A Time-Series DDP for Functional Proteomics Profiles

Luis Nieto-Barajas[1,2], Peter Müller[2], Yuan Ji[2],
Yiling Lu[3] and Gordon Mills[3]

[1]*Department of Statistics, ITAM, Mexico*

[2]*Department of Biostatistics,*

[3] *Department of Systems Biology,*

The University of Texas M. D. Anderson Cancer Center, USA

Abstract

Using a new type of array technology, the reverse phase protein array (RPPA), we measure time-course protein expression for a set of selected markers that are known to co-regulate biological functions in a pathway structure. To accommodate the complex dependent nature of the data, including temporal correlation and pathway dependence for the protein markers, we propose a mixed effects model with temporal and protein-specific components. We develop a sequence of random probability measures (RPM) to account for the dependence in time of the protein expression measurements. Marginally, for each RPM we assume a Dirichlet process (DP) model. The dependence is introduced by defining multivariate beta distributions for the unnormalized weights of the stick breaking representation. We also acknowledge the pathway dependence among proteins via a conditionally autoregressive (CAR) model. Applying our model to the RPPA data, we reveal a pathway-dependent functional profile for the set of proteins as well as marginal expression profiles over time for individual markers.

Key words: Bayesian nonparametrics, dependent random measures, Markov beta process, mixed effects model, stick breaking processes, time series analysis.

1. Introduction

1.1. Functional Proteomics

Functional proteomics is a systemic biological tool concerned with probing the protein functions and their interaction with therapeutic regimens. Instead of studying individual protein expression, system biologists now try to decode the behavior of an entire protein pathway, including protein-protein interaction. Typically, a form of stimulus (such as a drug and/or small interfering RNAs) is given to the biological system. The system is then observed over time and the expression of all the proteins in disease-related pathways is measured. Such experiments allow biologists and clinicians to directly investigate causal relationships between the stimulus and the protein markers as well as the pathway dependence among the markers.

We analyze data from a functional proteomics profiling experiment using reverse phase protein arrays (RPPA) (Tibes et al., 2006). RPPA is a high-throughput functional proteomic technology that quantifies the expression for targeted proteins. Researchers can print on each array more than 1,000 individual patient samples containing the whole cellular protein repertoire in batches. On the slides, each sample can be recognized as a square matrix of dots, which represent a dilution series used for quantification of protein expression. Each slide is probed with an antibody that represents one specific protein. Researchers choose specific pathways they want to target, usually containing 50-200 proteins, and produce the same number of arrays, with each array hybridized against one protein. Because of the reversed design (as compared to the microarray), RPPA's allow much larger sample sizes than the traditional microarrays and therefore allow for higher statistical power in detecting protein interactions.

The RPPA data presented here are time-course data for proteins on a signaling pathway over time after an initial intervention. We characterize the effect of the intervention by inference on the changing distribution of protein activation over time. The changing distribution

summarizes the treatment effect on the entire pathway. Taking a Bayesian perspective, the desired inference requires a prior probability model for this sequence of random distributions. Probability models for random distributions are traditionally known as nonparametric Bayes models. See, for example, Hjort et al. (2010) for a recent review. One of the most popular nonparametric Bayesian models is the Dirichlet process (DP) model (Ferguson, 1973). Extending the DP model we define a non-parametric Bayesian model for a time series of random probability measures. Additional protein-specific random effects account for dependence of proteins on the same pathway.

1.2. An RPPA study

We recently performed pathway inhibition experiments to study the therapeutic effects of FDA-approved drugs on ovarian cancer cell lines. Focusing on disease markers in the EGFR/PI3K pathway (Figure 1), we aim to investigate how Lapatinib, an EGF (epidermal growth factor) inhibitor, impacts cellular signaling networks. We formalize the research question as inference on the distribution on protein activation over time for the proteins on the EGFR/PI3K pathway. In other words, we are interested in the shift of the distribution over time.

In the experiment, an ovarian cell line was initially treated with Lapatinib, and then stimulated with EGF over time. At $D = 8$ time points, $t_d = 0, 5, 15, 30, 60, 90, 120$, and 240 minutes, about 30 proteins were probed using RPPA. For each protein, three replicate expression intensities were recorded. We normalized the data such that each protein has a median expression level of 1,000. We then log-transformed the data and computed difference scores as defined for the significance analysis of microarrays method (Tusher et al., 2001). The difference scores are between post-treatment and pre-treatment intensities for each protein at each time point. This pre-processing of the data is necessary in order to make expressions comparable across proteins. Figures 2 and 3 present histograms and time

series of these difference scores for the 30 proteins at each of the 8 time points, respectively. See also Figures B and C in the Supplementary Material for individual plots of each protein.

