Published in final edited form as:

J Appl Stat. 2014 ; 41(11): 2483–2492. doi:10.1080/02664763.2014.920776.

A Bayesian hierarchical model for inference across related reverse phase protein arrays experiments

Riten Mitra^{a,*}, Peter Müller^b, Yuan Ji^c, Yitan Zhu^c, Gordon Mills^d, and Yiling Lu^d

^aICES, University of Texas at Austin, Austin, TX, USA

^bDepartment of Mathematics, University of Texas at Austin, Austin, USA

^cCCRI, Northshore University HealthSystem, Chicago, IL, USA

^dDepartment of Systems Biology, UT MD Anderson Cancer Centre, Houston, TX, USA

Abstract

We consider inference for functional proteomics experiments that record protein activation over time following perturbation under different dose levels of several drugs. The main inference goal is the dependence structure of the selected proteins. A critical challenge is the lack of sufficient data under any one drug and dose level to allow meaningful inference on dependence structure. We propose a hierarchical model to implement the desired inference. The key element of the model is a shared dependence structure on (latent) binary indicators of protein activation.

Keywords

timecourse; graphical; Bayesian; hierarchical model; protein networks

1. Introduction

Time-course measurements capture the dynamic behavior of a system. Some interesting examples of time-course measurements arise from high-throughput proteomic experiments that record expressions of a set of protein markers over time. When the proteins belong to the same biological pathway, the behavior of a protein i at time t is influenced by the others, potentially including itself, at time $t - 1$. We propose a sparse Bayesian graphical model that captures this association. The proposed model has three distinct novel features. First, we combine inference across different conditions such as dose levels and different drugs by enforcing a common network structure across all conditions, allowing us to borrow information and significantly strengthen inference under any one of the conditions. Importantly, the data are not exchangeable across conditions, making a straightforward hierarchical model inappropriate. Second, instead of directly modeling the raw data, we model latent dichotomized binary signals that represent activation status of the proteins. Third, we propose a flexible graphical prior model that imposes a sparsity constraint on the network. Several methods have been devised to analyze time-course data in proteomic and genetic experiments. For example, Inoue *et al.* [7] combined clustering and Bayesian

*Corresponding author. riten82@gmail.com.

networks to identify the dependence of gene expression over time. Their approach was to simultaneously cluster gene expressions and model dependence between clusters across time. A network over the cluster parameters describes dynamically the evolution of the expression profiles over time. Telesca *et al.* [19] used Bayesian hierarchical network models to assess the temporal structure of microarray gene expression data. In a similar spirit, recent methods in the analysis of proteomic time-course experiments focused on discovering protein–protein interaction networks. For example, Bender *et al.* [1] used dynamic-dependent networks to model the signaling dynamics of 16 proteins in the ERBB pathway for a breast cancer cell line. They use a Boolean propagation mechanism which defines discrete state transitions for a given network structure. The optimal state transition is inferred through a hidden Markov model that is then estimated with likelihood-based methods. The limited number of repeat observations in our data preclude the application of the models discussed above for separate inference of dependence structure under each condition. Instead we use a Bayesian strategy to borrow strength across different experiments. Our proposed model can thus extract the inherent connections between proteins which exist across all different experimental conditions.

The rest of the article is organized as follows. In the next section, we describe the data structure. We develop the sparse hierarchical Bayesian graphical model in Section 4 and complete the inference model with a prior on latent activation states and a sampling model for the observed data in Section 3. In Section 6, we describe the implementation of posterior simulation and results for the time-course proteomics data. We conclude in Section 7 with a short review of the model and its relevance to the current state of biological research.

2. Data

Typically, a proteomics profiling experiment begins with a stimulus that targets the pathway of interest. The expression of all proteins in the pathway is then measured over time. Such experiments allow investigators to explore simultaneously the time-course behavior of a protein marker under the given stimulus, as well as the dependence between different markers.

Reverse phase protein arrays (RPPA) [20] are a particular example of functional proteomic experiments. RPPA experiments simultaneously measure expression for targeted proteins for a large number of samples. An array or slide in the experiment consists of up to thousands of individual patient samples. Each sample is replicated in batches of four-by-four dot matrices. The individual samples on a slide are then hybridized against an antibody that binds to one specific protein. Investigators wanting to study a particular pathway design a RPPA experiment with each array corresponding to a specific protein antibody in the pathway.

