that Y can be subdivided into two groups $Y^{(0)}$ and $Y^{(1)}$ each of size n_0 and n_1, where $n_0 + n_1 = n$. Given the graphical structures $\mathcal{G}^{(0)}$ and $\mathcal{G}^{(1)}$, let $pa_k(j)$ denote the parent nodes of vertex j, induced by graph $\mathcal{G}^{(k)}$, $(k = 0, 1)$, and $Y_j = (y_{1j}, \ldots, y_{nj})^T$, denote the j^{th} column of Y, $(j = 1, \ldots, p)$. The joint likelihood is defined as

$$p(Y \mid \cdot) = \prod_{k=0}^{1} \prod_{j}^{p} p\left(Y_j^{(k)} \mid Y_{pa_k(j)}^{(k)}, \mathcal{G}^{(k)}, \cdot\right) = \prod_{k=0}^{1} \prod_{j}^{p} \prod_{i=1}^{n_k} p\left(y_{ij}^{(k)} \mid Y_{pa_k(j)}^{(k)}, \mathcal{G}^{(k)}, \cdot\right).$$

In the multivariate Gaussian framework, we can express each of $p\left(y_{ij}^{(k)} \mid Y_{pa_k(j)}^{(k)}, \mathcal{G}^{(k)}, \cdot\right)$ as a conditional linear model of the form

$$y_{ij}^{(k)} \mid Y_{pa_k(j)}^{(k)}, \alpha_j, \boldsymbol{\beta}_j, \boldsymbol{\gamma}_j, \sigma_j^2, \mathcal{G}^{(k)} \sim N\left(\alpha_j + \sum_{l \in pa_k(j)} y_{il}^{(k)} \left(\beta_{lj} + \gamma_{lj}\, I\{k = 1\}\right), \sigma_j^2\right), \qquad (2.2)$$

for $i = 1, \ldots, n$, $j = 1, \ldots, p$, and $k = 0, 1$. Here α_j parametrizes a conditional mean value and σ_j^2 a conditional variance. In the foregoing formulation, β_{lj} indexes the strength of association between $y_{il}^{(0)}$ and $y_{ij}^{(0)}$, with the convention $\beta_{lj} = 0$ when $(l \to j) \notin \mathcal{E}_0$. The strength of association between $y_{il}^{(1)}$ and $y_{ij}^{(1)}$ is defined by $\beta_{lj} + \gamma_{lj}$, with $(\beta_{lj} + \gamma_{lj}) = 0$ whenever $(l \to j) \notin \mathcal{E}^{(1)}$. In this setting, the parameter γ_{lj} becomes the main quantity of interest as it directly informs the differences in association between subgroup random quantities.

Throughout the manuscript we will refer to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to index full sets of interaction coefficients and a will use a subscript to denote their column subsets. Details about how $\boldsymbol{\gamma}$ is used to index the differences between $\mathcal{E}^{(0)}$ and $\mathcal{E}^{(1)}$ and final inference about differential interactions are discussed in Section 2.2.

2.2 *Priors on continuous parameters α, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and σ_j^2*

The strength of association between random quantities in the baseline group is parametrized through the $\boldsymbol{\beta}$ coefficients. Differential parameters $\boldsymbol{\gamma}$ distinguish the strength of association between the baseline and differential groups. We explicitly address two inferential questions. First, are there differences in patterns of conditional dependence between baseline and differential groups? This question relates to the identification of the set $\{(\mathcal{E}^{(0)})^c \cap \mathcal{E}^{(1)}\} \cup \{\mathcal{E}^{(0)} \cap (\mathcal{E}^{(1)})^c\}$. Second, when considering edges that are shared between both baseline and differential groups, are there significant differences in the way these edges define conditional dependence patters? Here, we consider the set $(\mathcal{E}^{(0)} \cap \mathcal{E}^{(1)})$, but we are specifically interested in the size of γ_{lj}.

We encode these inferential goals directly in the prior. In particular, it is convenient to define latent trinary indicators $z_{lj} \in \{0, 1, 2\}$ distinguishing between three cases: (1) $z_{lj} = 0$, whenever the association implied by the edge $(l \to j)$ is invariant with group label; (2) $z_{lj} = 1$, whenever $(l \to j) \in \mathcal{E}^{(0)}$, but $(l \to j) \notin \mathcal{E}^{(1)}$; and (3) $z_{lj} = 2$, whenever the strength of association implied by $(l \to j)$ varies with group label, beyond the scenario of case (2). Note that $\mathcal{G}^{(0)}$ and $\mathbf{Z} = [z_{lj}]$ together deterministically determine $\mathcal{G}^{(1)}$. However, $\mathbf{Z}$ codes information beyond topological differences between $\mathcal{G}^{(0)}$ and $\mathcal{G}^{(1)}$ by also indicating common edges with varying strength across conditions.