We first undertake a frequentist analysis of the data using the Bioconductor package Linear Models for Microarray Data (LIMMA). Specifically, we propose a linear model with two covariates, protein as a factor with 30 levels (indexed by i), and a partition of the eight times points into four groups (denoted by c_{d^*}, $d^* = 1, 2, 3, 4$). In particular, $c_{1^*} = \{0\}$, $c_{2^*} = \{5, 15, 30\}$, $c_{3^*} = \{60, 90, 120\}$, and $c_{4^*} = \{240\}$. We set the base line time point 0 as the first group. Then whenever the time is doubled, we define a new group. Denote y_{di} the difference score of protein i at time d. A negative value of y_{di} implies suppression of protein expression after initial treatment of Lapatinib. Let $C = \{C_d,\ d = 1, \ldots, D\} = \{0, 1, 1, 1, 2, 2, 2, 3\}$ be the group label under the above partition. For each protein, LIMMA fits a linear model of the form,

$$y_{di} = \gamma + \beta_{i,C_d} + \epsilon_{di}.$$

We then assess the contrasts that compare the differences between $\beta_{i,0}$ (the fitted base-line mean at time 0) versus other β_{i,C_d} in which $C_d \neq 0$. This allows us to see if the protein expression changed over time groups. We apply the empirical Bayes shrinkage estimation to calculate moderated t-statistics (Smyth, 2004) and adjust the p-values based on the false discovery rate (FDR) approach in Benjamini and Hochberg (1995). These are all standard procedures in the LIMMA package.

We find that *four* proteins (p.p70S6K.T389, p.JNK.T183.Y185, PKCa, and p.PKC.S567) are differentially expressed between time group 1 and time group 0, *two* proteins (p.p70S6K.T389, pS6.S235.236) between time group 2 and time group 0, and only *one* protein (p.p70S6K.T389) between time group 3 and time group 0. This implies that as time elapses, fewer and fewer proteins are differentially expressed. Protein p.p70S6K.T389 is differentially expressed in all three comparisons, suggesting a consistent suppression over time. This is indeed the case shown in the data. Surprisingly, protein pS6.S235.236 is only iden-

tified as differentially expressed between time group 2 and time group 0, but not in other comparisons. Staring at the row data (Figure 3), this protein corresponds to the one with the largest differences in its expression at almost all the time points relative to time point 0. We need a more detailed modeling and decided to pursue a fully model-based Bayesian inference. As Section 5.3 shows, the alternative approach discovers interesting and plausible new findings that echo biological belief.

1.3. Random effects model

The Bayesian analysis aims to achieve three objectives, (i) to characterize the time-varying distribution of protein activation over time; (ii) to identify proteins that are suppressed due to deactivation of the pathway by Lapatinib; and (iii) to account for the protein–protein interaction in the pathway. To this end we propose the following random effects model. Let $F_d,\ d = 1, \ldots, D$ denote a time-varying distribution of protein activations. Let u_i denote protein-specific random effects. We use u_i to account for (known) dependence of proteins in the pathway. Let y_{di} denote the observed difference in expression for protein i at time d. We assume that y_{di} arises from the model

$$y_{di} = x_{di} + u_i + \epsilon_{di} \tag{1}$$

with independent residuals $\epsilon_{di} \sim N(0, \tau_d)$ and $x_{di} \sim F_d$, independently. The distributions $\{F_d,\ d = 1, \ldots, D\}$ characterize the time-varying activation of the pathway after adjusting for protein-specific effects.

The experiment includes data on some control proteins. These are proteins that are not expected to respond to Lapatinib treatment and are included for verification purposes. Due to the inclusion of these control proteins, we expect multi-modality in the distributions F_d. This is confirmed by Figure 2. Beyond this, it is difficult to predict specific features of the time-varying distributions. These considerations rule out the use of standard parametric models. Instead, we introduce a variation of the nonparametric dependent Dirichlet process

model, namely the time-series dependent Dirichlet process (tsDDP), to capture the temporal effects of Lapatinib on the entire protein pathway. For the protein-specific random effects u_i we use a conditionally autoregressive (CAR) model to account for the pathway dependence among proteins.

The contents of the rest of the paper is described as follows. In Section 2 we introduce a novel dependent Dirichlet process to model the temporal effects. Pathway dependence is modeled via a conditionally autoregressive model which is described in Section 3. In Section 4 we deal with implementation issues and derive an efficient algorithm for carrying out posterior inference. Results from the analysis of the RPPA data are presented in Section 5 and the paper concludes with some remarks in Section 6.