In this paper we analyze data from a pathway-inhibition RPPA experiment on ovarian cancer cell lines. The experiment aimed to discover the temporal behavior of $m = 66$ disease markers in the PI3K pathway. The experiment starts by treating an ovarian cancer cell line with a specific drug. The cell line was then observed at $T = 8$ time points. In the following

discussion we index the time points as $t = 1, \dots, T$. The actual times were 0, 5, 15, 30, 60, 90, 120, and 240 min.

A total of $m = 66$ protein markers, $i = 1, \dots, n$, were measured. For each protein, we obtained three replicates of their expression intensities. The same experiment was repeated using four different dose levels of the drug.

We preprocessed the raw data to make expressions comparable across proteins, following the same steps as described in [13]. The three replicate measurements were averaged to give a single measurement for each dose level, drug, protein and time. The preprocessed data are arranged in an $m \times 4 \times 3 \times T$ array $\mathbf{y} = [y_{ik\ell t}]$ of expressions. Here, $i = 1, \dots, m$ indexes the proteins, the index $k = \{1, 2, 3, 4\}$ represents dose levels, $\ell = 1, 2, 3$ represents drugs and $t = 1, \dots, 8$ indexes the times for the repeat measurements. That is, the measurement $y_{ik\ell t}$ is the expression of protein i at dose k , drug ℓ and time t .

In preliminary data analysis, we considered separate analyses for different drug/dose combinations. For each of these conditions, we inferred networks on the basis of just 8 data points. We found that the posterior distribution on the dependence structure was too diffuse for any useful inference. Substantial statistical power could be gained by borrowing strength across different conditions. In the following discussion we develop an approach to formalize such borrowing of strength.

3. A hierarchical model across related experiments

The small sample sizes make it impractical to proceed with separate analyses for each drug/dose combination. We instead proceed with a hierarchical model. The main innovation in the proposed hierarchical model is that the submodels share the dependence structure of underlying binary indicators of protein activation. In particular, we assume that the shared dependence structure only at the level of such latent binary indicators. Beyond the (binary) biological signal we allow independent residuals across experiments.

This modeling strategy is motivated by the observation that many molecular pathways involve substantial noise, and that the actual dependence acts only between latent activation states of the proteins. We denote the latent signal for protein i at time t for dose k and drug ℓ as $e_{ik\ell t} \in \{0, 1\}$ with $e_{ik\ell t} = 1$ indicating that the protein is activated. The reduction to a binary underlying signal is similar to the probability of expression model proposed in [16] for microarray data, who use the same modeling strategy for inference about levels and differences of gene expression. In contrast, here the central inference question is learning about the dependence structure.

Let $\mathbf{e}_{\bullet k, \ell, t} = (e_{ik\ell t}, i = 1, \dots, m)$ denote the latent indicators for a particular drug, dose and time and let $\mathbf{e}_{\bullet k, \ell} = (\mathbf{e}_{\bullet k, \ell, t}; t = 1, \dots, T)$ denote all data for an experiment with a particular drug/dose combination. We formalize the assumption of a shared dependence structure of $e_{ik\ell t}$ across drugs and doses by specifying the joint distribution $p(\mathbf{e}_{\bullet k, \ell})$ as a logistic autoregressive model $p(\mathbf{e}_{\bullet k, \ell} | \mathbf{e}_{\bullet k, \ell, t-1}, \boldsymbol{\beta})$ with the coefficients $\boldsymbol{\beta}$ common across all dose and drug combinations. Let $\text{logit}(p) = \log\{p/(1-p)\}$. We assume

$$\text{logit}\{p(e_{iklt}=1|e_{\bullet k, \ell, t-1}, \boldsymbol{\beta})\} = \beta_i + \sum_{i \rightarrow j} \beta_{ij} e_{i, k, \ell, t-1} e_{j, k, \ell, t} \quad (1)$$

and $\text{logit}\{p(e_{iklt}=1|\boldsymbol{\beta})\} = \beta_i$ at $t=1$. The model is indexed by auto-regressive coefficients $\boldsymbol{\beta}$. The key feature is that $\boldsymbol{\beta}$ is common across all experiments. We assume the same dependence structure of the latent indicators e_{iklt} across all experiments. In Equation (1), a non-zero coefficient β_{ij} implies that the activation of protein i at time $t-1$ impacts the activation of protein j at time t . It is convenient to represent the set of zero coefficients β_{ij} as a graphical model. Let $G = \{V, E\}$ denote a graph with nodes $V = \{1, \dots, m\}$ and edges $E \subset V \times V$ with $\{i, j\} \in E$ if and only if $\beta_{ij} \neq 0$. In the next section, we will complete the model with a prior $p(G)$ on the common dependence structure.