For any configuration of $\mathcal{G}^{(0)}$ and $\mathbf{Z}$ (implying also a configuration of $\mathcal{G}^{(1)}$), and for all $j = 1, ..., p$; let $\tilde{\mathbf{Y}}_{j0}$ denote the matrix including all parents of Y_j in $\mathcal{G}^{(0)}$. Similarly, let $\tilde{\mathbf{Y}}_{j1}$ denote the matrix including all parents of Y_j in $\mathcal{G}^{(1)}$, for which $z_{\ell j} = 2$. We define a differential autoregression matrix $\tilde{\mathbf{X}}_j$ s.t.

$$\tilde{\mathbf{X}}_j = \left(\begin{array}{c|c|c} \mathbf{1}_{n_0} & \tilde{\mathbf{Y}}_{j0}^{(0)} & \mathbf{0} \\ \mathbf{1}_{n_1} & \tilde{\mathbf{Y}}_{j0}^{(1)} & \tilde{\mathbf{Y}}_{j1}^{(1)} \end{array}\right), \text{ with associated coefficient vector } \mathbf{B}_j^T = (\alpha_j, \boldsymbol{\beta}_j^T, \boldsymbol{\gamma}_j^T).$$

Using this notation, the conditional likelihood in (2.2) is represented as

$$Y_j \mid \tilde{\mathbf{X}}_j, \mathcal{G}_0, \mathbf{Z}, \mathbf{B}_j, \sigma_j^2 \sim N\left(\tilde{\mathbf{X}}_j \mathbf{B}_j,\ \sigma_j^2 \mathbf{I}_n\right).$$

A conditional Zellner g-prior (Zellner, 1986) is then defined as

$$\mathbf{B}_j \mid \mathcal{G}_0, \mathbf{Z}, \sigma_j^2 \sim N\left(\mathbf{0},\ \omega_j \sigma_j^2 \left(\tilde{\mathbf{X}}_j^T \tilde{\mathbf{X}}_j\right)^{-1}\right).$$

Different choices of ω_j have been proposed in the model selection literature (Celeux *and others*, 2012). We consider the benchmark prior $\omega_j = max(n, \tilde{p}_j^2)$ (Fernandez *and others*, 2001), where $\tilde{p}_j$ is the number of columns in $\tilde{\mathbf{X}}_j$. Alternatively, data-dependent choices have been discussed in Consonni and La Rocca (2012).

Given this parametrization, posterior inference over differential patterns of interaction focuses on $p(\gamma_{lj} \mid Y)$, as this quantity is directly informative about the size of differences in partial correlation and about the significance of such differences, as measured by $p(\gamma_{lj} \neq 0 \mid Y)$. Finally, dispersions σ_j^2, $j = \{1, \ldots, p\}$, are modeled independently as a conjugate Inverse Gamma priors with hyper parameters $\frac{\delta_j}{2}$ and $\frac{\tau_j}{2}$, so that $\sigma_j^2 \sim IG\left(\frac{\delta_j}{2}, \frac{\tau_j}{2}\right)$.

2.3 *Model space priors*

Let $\mathcal{G}^\star = \{\mathcal{V}, \mathcal{E}^\star\}$, be a prior graph derived through direct expert opinion or constructed using public databases (for example, a DAG derived through Gene Ontology, `http://genontology.org`). In order to define a prior centered on $\mathcal{G}^\star$, we follow Telesca *and others* (2012) and define an exponentially decaying function centered around the prior graph $\mathcal{G}^\star$, so that $p(\mathcal{G}^{(k)} \mid \psi;\ \mathcal{G}^\star) \propto \psi^{d(\mathcal{G}^{(k)}, \mathcal{G}^\star)}, \quad \psi \in [0, 1]$; where the graph distance function is defined as $d(\mathcal{G}^{(k)}, \mathcal{G}^\star) = |\mathcal{E}^{(k)c} \cap \mathcal{E}^\star| + \delta |\mathcal{E}^{(k)} \cap \mathcal{E}^{\star c}|, \delta \geqslant 1$. This prior depends on inclusion probabilities ψ, on which we place a hierarchical $beta(a_\psi, b_\psi)$ prior, as suggests in Scott and Berger (2006) to control for error control in the multiple comparison problem. Marginalizing over ψ the model space prior becomes $p\left(\mathcal{G}^{(k)} \mid \mathcal{G}^\star\right) \propto B(a_\psi + d(\mathcal{G}^{(k)}, \mathcal{G}^\star), b_\psi)/B(a_\psi, b_\psi)$,

which is strictly decaying function with respect to $d(\mathcal{G}^{(k)}, \mathcal{G}^{\star})$. Given the prior graph $\mathcal{G}^{\star}$, we model $\mathcal{G}^{(0)}$ and $\mathcal{G}^{(1)}$ as partially exchangeable, thus we define the overall model space prior as $p\left(\mathcal{G}^{(0)}, \mathcal{G}^{(1)} \mid \psi;\ \mathcal{G}^{\star}\right) = p\left(\mathcal{G}^{(0)} \mid \psi;\ \mathcal{G}^{\star}\right) p\left(\mathcal{G}^{(1)} \mid \psi;\ \mathcal{G}^{\star}\right)$. We note that this prior becomes a simple sparsity prior when $\mathcal{E}^{\star}$ is empty, making it useful for de-novo analyses. Our simulation (see supplement) do however outline the advantage of partially informative model space priors, even in the context of partial misspecification.