2. Time-Varying Protein-Activation

2.1 Dependent DP Models

One of the inference goals is modeling of the time-varying distributions, $\mathcal{F} = \{F_d,\ d = 1, \ldots, D\}$, of protein-activation. We introduce a novel model for a time series of random probability measures (RPM's), $p(\mathcal{F})$. In particular, we propose dependent RPMs with Dirichlet process (DP) marginal distributions (Ferguson, 1973). We define the joint model on the sequence of RPMs $\mathcal{F}$ by assuming discrete RPMs with common locations of point masses and varying, but dependent, weights. This construction allows us to specify different strengths of dependence in different parts of the sample space. This feature becomes relevant when, for example, two RPMs are judged to be similar in the center of the distributions, but outliers are likely to be quite different, as is the case for the RPPA data.

The DP prior (Ferguson, 1973) is indexed by two parameters, a total mass parameter c and base measure G. We write $F \sim \mathcal{DP}(c, G)$. The base measure defines the expectation, i.e., $\mathrm{E}(F) = G$ and the total mass parameter can be interpreted as a precision parameter;

the larger c, the closer the random F is to G.

MacEachern (1999) introduced a general idea to define dependent DP (DDP) as a way of extending the DP model to multiple random measures $\mathcal{F} = \{F_x,\ x \in X\}$. Since the DP generates almost surely (a.s.) discrete RPM, we can write $F_x(\cdot) = \sum_h w_{xh}\, \delta_{\mu_{xh}}(\cdot)$, where δ_m denotes a point mass at m and w_{xh} and μ_{xh} are random quantities. The key idea behind the DDP is to introduce dependence across the measures F_x by assuming that the distribution of the point masses μ_{xh} are dependent across different levels of x but still independent across h. Alternatively the dependence across x can be introduced on the weights w_{xh}, keeping the model unchanged marginally, for fixed x. There are several ways to achieve this. For instance, in the basic version of the DDP the weights are assumed constant across x, i.e., $w_{xh} = w_h$, and the dependence of μ_{xh}'s across x could be obtained by using a Gaussian process (MacEachern, 1999). Alternatively, DeIorio et al. (2004) indexed F_x with categorical factors x to define an analysis of variance type dependence of the μ_{xh} across x. Moreover, Gelfand et al. (2005) worked with spatial data and defined dependence through a spatial Gaussian process. In a time series setting, taking x to be time, Caron et al. (2007) and Rodriguez and Ter Horst (2008) considered dynamic DP models by introducing first order dependence in the μ_{xh} via a generalized Polya urns scheme and a dynamic linear model respectively. Recently, Barrientos et al. (2011) have shown that, however the dependence is introduced in the model, weights and/or locations, the DDP has full weak support.

Our proposed model is a special case of the general DDP model introduced in MacEachern (1999). Most definitions of DDP models, including all examples above except for Griffin and Steel (2006), introduce the desired dependence on the locations μ_{xh}. In contrast, we proceed by introducing the dependence through the weights. We believe dependence on the weights and common locations to be more natural for time series analysis than common weights and dependent locations. From a data analysis point of view, the former corresponds to discretizing the sample space, with a common discretization over the sequence of random

measures. The latter can be interpreted as a discretization on the probability scale.

In two other independent works, Taddy (2010) and Griffin and Steel (2011) proposed stick breaking autoregressive models, which include the DP as special case. In particular, Griffin and Steel (2011) construct correlated RPMs F_d indexed by continuous time d. They start with a stick breaking construction at time $d = 0$. As time progresses additional point masses are added, in a way that leaves the marginal distribution of F_d unchanged and introduces the desired correlation. The arrival times when new point masses are added are latent. This is in contrast to the proposed model that will use common locations and time varying weights only. Taddy (2010) uses a construction that is closer related to our proposed process. He works with stick-breaking RPMs and introduces the desired dependence via a beta autoregression on the fractions of the stick breaking constructions by means of two sets of latent variables. The construction implicitly assumes equally spaced time points. On the other hand, we will use a latent binomial process to induce the desired correlation and we can easily accommodate unequal time points.

2.2 Time-Series DDP

We define a probability model for a sequence of dependent measures $\{F_1, F_2, \ldots, F_D\}$. Recall Sethuraman's (1994) stick breaking representation for an RPM

$$F_d = \sum_{h=1}^{\infty} w_{dh}\, \delta_{\mu_{dh}} \tag{2}$$

for $d = 1, \ldots, D$. The random cumulative distribution function has steps of size w_{dh} at locations μ_{dh}. Let $\text{Be}(a, b)$ denote a beta distribution with mean $a/(a + b)$. Sethuraman showed that a constructive definition of $F_d \sim \mathcal{DP}(c, G)$ is achieved by taking $\mu_{dh} \overset{\text{iid}}{\sim} G$ and $w_{dh} = v_{dh} \prod_{j<h}(1 - v_{dj})$ for independent $v_{dh} \overset{\text{iid}}{\sim} \text{Be}(1, c)$.