We still need to link the latent indicators e_{iklt} with the observed protein activation y_{iklt} . This defines the top level sampling model. To ensure a meaningful interpretation of e_{iklt} as indicators for activation we specify the sampling distribution for the observed expression y_{iklt} as a mixture of normals. The mixture distribution formalizes the interpretation of e_{iklt} as an indicator for protein activation.

$$p(y_{iklt}|e_{iklt}) \propto \begin{cases} N(\mu_{1i}, \sigma_{1i}^2), & e_{iklt}=0. \\ N(\mu_{2i}, \sigma_{2i}^2), & e_{iklt}=1. \end{cases} \quad (2)$$

In the assumed sampling model (2) we use normal residual distributions, mainly for computational convenience. We carried out a test-of-fit to validate the assumption. We used a Bayesian chi-square test of fit [9], and found no substantial evidence against the normal assumption. While there is scope for generalization, the normal assumption is acceptable as a simplifying model choice. We write $\theta_i = (\mu_{1i}, \mu_{2i}, \sigma_{1i}, \sigma_{2i})$ for the parameters that index the sampling model for the i th protein and $\boldsymbol{\theta} = (\theta_i, i = 1, \dots, m)$ for the combined parameter vector for the sampling model.

4. A network model for G

We complete the model construction with a prior probability model for the common dependence structure, common across all drug/dose combinations. Recall that $G = \{V, E\}$ summarizes the selection of covariates in the auto-regressive model for the binary vector $e_{\bullet k, \ell, t}$ across t . Inference on G addresses the central inference question of describing the dependence of a set of protein marker expressions at time t on the expressions at time $t-1$. The presence of a directed edge in G from node i to j implies that the expression of protein i at time $t-1$ impacts the expression of protein j at time t . Note that with this interpretation, the inclusion of edges (i, i) , is meaningful (and even likely). We assume that this graph G remains the same across all drugs and dose levels. This is the key assumption that will enable us to borrow strength across the related experiments.

In biomedical experiments it is now increasingly common to encounter graphs with large numbers of nodes, representing proteins, genes and other entities. For example, when we consider the full set of proteins in the PI3k pathway, the number of nodes is 66 and the

number of all possible edges is 66^2 , as the graph allows edges in both directions as well as self-cycles. Under Bayesian inference, graphs could be treated as random variables and characterized by prior distributions. Some of the popular Bayesian approaches to inference on graphs include Gaussian graphical models [3,4,10], Bayesian networks [6,8] and directed acyclical graphs [11]. In general, the choice of the prior model plays an important role. For small sample sizes, models using a vague prior for a high-dimensional graph can lead to identifiability problems in the posterior model, which in turn typically result in slow mixing Markov chain Monte Carlo (MCMC) simulations. Frequentist inference tackles this problem by imposing an additional sparsity constraint, for example, in the optimization of the likelihood function. Such constraints are modeled through Lasso and penalized likelihood methods [18,21].

Alternatively, a similar penalty can be introduced in a Bayesian approach with a prior that favors sparsity. For example, in a model that assumes that edges are a priori independent with an edge inclusion probability π , a prior favoring sparse networks is constructed by a suitable hyperprior on π . Let $G_{ij} = I\{(i, j) \in E\}$ denote an indicator for the graph including an edge from i to j . We assume

$$\begin{aligned} p(G_{ij}=1) &= \pi, \\ \pi &\sim \text{Beta}(a, b). \end{aligned} \quad (3)$$

The hyper-parameter π represents the probability of edge inclusion.

The desired sparsity of the network can be calibrated by tuning the hyper-parameters a and b . We set $a = 2$, $b = 10$ for inference.