3. Posterior Simulation

To obtain draws from the posterior distribution $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \mathcal{G}^{(1)}, \mathcal{G}^{(0)} \mid Y)$, we use a Gibbs sampling algorithm similar to Smith and Kohn (1996). with slight modification in the proposal strategy. The conditionally conjugate structure of priors and likelihood allows for the computation of models marginal likelihood (see appendix D for details) and efficient transitions in the model space (Eklund and Karlsson, 2007).

The proposed transition sequence extends the method proposed by Fronk and Giudici (2004). to allow for transitions over the additional structure $\mathcal{G}^{(1)}$. We note that, in our formulation, $\mathcal{G}^{(1)}$ is fully determined by $\mathcal{G}^{(0)}$ and latent components $\mathbf{Z}$. It follows that, systematic or random scans through the following transition sequence define an ergodic Markov chain that we can use to sample from posterior quantities of interest. In particular, we consider the following transition sequence

$$(i) \quad \mathcal{G}^{(0)}, \mathbf{Z} \mid \mathcal{G}^{(1)} \quad (3.1) \qquad (ii) \quad \mathbf{Z} \mid \mathcal{G}^{(0)} \quad (3.2)$$

$$(iii) \quad \sigma^2 \mid \mathcal{G}^{(0)}, \mathbf{Z} \quad (3.3) \qquad (iv) \quad \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathcal{G}^{(0)}, \mathbf{Z}, \sigma^2 \quad (3.3)$$

Details about each transition are explained in the corresponding sections.

3.1 *Updating the baseline DAG* $\mathcal{G}^{(0)}$

To update $\mathcal{G}^{(0)}$, we select an edge $(l \to j)$ at random, i.e., using a uniform distribution over all possible edges. If $(l \to j) \notin \mathcal{E}^{(0)}$ we propose its addition to $\mathcal{E}^{(0)}$ (birth); if $(l \to j) \in \mathcal{E}^{(0)}$ we propose removal (death); if $(l \leftarrow j) \in \mathcal{E}^{(0)}$ but $(l \to j) \notin \mathcal{E}^{(0)}$ then propose to remove $(l \leftarrow j)$ and add $(l \to j)$ (switch). This transition can be represented through changes in the differential autoregressive matrix $\tilde{\mathbf{X}}_j$ consistent with the proposed $\mathcal{G}^{(0)\prime}$ and $\mathbf{Z}'$. Acceptance of a move is defined as

$$R(\mathcal{G}^{(0)}, \mathcal{G}^{(0)\prime}) = 1 \wedge \frac{p\left(Y_j \mid \tilde{\mathbf{X}}'_j\right) p(\mathcal{G}^{(0)\prime}) p(\mathbf{Z}') q(\mathcal{G}^{(0)} \mid \mathcal{G}^{(0)\prime}) q(\mathbf{Z} \mid \mathbf{Z}')}{p\left(Y_j \mid \tilde{\mathbf{X}}_j\right) p(\mathcal{G}^{(0)}) p(\mathbf{Z}) q(\mathcal{G}^{(0)\prime} \mid \mathcal{G}^{(0)}) q(\mathbf{Z}' \mid \mathbf{Z})}.$$

The moves are deterministic except for the case when $(l \to j) \notin \mathcal{E}^{(0)}$ but $(l \to j) \in \mathcal{E}^{(1)}$ where we propose z_{lj} as 0 or 2 with equal probability. For the list of all possible moves see supplementary table 3 . Computational efficiency can be gained by pre-computing $\tilde{\mathbf{X}}_j^T \tilde{\mathbf{X}}_j$ and $\tilde{\mathbf{X}}_j^T \mathbf{Y}_j$ for saturated models. Details are discussed in Eklund and Karlsson (2007).