We introduce dependence between (F_d, F_{d+1}) by the following construction. We assume common locations $\mu_{dh} = \mu_h$ for all $d = 1, \ldots, D$, and introduce the desired dependence through the weights. We introduce latent binomial random variables (r.v.) that link v_{dh} and

$v_{d+1,h}$, and thus $(w_{dh}, w_{d+1,h})$.

Specifically, the joint distribution $p(F_1, \ldots, F_D)$ is defined by considering the stick breaking representation (2) for each F_d, $d = 1, \ldots, D$, with common locations

$$\mu_{dh} = \mu_h \overset{\text{iid}}{\sim} G,$$

together with a joint distribution for $(v_{1h}, \ldots, v_{Dh})$ assumed to be a Markov process similar to the one considered by Nieto-Barajas and Walker (2002). It is defined conditional on latent variables z_{dh}:

$$\begin{aligned} v_{1h} &\sim \text{Be}(1, c), \\ z_{dh} \mid v_{dh} \sim \text{Bin}(m_{dh}, v_{dh}), &\quad v_{d+1,h} \mid z_{dh} \sim \text{Be}(1 + z_{dh}, c + m_{dh} - z_{dh}) \end{aligned} \tag{3}$$

for $d = 1, \ldots, D-1$.

We refer to the joint model for $(F_1, \ldots, F_d)$ as *time series DDP*. We write $(F_1, \ldots, F_d) \sim ts\mathcal{DDP}(c, G, \mathbf{m})$, where $\mathbf{m} = \{m_{dh}\}$ for $d = 1, \ldots, D$ and $h = 1, 2 \ldots$ Let $\mathbf{v} = \{v_{dh}\}$ and $\mathbf{z} = \{z_{dh}\}$. To verify the marginal distribution $p(F_d)$, we note that the Markov process for $(\mathbf{v}, \mathbf{z})$ is fully stationary with $p(v_{dh}) = \text{Be}(1, c)$ as the marginal distribution for v_{dh}, which together with the independent and identically-distributed model on the point masses μ_h implies that $F_d \sim \mathcal{DP}(c, G)$, marginally, for $d = 1, \ldots, D$.

Equations (3) define a joint distribution for v_{dh} and z_{dh}. The role of the latent variable z_{dh} is to introduce dependence between the pair $(v_{dh}, v_{d+1,h})$. The strength of the dependence is specified by m_{dh}. A larger binomial sample size m_{dh} implies higher correlation. In fact, $\text{Corr}(v_{dh}, v_{d+1,h}) = m_{dh}/(1 + c + m_{dh})$. In the limiting case when $m_{dh} \to \infty$, we get equal weights (i.e., $v_{dh} = v_{d+1,h}$ w.p. 1) and, doing so for all d and h, we obtain identical random measures $F_d = F_1$ for $d = 2, \ldots, D$ w.p. 1. However, if $m_{dh} = 0$, v_{dh} and $v_{d+1,h}$ become independent. If we set $m_{dh} = m_d$ for all h we produce the same degree of dependence in the pairs $(v_{dh}, v_{d+1,h})$. In the case of the RPPA data, the observation times t_d are not

equally spaced. We address this by considering $m_{dh} = m_h \Delta_d$ or simply $m_{dh} = m\Delta_d$, where $\Delta_d = 1/(t_{d+1} - t_d)$ and t_d's are the actual times at which the y_{di}'s were observed. If the elapsed time from distributions d and $d+1$ is short, then Δ_d is large and implies a strong dependence, whereas a longer elapsed time between distributions implies a weaker dependence.

One way of assessing the dependence structure induced by our model is by considering the correlation between the probability masses of distributions F_d and F_{d+1}. This correlation is shown in the following Proposition.

Proposition 1 *Let $A \subset \mathbb{R}$ be measurable. For $(F_1, \ldots, F_D) \sim ts\mathcal{DDP}(c, G, \mathbf{m})$, let $\rho_d(A) \equiv Corr\{F_d(A), F_{d+1}(A)\}$. Then $\rho_d(A) > 0$ and is given by*

$$\rho_d(A) = (1+c)\sum_{h=1}^{\infty} a_{dh} \prod_{i=1}^{h-1} b_{di} + \frac{G(A)}{1-G(A)} \left[\sum_{h=1}^{\infty} \{2 - (1+c)a_{dh}\} \prod_{i=1}^{h-1} b_{di} - (1+c) \right],$$

where for $i = 1, 2, \ldots,$ and $d = 1 \ldots, D-1$,

$$a_{di} = \frac{2(1+m_{di}) + c}{(1+c+m_{di})(1+c)(2+c)} \quad \textit{and} \quad b_{di} = \frac{c-1}{1+c} + a_{di}.$$

See (4) below for a highly simplified expression in an important special case.