Sparsity priors mitigate the curse of dimensionality in graphical inference. An alternative approach to tackle inference in large graphs is the use of informative priors that are based on some consensus network. This network is used as a centering graph. In simulation studies (not shown) we find that these models can recover the truth adequately well if the centering graph is close to the simulation truth. When sufficient prior information based on biological expertise or related data is available we can construct informative priors and consider inference for a larger number of vertices. Let $G_0 = (V, E_0)$ denote an informed prior guess of the dependence structure, based on expert opinion. We construct the graphical prior $p(G)$ based on the number of changes relative to G_0 , $d(G, G_0) = |E \cap E_0^c| + |E^c \cap E_0|$, using a geometric distribution on d . That is, we assume

$$p(G) \propto \rho^{d(G, G_0)}. \quad (4)$$

We shall refer to ρ as the concentration parameter. The prior $p(G)$ implies that we place less weight on graphs which are more distant from G_0 . The weights reduce exponentially with increasing values of d . Since $p(G)/p(G_0) = \rho^{d(G, G_0)}$ is linked to easily interpreted prior odds, it is possible to calibrate ρ to reflect the prior confidence on a consensus network G_0 .

5. Model summary and hyperpriors

In summary, the two key elements of the proposed model are: (1) the dependence network G is common to all dose levels, and, (2) we model dependence at the level of latent binary indicators e_{iklt} of protein activation, rather than the raw expression measurements y_{iklt} . The overall model structure is summarized in Figure 1.

For reference we state the full joint model,

$$p(\mathbf{e}, \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}, G) \propto p(\mathbf{e} | \boldsymbol{\beta}, G) p(\mathbf{y} | \mathbf{e}, \boldsymbol{\theta}) p(\boldsymbol{\beta} | G) p(\boldsymbol{\theta}) p(G | \pi) p(\pi). \quad (5)$$

The last two factors $p(G | \pi) p(\pi)$ are the prior model (3). The first factor is the product over t of the conditional distributions $p(e_{iklt} | e_{i,k,\ell,t-1})$. That is,

$$p(\mathbf{e} | \boldsymbol{\beta}, G) = \prod_{i,k,\ell} p(e_{ik\ell 1}) \prod_{t=2}^D p(e_{iklt} | e_{i,k,\ell,t-1}),$$

with the factors in the product being logistic probabilities of Equation (1). The factor $p(\boldsymbol{\beta} | G)$ in Equation (5) is the prior on the coefficients in the logistic model. We assume independent normal priors $p(\beta_{ij} | G) \propto N(0, 50)$ and $\beta_i \propto N(0, 50)$, essentially a flat prior on $\boldsymbol{\beta}$. Finally, we specify the priors for the parameters in $\boldsymbol{\theta}$ as $\mu_{ik} \sim N(0, \tau)$, $k = 1, 2$ and $\tau = 100$. For efficient MCMC performance, we fix σ_{ik} to $s_e(i)/2$ where $s_e(i)$ is the empirical standard deviation of the expressions for protein i .

These vague prior distributions for the sampling model parameters are chosen to minimize the impact of prior information on the desired inference on G . We found posterior inference on G to be robust with respect to the particular choices.

We set the hyper-parameters of the graphical prior to $a = 2$ and $b = 10$. This represents a prior judgment on the sparsity of the graph with $E(\pi) = 0.16$ and $SD(\pi) = 0.1$. Posterior inference about the most important edges in the graph is not sensitive to this choice. The sum, $a + b$, can be considered a prior equivalent sample size (in consideration of the nature of posterior updating). Here, the relevant sample size is the number of edges in the graph. The chosen $a + b = 12$ is small compared with the number of edges even in the moderate size graph in this application.

6. Results

We carry out inference under model (5). We set the hyper-parameters of the graphical prior to $a = 2$ and $b = 10$. Inference proceeds by MCMC posterior simulations. We use $[x | y, z]$ to generically indicate a transition probability that modifies x and is indexed by currently imputed values of y and z . MCMC posterior simulation proceeds by iterating over the following transition probabilities:

$$[e_t | e_{t-1}, G, \beta], [\beta, \alpha | e, G], [\theta | e, y], [G | \beta, e].$$

The full conditional posterior of e_t given e_{t-1} and the other parameters is straightforward as a set of m (conditionally) independent binary random variables. The update of e_t includes a loop over k and ℓ , and is repeated for $t = 1, \dots, T$. The sampling model parameters θ are updated by draws from the complete conditional posterior distributions. The coefficients β_{ij} and β_i are updated by a Metropolis–Hastings transition probabilities. For example, for β_{ij} we use a random walk proposal distribution $\beta_{ij} \sim N(\beta_{ij}, .4)$. Finally, updating G requires trans-dimensional MCMC. Adding an edge introduces an additional coefficient β_{ij} , thus incrementing the dimension of the parameter vector. We implement a reversible jump transition probability with birth and death steps to add new edges or delete current edges.