3.2 *Updating the differential model space through latent indicators* $\mathbf{Z}$

Given the baseline graph $\mathcal{G}^{(0)}$ we propose to move over the differential model space updating the latent variables z_{lj}. Updates in the state of $\mathbf{Z} = [z_{lj}]$ will also define changes in $\mathcal{G}^{(1)}$. We select an edge $(l \to j)$ at random. Depending on the current state of $\mathcal{G}^{(0)}$ and z_{lj}, we consider the proposal transitions summarized in Table 2. The proposed extension on simple birth, death and switch moves allows for transitions that do not alter the structure of $\mathcal{G}^{(1)}$ but change the magnitude of association (move#4,7). As before, all moves define a new differential autoregression matrix $\tilde{\mathbf{X}}'_j$. The acceptance ratio is defined as

$$R_z(\mathcal{G}^{(1)}, \mathcal{G}^{(1)\prime}) = 1 \wedge \frac{p\left(Y_j \mid \tilde{X}'_j\right) p(\mathcal{G}^{(1)\prime}) q(\mathbf{Z} \mid \mathbf{Z}')}{p\left(Y_j \mid \tilde{X}_j\right) p(\mathcal{G}^{(1)}) q(\mathbf{Z}' \mid \mathbf{Z})}.$$

3.3 *Updating σ^2, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$*

Given $\mathcal{G}^{(0)}$ and $\mathbf{Z}$, sampling σ_j^2 from their marginal posterior distributions and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ given σ^2 from their conditional posterior distributions is achieved through direct simulation. Detailed calculations are reported in the supplemental appendix.

3.4 *Posterior Summaries*

Posterior probabilities $p(\mathcal{G} \mid Y)$, $p(\mathbf{z} \mid Y)$ and corresponding MCMC samples characterize our knowledge about baseline and differential interactions in light of the data. Based on these quantities, the main inferential goal is to select representative baseline and differential graphs, say $\mathcal{G}^{(0)*}$ and $\mathcal{G}^{(1)*}$. While posterior probabilities do summarize evidence about interaction structures, selecting a point estimate in these model spaces requires further consideration.

Given a joint model on edge and parameter inclusion probabilities, in the Bayesian framework, selection of point estimators for interaction structures $\mathcal{G}^{(0)}$ and $\mathcal{G}^{(1)}$ usually translates into the appropriate definition of a cutoff value for posterior inclusion probabilities (Scott and Berger, 2006; Müller *and others*, 2006). A cutoff threshold is often determined in order to ensure optimization of a chosen loss function.

An alternative operational strategy is to select a point estimator, on the basis of multiple comparison arguments. An often used error rate is the false discovery rate (FDR) (Benjamini and Hochberg, 1995). Rules as discussed in Benjamini and Hochberg (1995) control the frequentist expectation of the error rate across repeat experimention. Several authors chose instead to control the posterior expectation of the same error rate. See, for example, Newton (2004). From a Bayesian perspective, Müller *and others* (2006) illustrate a decision-theoretic procedure aimed at minimizing the posterior expected false negative rate ($\overline{\text{FNR}}$), while controlling for the posterior expected false discovery rate ($\overline{\text{FDR}}$) at a level α. The rest of this article is based on results obtained under Bayesian FDR control proposed by Müller *and*

others (2006). Details, along with simulations, can be found in the supplement. Our studies show that the procedure controls for empirical FDR and FNR at the specified error rate.

4. CASE STUDY: Acute Myeloid Leukemia

We apply our model to the data from a study of Acute Myeloid Leukemia (AML) obtained using the reverse phase protein arrays (RPPA) (Tibes *and others*, 2006). RPPA is a high-throughput proteomic technology that provides a quantification of the expression for specifically targeted proteins selected from molecular pathways. Our study is based on RPPA of bone marrow specimens collected from 435 newly diagnosed AML patients; 332 primary refractory patients and 103 relapsed patients. We will call the refractory patients the baseline group and the relapsed patients the differential group.

Scientific interest centers on the identification of differential signal transduction pathways (STPs), which are suspected to contribute to leukemogenesis by perturbing the rates of proliferation, differentiation, and apoptosis (Kornblau *and others*, 2006). The objective of this study is to investigate the difference in interactions of important protein markers related to AML for the refractory patients and the relapsed patients. We selected 29 proteins in signal transduction, apoptosis, and cell cycle regulatory pathways and studied their expression profiles in all 435 samples. An attractive feature of the AML data under study is that the number of samples (n = 435) is much greater than the number of proteins (p = 29), which provides an opportunity for principled inference about differential interaction structures on the basis of a highly structured stochastic system.

The prior signaling network was developed based on published articles from PubMed searches as well as from the connections map in Signal Transduction Knowledge Environment (http://www.stke.org). Other prior distributions on the parameters were selected as vague as possible to show that it is suitable for initial studies since the likelihood will dominate

the posterior when the sample size is large. The two parameters of the dispersion parameter σ_l^2 where set to 0.5 and 0.5. The prior on ψ is set to $Beta(1, p^2)$. We ran our algorithm for 2 million iteration saving every 100th sample while discarding the initial 1 million samples.

For the decision rule we employed the criteria of Bayesian FDR control method proposed by (Müller *and others*, 2006) that simultaneously controls for the expected posterior false discovery rates and false negative rates. We control the expected posterior FDR at 1% for both baseline and differential associations separately.

Figure 2 is a network representation of the estimated graph in the form of essential graph (Chickering, 2002) for the refractory and relapsed patients. The bottom three figures show edges that patients in two conditions agree on, that exists only in the refractory patients and that exists only in the relapsed patients. A solid line indicates a positive coefficient estimate and a dotted line indicates a negative coefficient.