The proof is postponed to the Appendix. The correlation between two adjacent RPMs is larger in regions where the prior mean G assigns more probability. That is, the prior model places the strongest dependence in the most probable regions according to G. The expression greatly simplifies when $m_{dh} = m_d$ for all h. It can be shown that the correlation in this case simplifies to

$$\rho_d(A) = \frac{1+c}{1 + 2c(2+c+m_d)/(2+c+2m_d)}. \tag{4}$$

Furthermore, if $m_d = 0$ then

$$\rho_d(A) = (1+c)/(1+2c).$$

Recall that a value of $m_{dh} = 0$ implies independence between $(v_{dh}, v_{d+1,h})$. However, formula (4) indicates that even if v_{dh} is independent of $v_{d+1,h}$ $\forall h$, F_d is still correlated with F_{d+1} due

to the common point masses μ_{dh}. In fact, for the same simplified case when $m_{dh} = m_d = 0$, if $c \to 0$ implies that $\rho_d(A) \to 1$, whereas if $c \to \infty$ implies that $\rho_d(A) \to 1/2$. In other words, if $m_{dh} = 0$ then $\rho_d(A) \in (1/2, 1)$. Note that a moderate to strong dependence in the measures F_d and F_{d+1} does not imply a strong dependence in the observations x_{di} and $x_{d+1,j}$. Consider the extreme case when $m_{dh} \to \infty$, i.e. $F_d = F_{d+1}$ w.p. 1, then $x_i, x_j | F_d \overset{\text{iid}}{\sim} F_d$, with $F_d \sim \mathcal{DP}(c, G)$, implies that $\text{Corr}(x_i, x_j) = 1/(c+1)$ which can cover the whole range in $(0, 1)$. Thus, for finite m_{dh} this correlation should be lower, i.e. $\text{Corr}(x_{di}, x_{d+1,j}) \leq 1/(c+1)$. For most time-course problems one would assume at least moderate correlations in the measures. The restriction to $\rho \geq 1/2$ is therefore reasonable.

Remark 1: The implied correlations $\rho_d(A)$ of the dependent measures in a DDP model highlight a fundamental difference between alternative constructions: (1) *common* weights and *dependent* locations (as in MacEachern (1999)); (2) *common* locations and *dependent* weights, as in the proposed model (3). Consider the extreme cases of the two models: (a) *common* weights with *independent* locations; (b) *common* locations with *independent* unstandardized weights v_{dh}. Model (a) implies that $\rho(A) = 0$. Model (b) implies that $\rho(A) \geq 1/2$. The common locations in the latter model give rise to a natural dependence, even with independent weights.

In summary we assume

$$\begin{aligned} x_{di} \mid F_d &\overset{\text{iid}}{\sim} F_d \\ (F_1, \ldots, F_D) &\sim ts\mathcal{DDP}(c, G, \mathbf{m}). \end{aligned} \tag{5}$$

Since the distributions $\{F_d\}_{d=1}^D$ are dependent, adjacent effects $(x_{di}, x_{d+1,j})$ are also (marginally) dependent after integrating with respect to the RPMs F_d's. Equations (1) and (5) can be viewed as a dynamic state space model (Harrison and Stevens, 1976).

3. Pathway Dependence

The second important component in model (1) are the protein-specific effects u_i's. A dependent prior on $\mathbf{u} = (u_1, \ldots, u_n)$ accounts for dependence of proteins on the common pathway. We assume known dependence structure, summarized in the consensus pathway shown in Figure 1. We represent this (known) dependence as a conditional autoregressive (CAR) model, originally studied by Besag (1974). The general specification of the CAR model for a vector $\mathbf{u}' = (u_1, \ldots, u_n)$ is

$$\mathbf{u} \sim \mathrm{N}\left(\mathbf{0}, D_\lambda(I - \alpha B)\right), \tag{6}$$

where I is the identity matrix, B is an $n \times n$ matrix such that $b_{ii} = 0$, $D_\lambda = \mathrm{diag}(\lambda_i)$ and α is such that $\alpha\omega_i < 1$ for $i = 1, \ldots, n$, with ω_i the eigenvalues of B (Wall, 2004). Different choices of matrices D_λ and B produce different CAR models. In particular, Cressie (1973) suggests to use

$$D_\lambda = \lambda I \;\text{ and }\; B = W,$$

where W denotes the adjacency matrix of the protein interaction, such that $w_{ii} = 0$ and $w_{ii'} = 1$ if proteins i and i' are neighbors ($i \sim i'$). We will use this latter specification of Cressie.

We define the proximity matrix W as follows: Considering the pathway diagram in Figure 1, we first remove the feedback loop from p70 to EGFR to interpret the graph as a directed acyclic graph (Thulasiraman and Swamy, 1992). We then marginalize with respect to the proteins that are not recorded in the data and finally use moralization to define a neighborhood structure for $\mathbf{u}$. Any protein that was recorded in the data, but does not feature on the pathway, is simply included without neighbors.