The posterior estimates are computed using MCMC posterior simulation with a total of 25,000 iterations and an initial burn-in of 13,000 iterations. We use the 12,000 post burn-in MCMC samples to evaluate posterior inclusion probability $\hat{P}_{ij} = p(G_{ij} = 1 | y)$ for each possible edge (i, j) in the graph. We then report an estimate of graph G by including all edges with $\hat{P}_{ij} > c$ where the threshold c is chosen to achieve a posterior expected false discovery rate close to 0.05. Figure 2 shows the estimated posterior network for the entire set of proteins. Table 1 lists the name of proteins. The numbers on the left column correspond to the node labels in Figure 2.

We define connectivity c_i of a protein node i as the number of edges that connect to it.

Posterior summaries of connectivities are useful indicators of the importance of the node in the pathway. TSC2 pT1462 ($i = 22$ in the figure), AKT ($i = 3$) ER ($i = 64$), p70S6K ($i = 10$), and STAT3 ($i = 42$) are some proteins with high posterior mean connectivities. Of these, TSC2 pT1462 has the highest mean connectivity $E(c_{22} | y) = 10$, with $p(c_{22} = c | y) = 0.45$, 0.52 and 0.03 for $c = 9, 10, 11$.

The estimated posterior network in Figure 2 infers many genomic interactions validated by recently biological research. First, AKT, a serine threonine-specific protein kinase, is one of the major targets of PI3K and plays a key role in the PI3K signaling pathway. It regulates TSC2 and p70S6K in our estimated network. The positive regulation of AKT on phosphorylated p70S6K is supported by an experiment showing constitutive activation of AKT increases phosphorylation of p70S6K protein [17]. The experimental result that constitutive activation of AKT significantly induces the degradation of TSC1/TSC2 protein complex in a dose-dependent manner validates the negative regulation of AKT on TSC2 predicted by our posterior network [2]. The positive regulation of SRC on AKT is also consistent with the fact that AKT is transcriptionally up-regulated by SRC ($i = 46$) through a cascade of molecular interactions [15]. Second, STAT3 is a highly connected node in the estimated network. Its activation is important for insulin/IGF1-induced anchorage-independent growth and survival of certain ovarian cancer cells. The predicted positive regulation of IGFR on phosphorylated STAT3 protein is consistent with the fact that IGFR mediates STAT3 activation and its target gene expression [22]. Another positive regulation on STAT3 inferred by the network is from c-kit (protein 47 in Table 1), which has been

found mutant in a variety of malignancies and induces STAT3 activation [14]. Third, estrogen receptors (ER, $i = 63$) are specific receptors for estrogen hormone that is a suspected etiologic factor causing ovarian carcinoma. The positive regulation of ER on phosphorylated MAPK in the estimated network is biologically plausible because ER protein is necessary for the activation of MAPK signaling pathway induced by progesterone receptor [12]. Finally, the protein kinase C (PKC) family including PKCa ($i = 59$) has been shown to phosphorylate GSK3, a downstream target in the PI3K signaling pathway, which supports the positive regulation of PKCa on phosphorylated GSK3 in our estimated network [5].

Finally, we conducted a set of sensitivity analysis, varying the values of the sparsity hyper-parameters. The network topologies (not shown) vary. However, a subset of proteins appears in the list of nodes with top degree, consistently across different choices of hyper-parameters. Examples of these are the proteins 22 and 63. In summary, for a small number of observations, inference is sensitive to choice of the hyper-parameters, as one would expect. However, posterior inference has sufficient power to identify the proteins that play a prominent role in the pathway.