While we maintain that our findings are exploratory, we have observed that selected differential interaction patterns have been confirmed in the literature as potential indicators of more aggressive forms of AML. For example, we find a the striking differential role of the YAP protein (Yes associated protein) as a hub node in the relapse sample. This protein plays a key role in the Hippo pathway, usually associated with the regulation of proliferation and apoptosis. While the role of Hippo has been established in solid tumors, only recent reports (Kornblau *and others*, 2013) associate YAP with differential prognostics in AML.

Figure 3 is the estimated posterior inclusion probability. The figures for the estimated mixing proportion and the posterior density plot of the coefficients can also be found in the supplementary materials. A comprehensive bio-medical interpretation of our findings is perhaps out of the scope or this paper, but it is our hope that our illustration shows the potential and practical relevance of the proposed method.

To validate our modeling approach, we compared our proposal to a model ignoring sample

labeling and fitting a Gaussian DAG model without differential interactions. The result showed that the difference in average log marginal likelihood was 325.19 and the difference in log conditional predictive ordinate was 75.04, both suggesting strong evidence for the differential association.

5. DISCUSSION

We proposed a probability model for inference on differential interaction in Gaussian DAGs. The proposed framework is likely to be particularly useful when primary interest focuses on potential contrasts characterizing the association structure between known subgroups of a given sample. Although we worked on a case where there are only two subgroups, the method is directly generalizable to the case of k subgroups. Monte Carlo studies (supplementary material) show the model to be robust against prior graph misspecification. Inference yields well calibrated empirical FDR and FNR scaling to larger graphs and sample sizes. We evaluated our method analyzing data generated from a synthetic experiment and showed that our inferences have desirable operative characteristics (see the supplementary materials for details). The application of the proposed model to the analysis of RPPA data in AML identified interesting differential regulation patterns, distinguishing refractory from relapsed patients. While we are well aware that our model belongs to the class of hypothesis generation tools, we remark that the proposed methodology avoids the use of step-wise analyses and ad hock penalization choices, providing a principled tool for inference on differential networks.

Although we avoided all the costly matrix inversion steps in our computation, when dealing with large and dense graphs, the sheer size of the space to explore could make the computation challenging. Alternative strategies may be necessary for these large problems.

The proposed framework of differential network inference could be extended beyond the

multivariate Gaussian distribution. Models space and interaction parameters could, for example be applied to the approach of Telesca *and others* (2012), who show how to incorporate heavy tails in the observations by the use of a mixture model. As for the case of discrete and mixed data, the copula Gaussian graphical model framework proposed by Dobra and Lenkoski (2011) could be easily expanded using a modeling strategy similar to the one proposed in this paper.

Extension beyond DAGs may be desirable in many applied settings. For example, in the setting of Reciprocal Graphs (Koster, 1996), used in Telesca *and others* (2011) one may allow baseline and differential models, to be defined in terms of undirected edges as well as the directed ones, with the possibility of including cycles and reciprocal relations. We should also point out that the same idea could of course be applied to undirected graphical models. While these extension are conceptually trivial, coherent multivariate representation and computational constraints may require extensive additional work.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(1), 289–300.

Celeux, G., El Anbari, M., Marin, J.M. and Robert, C.P. (2012). Regularization in regression: comparing Bayesian and frequentist methods in a poorly informative situation. *Bayesian Analysis* **7**(2), 477–502.

Chickering, D M. (2002). Learning equivalence classes of Bayesian-network structures. *J. Mach. Learn. Res.* **2**(445-498).

Consonni, G and La Rocca, L. (2012). Objective Bayes factors for Gaussian directed acyclic graphical

models. *Scandinavian Journal of Statistics* **39**(4), 743–756.

Danaher, P., Wang, P. and Witten, D.M. (2011). The joint graphical lasso for inverse covariance estimation across multiple classes. *arXiv preprint arXiv:1111.0324*.

Dobra, Adrian, Hans, Chris, Jones, Beatrix, Nevins, J.R.Joseph R., Yao, Guang and West, Mike. (2004). Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis* **90**(1), 196 – 212.

Dobra, A. and Lenkoski, A. (2011). Copula Gaussian graphical models and their application to modeling functional disability data. *The Annals of Applied Statistics* **5**(2A), 969–993.

Eklund, Jana and Karlsson, Sune. (2007). *Computational Efficiency in Bayesian Model and Variable Selection*. Central Bank of Iceland, Economics Department.

Fernandez, C., Ley, E. and Steel, M.F.J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics* **100**(2), 381–427.

Friedman, Jerome H, Hastie, T J and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441.

Fronk, Eva-Maria and Giudici, Paolo. (2004). Markov Chain Monte Carlo model selection for DAG models. *Statistical Methods & Applications* **13**, 259–273. 10.1007/s10260-004-0097-z.