To ensure propriety of model (6), the association parameter needs to satisfy the condition (Cressie, 1973), $\alpha \in (\gamma_{min}^{-1}, \gamma_{max}^{-1})$, where γ_{min} and γ_{max} are the minimum and maximum eigenvalues of the adjacency matrix W. In our case, the range of values for α is $(-0.4374, 0.1734)$. We will consider a prior distribution for α in its range of values by translating a beta distribution. That is, if we want $\alpha \in (c_1, c_2)$ we will take $\alpha = c_1 + (c_2 - c_1)X$ with $X \sim \mathrm{Be}(a, b)$.

In notation we will say that $\alpha \sim \text{TBe}(a, b, c_1, c_2)$. The model is completed by considering a prior distribution on the precision parameter λ of the form $\lambda \sim \text{Ga}(\alpha_\lambda, \beta_\lambda)$.

4. Posterior inference

An important feature of the proposed model is the easy implementation of posterior MCMC simulation. The key problem is to sample from the posterior distribution of the random distributions F_d. To this end we propose a *collapsed* Gibbs sampler based on marginalizing with respect to the non-allocated point masses in the stick breaking representation of F_d, similar to Papaspiliopoulos and Roberts (2008).

Recall that our model is defined by equations (1), (5) and (6) together with its hyperprior distributions on prior parameters c, $\mathbf{m}$, λ and α. Additionally, a prior distribution is assumed for the residual precision, $\tau_d \sim \text{Ga}(\alpha_\tau, \beta_\tau)$.

For the following discussion it is convenient to rewrite $x_{di}|F_d \sim \sum_h w_{dh}\, \delta_{\mu_h}$ with a latent indicators k_{di}

$$\Pr(k_{di} = h) = w_{dh}, \text{ and } x_{di} = \mu_{k_{di}}.$$

We also introduce notation that will be helpful. Let $\mathbf{k} = \{k_{di},\ i = 1, \ldots, n_d,\ d = 1, \ldots, D\}$ denote the set of latent indicators and $\mathbf{k}^* = \{k_1^*, \ldots, k_s^*\}$ denote the set of unique values in $\mathbf{k}$. Also, let $\boldsymbol{\mu}^* = \{\mu_h,\ h \in \mathbf{k}^*\}$ denote the set of point masses that are linked with the latent variables x_{di}'s. We set $H = \max\{\mathbf{k}^*\}$, i.e., H is the highest index of a point mass μ_h that is sampled as a latent x_{di}. Recall from Proposition 1 that c and m_{dh} control the level of dependence. Let $\Delta_d = 1/(t_{d+1} - t_d)$ and $m_{dh} = m_d = m\Delta_d$. We use the data to set values for the remaining parameters c and m by including them in the unknown parameter vector and completing the model with a hyperprior on (c, m). We assume $c \sim \text{Ga}(\alpha_c, \beta_c)$ and $m \sim \text{TPo}(\xi, \kappa)$ independently a priori, where $\text{TPo}(\xi, \kappa)$ denotes a truncated Poisson distribution with intensity ξ and upper bound κ. Moreover, the centering measure G is

considered to be a $\mathrm{N}(\mu_0, \tau_0)$.

For later reference we denote the complete parameter vector as

$$\boldsymbol{\theta} = (\mathbf{x}, \mathbf{k}, \boldsymbol{\mu}^*, \mathbf{z}, \mathbf{v}, \boldsymbol{\tau}, c, m, \mathbf{u}, \lambda, \alpha)\,.$$

Here $\mathbf{v} = (v_{dh},\ d = 1, \ldots, D,\ h = 1, \ldots, H)$. We only include the weights for point masses 1 through H and locations for allocated point masses, $h \in \mathbf{k}^*$. The remaining parameters, including weights beyond H and non-allocated locations μ_h, $h \notin \mathbf{k}^*$, are not required in the following posterior simulation. In other words, we can analytically marginalize with respect to these parameters, allowing us to define posterior simulation for $p(\boldsymbol{\theta} \mid \text{data})$. Note that $\boldsymbol{\theta}$ includes $s \leq H$ locations $\boldsymbol{\mu}^*$, but exactly H weights v_{dh} (for each d), i.e., locations of allocated point masses only, but weights of *all* point masses with indices up to H.

Posterior inference is implemented via Markov chain Monte Carlo (MCMC) simulation (Tierney, 1994). The posterior conditional distributions can be found in the Appendix in the Supplementary Material. Fortran code for implementing this model can be obtained upon request from the first author.