As an alternative strategy, we performed an analysis using the informative prior (4) centered on a consensus network obtained from our collaborators. The motivation for the alternative inference were concerns that MCMC posterior simulation under the sparsity prior could suffer from slow mixing Markov chain simulation over the large and complex space of possible models G under this prior. As a centering network G_0 that appears in Equation (4) we used a network using a subset of 12 proteins. This network was based on expert opinion of our collaborators. Figure 3 shows the estimated posterior network based under this informative prior. We used $\rho = 0.15$

7. Conclusion

Advances in functional proteomic technologies have paved the way to exploring protein functions and their interactions with therapeutic agents in molecular pathways. In addition to studying the marginal expressions many recent approaches also focus on discovering large protein–protein interaction networks. Network modeling has become an important tool to study these interactions, and hence necessary for developing novel drugs and therapies. We presented a Bayesian graphical model-based approach that infers signaling pathways from time-course experiments and discover interactions between proteins at consecutive time points. The novelty in our approach arises from (a) using a sparse prior on graphs to accommodate large networks and (b) a flexible model that borrows strength across different dose levels. It is sometimes reasonable to assume that the network describing the dependence between the time points t and $t - 1$ should itself evolve with time. Stochastic dynamic network models could be considered in such cases. But these approaches could possibly lead to over-fitting in the current application with only 8 time points per protein. However, the proposed model has an easy extension to time-dependent networks in studies with large sample sizes. We show an application of the method to data measuring abundance of phosphorylated proteins in a human ovarian cell line, generated from RPPA. Our results

are validated by several well-known biological connections between proteins in ovarian cell lines.

References

1. Bender C, Henjes F, Fröhlich H, Wiemann S, Korf U, Beißbarth T. Dynamic deterministic effects propagation networks: Learning signalling pathways from longitudinal protein array data. *Bioinformatics*. 2010; 26:i596–i602. [PubMed: 20823327]
2. Dan HC, Sun M, Yang L, Feldman RI, Sui XM, Ou CC, Nellist M, Yeung RS, Halley DJ, Nicosia SV, Pledger WJ, Cheng JQ. Phosphatidylinositol 3-kinase/akt pathway regulates tuberous sclerosis tumor suppressor complex by phosphorylation of tuberlin. *J Biol Chem*. 2002; 277:35364–35370. [PubMed: 12167664]
3. Daniels M, Pourahmadi M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*. 2002; 89:553–566.
4. Dobra A, Hans C, Jones B, Nevins JR, West M. Sparse graphical models for exploring gene expression data. *J Multivariate Anal*. 2004; 90:196–212.
5. Fang X, Yu S, Tanyi JL, Lu Y, Woodgett JR, Mills GB. Convergence of multiple signaling cascades at glycogen synthase kinase 3: Edg receptor-mediated phosphorylation and inactivation by lysophosphatidic acid through a protein kinase c-dependent intracellular pathway. *Mol Cell Biol*. 2002; 22:2099–2110. [PubMed: 11884598]
6. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol*. 2000; 7:601–620. [PubMed: 11108481]
7. Inoue L, Neira M, Nelson C, Gleave M, Etzioni R. Cluster-based network model for time-course gene expression data. *Biostatistics*. 2007; 8:507–525. [PubMed: 16980695]
8. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan N, Chung S, Emili A, Snyder M, Greenblatt J, Gerstein M. A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science*. 2003; 302:449–453. [PubMed: 14564010]
9. Johnson VE. A Bayesian χ^2 test for goodness-of-fit. *Ann Statist*. 2004; 32:2361–2384.
10. Jones B, Carvalho C, Dobra A, Hans C, Carter C, West M. Experiments in stochastic computation for high-dimensional graphical models. *Statist Sci*. 2004; 20:388–400.
11. Lauritzen S, Spiegelhalter D. Local computations with probabilities on graphical structures and their application to expert systems. *J R Statist Soc Ser B (Methodological)*. 1988; 50:157–224.
12. Migliaccio A, Piccolo D, Castoria G, Di Domenico M, Bilancio A, Lombardi M, Gong W, Beato M, Auricchio F. Activation of the src/p21ras/erk pathway by progesterone receptor via cross-talk with estrogen receptor. *EMBO J*. 1998; 17:2008–2018. [PubMed: 9524123]
13. Nieto-Barajas L, Müller P, Ji Y, Lu Y, Mills G. A time-series ddp for functional proteomics profiles. *Biometrics*. 2012; 68:859–868. [PubMed: 22221181]
14. Ning ZQ, Li J, McGuinness M, Arcenci RJ. Stat3 activation is required for asp (816) mutant c-kit induced tumorigenicity. *Oncogene*. 2001; 20:4528–4536. [PubMed: 11494148]
15. Park S, Kim D, Kaneko S, Szewczyk KM, Nicosia SV, Yu H, Jove R, Cheng JQ. Molecular cloning and characterization of the human akt1 promoter uncovers its up-regulation by the src/stat3 pathway. *J Biol Chem*. 2005; 280:38932–38941. [PubMed: 16174774]
16. Parmigiani G, Garrett S, Anbazhagan R, Gabrielson E. A statistical framework for expression-based molecular classification in cancer. *J R Statist Soc Ser B*. 2002; 64:717–736.
17. Qiu Q, Yang M, Tsang BK, Gruslin A. Both mitogen-activated protein kinase and phosphatidylinositol 3-kinase signalling are required in epidermal growth factor-induced human trophoblast migration. *Mol Human Reprod*. 2004; 10:677–684.
18. Rothman A, Bickel P, Levina E, Zhu J. Sparse permutation invariant covariance estimation. *Electron J Stat*. 2008; 2:494–515.
19. Telesca D, Inoue L, Neira M, Etzioni R, Gleave M, Nelson C. Differential expression and network inferences through functional data modeling. *Biometrics*. 2009; 65:793–804. [PubMed: 19053995]
20. Tibes R, Qiu Y, Lu Y, Hennessy B, Andreoff M, Mills GB, Kornblau SM. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia

- specimens and hematopoietic stem cell. *Mol Cancer Ther.* 2006; 5:2512–2521. [PubMed: 17041095]
21. Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model. *Biometrika.* 2007; 94:19–35.
22. Zhang W, Zong CS, Hermanto U, Lopez-Bergami P, Ronai Z, Wang LH. Rack1 recruits stat3 specifically to insulin and insulin-like growth factor 1 receptors for activation, which is important for regulating anchorage-independent growth. *Mol Cell Biol.* 2006; 26:413–424. [PubMed: 16382134]

Figure 1.

Model structure. The autoregressive dependence structure of the (latent) indicators e_{iklt} for protein activation is determined by the graph G . Conditional on e_{iklt} we assume a model for the observed protein expression measurements y_{iklt} . The model uses additional parameters β_{ij} to index the prior model for e_{iklt} , and θ_i for the sampling model.

Figure 2.

Posterior estimated network for interactions of proteins in the Pi3k inhibitor pathway. The solid (blue) and dashed (red) edges denote positive and negative associations, respectively. An arrow from node i to node j indicates that protein i affects protein j at the subsequent time point. Please see Table 1 for the protein names.

Figure 3.

The estimated posterior graph based on the informative prior. The prior is centered at a consensus network of 12 proteins provided by the collaborators. The posterior graph shows a less spread than the estimated graph under the non-informative network prior, in terms of the number of connected nodes.

Table 1

Index of protein names.

	Protein	Protein	Protein
1	AKT pS473	23 TSC2	45 Src pY527
2	AKT pT308	24 HER2 pY1428	46 Src
3	AKT	25 HER2	47 c-kit
4	GSK3 pS21/9	26 IGF1R pY1135	48 p21
5	GSK3	27 IGF1R-1	49 p27
6	FOXO3a	28 IGF1R	50 p53
7	MTOR pS2448	29 EGFR	51 Cyclin D1
8	MTOR	30 MEK1-2 pS217/221	52 Cyclin E1
9	p70S6K pT389	31 MEK1-2	53 Cyclin B1
10	p70S6K	32 MEK1/2	54 Rb pS807/811
11	S6 pS240/244	33 MAPK pT202Y204	55 Rb
12	S6 pS235/236	34 ERK2	56 PCNA
13	S6	35 P38 pT180Y182	57 PARP
14	p110a	36 P38	58 PKCa pS657
15	p85	37 JNK pT183Y185	59 PKCa
16	PTEN	38 JNK	60 b-Catenin
17	LKB1	39 cJUN pS73	61 E-Cadherin
18	ACC pS79	40 cJUN	62 Cavinolin
19	ACC	41 STAT3 pT727	63 ER
20	AMPK pT172	42 STAT3 pY705	64 PR
21	AMPK	43 STAT3	65 PAI-1
22	TSC2 pT1462	44 STAT5 pY694	66 Rab25