Guo, Jian, Levina, Elizaveta, Michailidis, George and Zhu, Ji. (2011). Joint estimation of multiple graphical models. *Biometrika* **98**(1), 1–15.

Ickstadt, K., Bornkamp, B., Grzegorczyk, M., Wieczorek, J., Sheriff, M., Grecco, H. and Zamir, E. (2010). Nonparametric Bayesian networks. *Bayesian Statistics* **9**, 283–316.

Kornblau, SM, Hu, CW, Qiu, Yi Hua, Yoo, SY, Zhang, N, Kadia, TM, Ferrajoli, A, Coombes, KR and Qutub, AA. (2013). Hippo pathway (hp) activity in acute myelogenous leukemia (aml): Different prognostic implications of taz versus yap inactivation by phosphorylation. *Blood* **15**(21), 1337.

Kornblau, Steven M., Womble, Matthew, Qiu, Yi Hua, Jackson, C. Ellen, Chen, Wenjing, Konopleva, Marina, Estey, Elihu H. and Andreeff, Michael. (2006). Simultaneous activation of multiple signal transduction pathways confers poor prognosis in acute myelogenous leukemia. *Blood* **108**(7), 2358–2365.

Koster, J. T. A. (1996). Markov properties of nonrecursive causal models. *The Annals of Statistics* **24**(5), pp. 2148–2177.

Kyung, M., Gill, J., Ghosh, M. and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis* **5**(2), 369–411.

Lauritzen, S.L. (1996). *Graphical models*, Volume 17. Oxford University Press, USA.

Müller, P., Parmigiani, G. and Rice, K. (2006). FDR and Bayesian multiple comparisons rules. *Johns Hopkins University, Dept. of Biostatistics Working Papers*, 115.

Newton, Michael A. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics (Oxford)* **5**(2), 155–176.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial intelligence* **29**(3), 241–288.

Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge Univ Press.

Rodriguez, A., Lenkoski, A. and Dobra, A. (2011). Sparse covariance estimation in heterogeneous samples. *Electronic Journal of Statistics* **5**, 981–1014.

Scott, James G. and Berger, James O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference* **136**(7), 2144 – 2162.

Smith, Michael and Kohn, Robert. (1996). Nonparametric regression using bayesian variable selection. *Journal of Econometrics* **75**(2), 317–343.

Telesca, D., Muller, P., Kornblau, S., Suchard, M. and Ji, Y. (2011). Modeling protein expression and protein signaling pathways. *COBRA Preprint Series*, 87.

Telesca, D., Müller, P., Parmigiani, G. and Freedman, R.S. (2012). Modeling dependent gene expression. *The Annals of Applied Statistics* **6**(2), 542–560.

Tibes, Raoul, Qiu, YiHua, Lu, Yiling, Hennessy, Bryan, Andreeff, Michael, Mills, Gordon B. and Kornblau, Steven M. (2006). Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Molecular Cancer Therapeutics* **5**(10), 2512–2521.

Zellner, A. (1986). *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Chapter On assessing prior distributions and Bayesian regression analysis with g-prior distributions. Amsterdam: North-Holland,, pp. 233–243.

6. Supplementary Materials

The reader is referred to the on-line Supplementary Materials for simulations, technical appendices, and additional figures.

7. Acknowledgements

Yuan Ji's research is in part supported by NIH R01 CA132897. Peter Mueller's research is in part supported by NIH R01 CA132897 and NIH/NCI R01CA075981. Donatello Telesca's research was in part supported by NIH grant R01 CA158113-01. The authors would like to acknowledge two anonymous referees for their important feedback on the original manuscript.

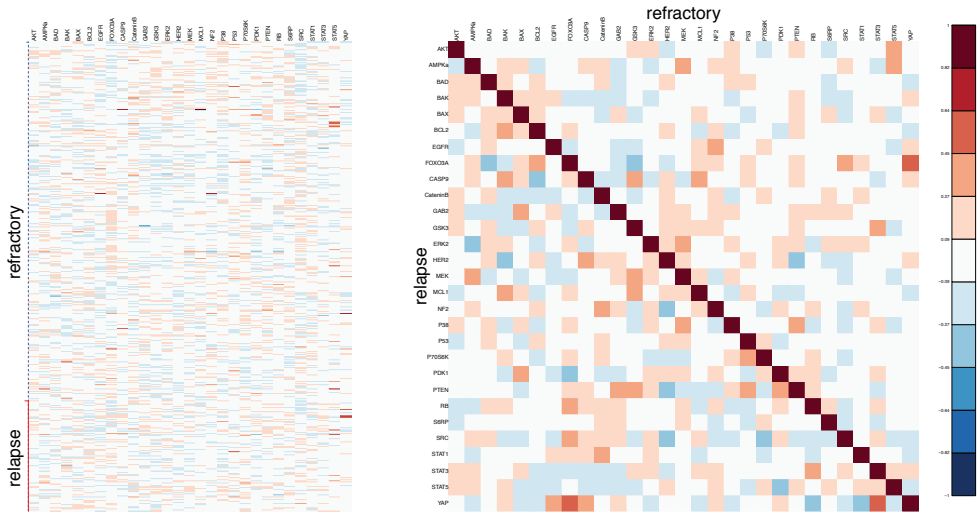

Fig. 1. The observed expression levels of targeted proteins for AML patients quantified using RPPA (left) and a image plot of the sample partial correlation coefficients for refractory patients in the upper triangle and relapsed patients in the lower triangle (right).