Ishwaran and James (2001) proposed a general algorithm for carrying out posterior inference in stick breaking processes that is similar to ours in the sense that it works with the posterior distribution of the whole process. Their method is called the blocked Gibbs sampler, which is an approximation to the posterior distribution of the process. Our algorithm involves no approximation and is computationally efficient due to the marginalization wherever possible.

5. Data Analysis

5.1 MCMC implementation and hyperparameters

We carry out inference for the time-course RPPA data described in the Introduction by implementing the random effects model (1), (5) and (6). Recall that $n_d = 30$ proteins were

measured for $D = 8$ time points. The recorded difference scores y_{di} are shown in Figure 3.

We implement MCMC simulation and obtain a Monte Carlo posterior samples of all the parameters in the model, $\boldsymbol{\theta}$, and in particular of the RPM's F_d. The desired inference on the overall effect of the intervention on the entire pathway is summarized by $E(F_d \mid data)$ over time $d = 1, \ldots, D$.

The eight observation times were $0, 5, 15, 30, 60, 90, 120$, and 240 minutes. We use time units of 5 minutes and took $m_{dh} = m\Delta_h$ with $\Delta_1 = 1$, $\Delta_2 = 1/2$, $\Delta_3 = 1/3$, and $\Delta_d = 1/6$ for $d = 4, 5, 6$, and $\Delta_7 = 1/24$. We complete the model specification with $\mu_0 = 0$, $\tau_0 = 0.1$, $\alpha_\tau = \beta_\tau = \alpha_\lambda = \beta_\lambda = 0.1$. For comparison we consider several choices for the hyperpriors on c, m and α. In particular, for the parameter m we take truncated Poisson distributions with upper bound of 5. Remember that the parameter m acts as a binomial sample size in the conditional distribution of the latent $\mathbf{z}$ (see equation (3)), which is turn plays the role of latent observations in the posterior distribution of the unnormalized weights $\mathbf{v}$ (see conditional distribution (c) in Section 3). So, since the actual sample size for each time t_d is relatively small (30 proteins), constraining $\mathbf{z}$ to take small values avoids that prior information swamp the data. On the other hand, according to Carlin and Banerjee (2003), CAR models have better performance for values of α close to the upper limit. We will consider noninformative and informative priors for α to compare.

We ran the Gibbs sampler for 300,000 iterations with a burn-in of 50,000. After burn-in we saved the imputed state every 10th iteration. We assessed model fit using the logarithm of the pseudomarginal likelihood (LPML) statistic (Gelfand et al., 1992), which is a summary of the individual conditional predictive ordinates (CPO). Alternative ways of model comparison rely on the use of Bayes factors. However, according to Basu and Chib (2003), Bayes factors for Dirichlet processes mixture models are technically difficult to compute and instead they suggest to compute CPOs for model comparison.

We considered models with combinations of Ga$(1, 1)$ or Ga$(0.1, 0.1)$ priors on c, and

TPo$(1,5)$ or TPo$(5,5)$ for m and TBe$(1,1,c_1,c_2)$ or TBe$(18,2,c_1,c_2)$ with $c_1 = -0.43$ and $c_2 = 0.17$ on α. We find LPML values (not shown) between 54 and 61 with the highest LPML reported for $c \sim \text{Ga}(1,1)$, $m \sim \text{TPo}(1,5)$, $\alpha \sim \text{TBe}(18,2,c_1,c_2)$. The LPML indicates that the model fit is not very sensitive to the hyper-prior choices of c and m, however it is sensitive to the prior choice of α. We find better fit under an informative prior for α, TBe$(18,2,c_1,c_2)$, versus a flat prior TBe$(1,1,c_1,c_2)$. This confirms Carlin and Banerjee (2003) who observed that values of α closer to the upper limit better capture the dependence and result in a better fit.

5.2 Alternative models

Before we proceed with data analysis, we compare the proposed model versus other alternatives. The first alternative model (Model M1) assumes a temporal random effect common to all individuals of the form $x_{di} = x_d \overset{\text{iid}}{\sim} \text{N}(0,\phi)$, with $\phi \sim \text{Ga}(\alpha_\phi, \beta_\phi)$. Model M2 also assumes a common temporal effect for all individuals but with a quadratic shape (suggested by the data) of the form $x_{di} = x_d = \omega_0 + \omega_1 t_d + \omega_2 t_d^2$, with $\omega_k \overset{\text{iid}}{\sim} \text{N}(0,\phi)$ for $k = 1,2,3$. Since the proposed model assumes a time series of random distributions $\{F_d\}$, a natural simplification could be to consider a common random distribution for all d's. Thus, the third alternative model, Model M3, uses $F_d \equiv F$ with $F \sim \mathcal{DP}(c,G)$. The last two competing models belong to the general class of DDP models (MacEachern, 1999) but in contrast to our proposal, impose the dependence in the locations instead of the weights. Model M4 is a linear DDP, a straightforward extension of DeIorio et al. (2004), with $x_{di} = a_{di} + b_{di} t_d$, where $(a_{di}, b_{di})|F \overset{\text{iid}}{\sim} F$ and $F \sim \mathcal{DP}(c,G)$, with $G = \text{N}(0,\phi_a) \times \text{N}(0,\phi_b)$. Finally, Model M5 simplifies the latter model and borrows strength across time by considering a common intercept and slope in time for each individual, that is, $x_{di} = a_i + b_i t_d$, where $(a_i, b_i)|F \sim F$ and $F \sim \mathcal{DP}(c,G)$ with G as in Model M4.