Extra Edges
CASP9 → YAP, STAT3 → YAP, SRC → YAP, P38 → YAP, FOXO3A→ YAP,RB → YAP, GSK3→ AKT,PTEN → GSK3, CASP9→ GSK3, BAX → GSK3, BCL2→ BAD
Canceled Edges
NF2 → EGFR, PDK1 → P53, PDK1 → STAT3, BCL2 → PDK1, BCL2 → HER2, GAB2 → HER2, PDK1 → HER2, P38 → BAD, PRC −BAX, FOXO3A− GAB2

Table 1. The list of differential edges.

$\mathcal{G}^{(0)}$	current z_{lj}	proposed z_{lj}	probability	move #	type
$(i,j) \notin \mathcal{E}^{(0)}$	$z_{lj}=0$	$z'_{lj}=2$	1	1	birth
	$z_{lj}=2$	$z'_{lj}=0$	1	2	death
$(i,j) \in \mathcal{E}^{(0)}$	$z_{lj}=0$	$z'_{lj}=1$	1/2	3	death
		$z'_{lj}=2$	1/2	4	differential
	$z_{lj}=1$	$z'_{lj}=0$	1/2	5	birth
		$z'_{lj}=2$	1/2	6	birth
	$z_{lj}=2$	$z'_{lj}=0$	1/2	7	differential
		$z'_{lj}=1$	1/2	8	death

Table 2. Proposal transition scheme for exploration of the differential model space to update z_{lj}. The transition probabilities 1 through 8 include four pairs of moves that are each other's inverse: (1,2), (3,5), (4,7) and (6,8).

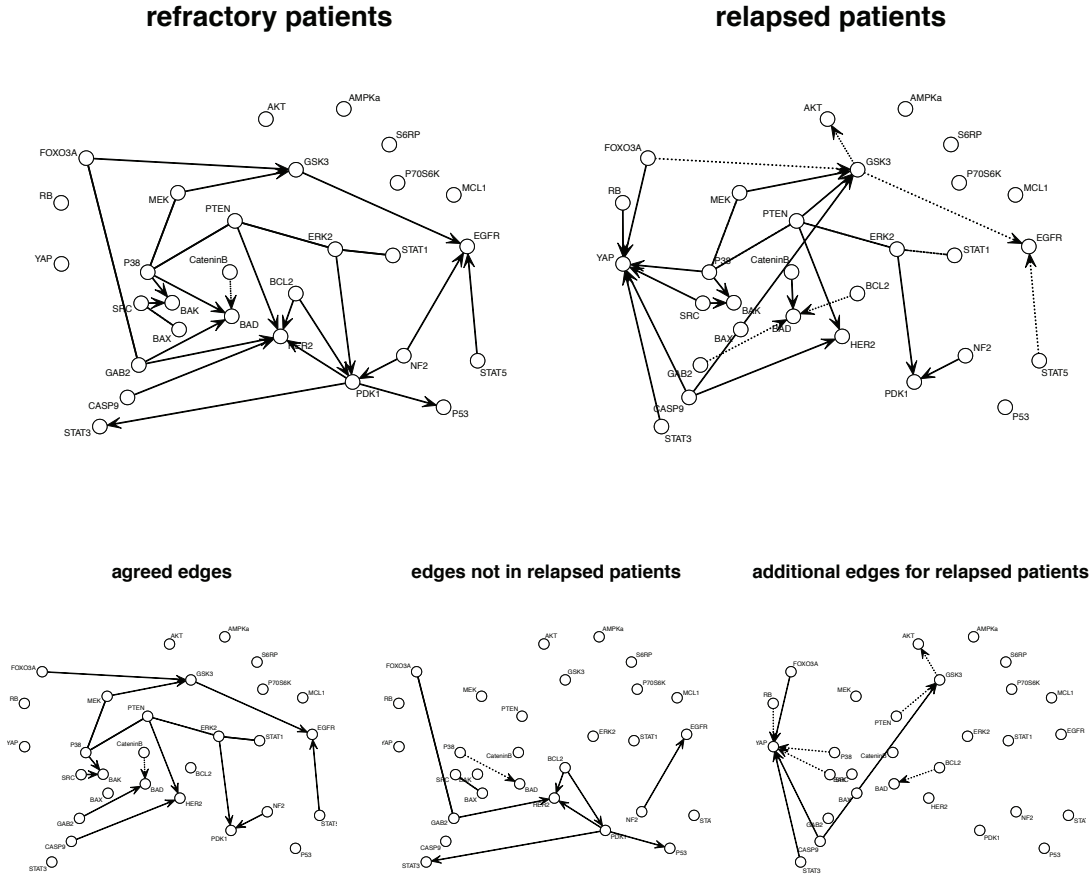

Fig. 2. Network representation of the estimated protein network for refectory patients and relapsed patients for associations chosen at α level of 0.01 using the decision criteria of Müller *and others* (2006) . The positive associations are shown as a solid line and negative associations are shown as a dotted line. The bottom three plots classify the edges into three categories: the edges that two groups agree on, the edges that does not exist in the differential graph, and edges that only exist in the differential graph.

8. Figures and Tables

[*Revised November, 2014*]

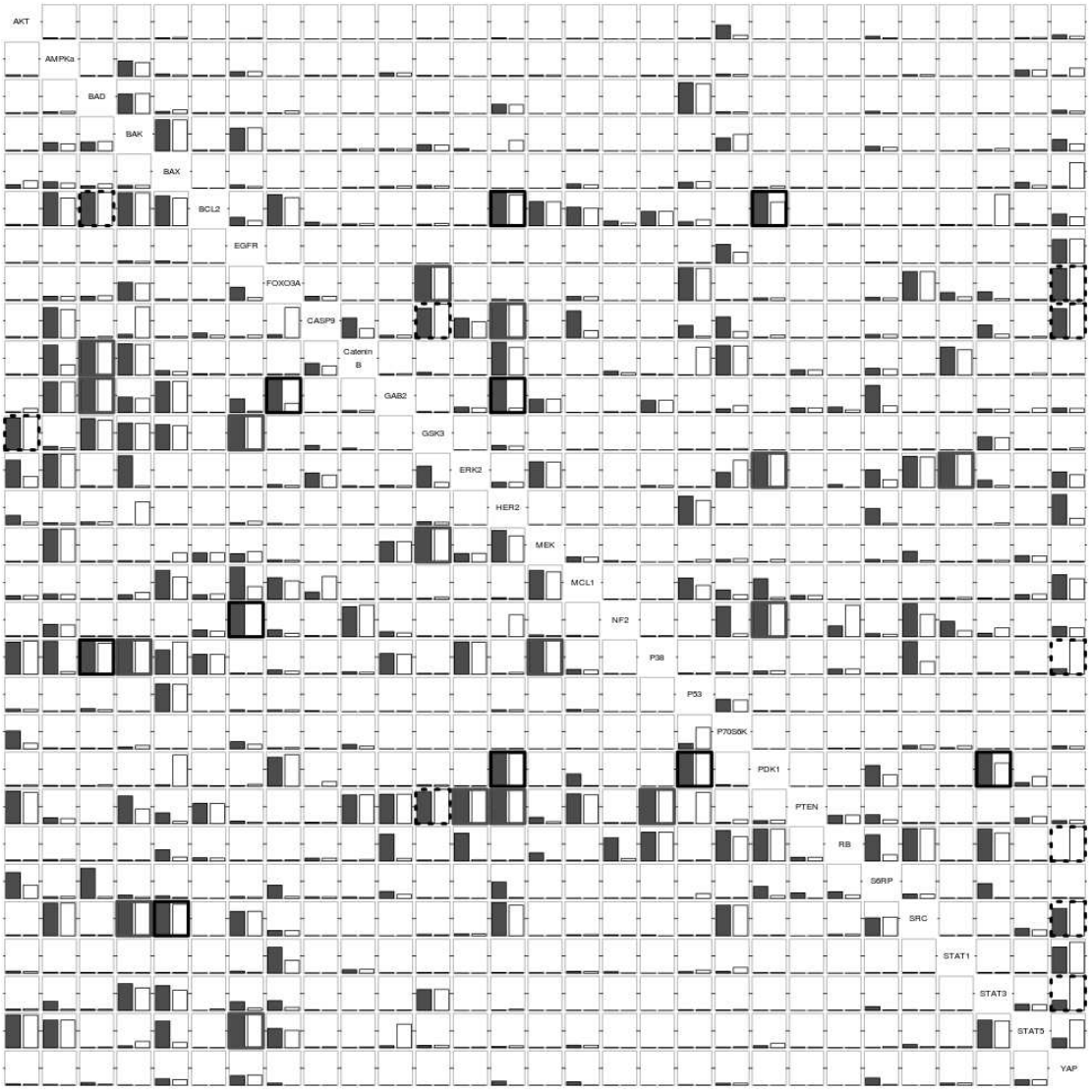

Fig. 3. Barplot of the estimated edge inclusion probability for the refractory patients (left) and the relapsed patients (right) for each edge.