For each of the five alternative models we implement posterior MCMC simulation. In all

cases, the priors for the common parameters are: $\tau_d \sim \text{Ga}(0.1, 0.1)$, for $d = 1, \ldots, 8$, $\lambda \sim \text{Ga}(0.1, 0.1)$, $\alpha \sim \text{TBe}(18, 2, -0.4374, 0.1734)$. The samplers were run with the same specifications of length, burn-in and thinning as our model. We also took $\alpha_\phi = \beta_\phi = \phi_a = \phi_b = 0.1$. In M2 we fixed $\phi = 0.1$. In M3, M4 and M5 we assumed $c \sim \text{Ga}(1, 1)$. Again, LPML statistics were computed to assess model fitting. We found $LPML = 34.98, 38.17, -1.28, 14.80, 39.75$ for models M1 through M5, respectively. The the DDP model M5 provides the best fit. The second best fitting is the parsimonious quadratic model M2. However, cross examining the LPML values obtained by our model and the five competing models, reveals that the proposed new model compares favorably against all alternative models considered.

5.3 Results

We now report posterior inference under the proposed model. In the sequel we consider the model with prior specifications: $c \sim \text{Ga}(1, 1)$, $m \sim \text{TPo}(1, 5)$ and $\alpha \sim \text{TBe}(18, 2, -0.43, 0.17)$. For F_d, the most important tuning parameters in the $ts\mathcal{DDP}$ model are c and m. They control the discreteness of and the dependence between the random distributions $\{F_d\}$ respectively. The posterior distributions of these two parameters are included in Figure A in the Supplementary Material. The posterior mean of c is 1.24 with a 95% credibility interval $(0.50, 2.37)$. Having a small value of c means that the random distributions F_d's have large variability around the centering measure. At the same time, this value also suggests that most of the probability mass of F_d is concentrated on a few point masses. In fact, the number of clusters (different μ_h's in the stick-breaking representation (2)) ranges from 6 to 11 with 95% probability and mode in 8.

The posterior distribution of m favors larger values of m with a distinct mode in 5. Recall that m controls the degree of dependence in the unnormalized weights in the stick-breaking representation that defines the $ts\mathcal{DDP}$. The larger m, the stronger the dependence. To better understand the impact of c and m in the posterior inference, we computed the

pairwise correlations between adjacent distributions F_d's, using Proposition 1. Figure 4 shows the evolution of the correlation $\rho_d(A) = \text{Corr}\{F_d(A), F_{d+1}(A) \mid m, c\}$, for $d = 1, \ldots, 7$, whose expression in given in (4). For the distributions separated by 5 minutes, F_1 and F_2, there is a high correlation of around 0.883, whereas for those distributions with a separation of 120 minutes, F_7 and F_8, the correlation reduces to 0.681.

Figure 5 summarizes inference on F_d. Inference on $\mathcal{F} = \{F_d\}$ shows the time varying effect of the intervention on the entire pathway. The figure plots $E(F_d|\text{data})$, arranged by t_d, $d = 1, \ldots, 8$. Two important features are visible in the figure. In panel (a), which represents $F_d(x)$ over $x < 0$, we see increasing suppression over the times $t_1 = 0$ through $t_5 = 60$. Starting with $t_6 = 90$ the effect is wearing off. The treatment does not equally impact all proteins in the pathway. A sizeable fraction of around 90% of the proteins remains unaffected. This is represented by the (practically) invariable probability mass around $x \geq 0$ (i.e., no difference to baseline $t_1 = 0$) in panel (b).

For the protein specific random effects, there are two parameters which play an important role in the determination of the CAR model. The parameter λ is the precision of the random effects around a linear combination of the neighbors' random effects (Carlin and Banerjee, 2003). The posterior mean of λ is 23.71 with a 95% credible interval $(10.49, 46.59)$, which suggests a moderate to large precision. The other parameter is α. This parameter is usually referred to as the spatial association parameter (Carlin and Banerjee, 2003). Remember that, to ensure the propriety of the prior for $\mathbf{u}$, α is bonded to lie between -0.43 and 0.17, reciprocals of the minimum and maximum eigenvalues of the adjacency matrix. After assigning an informative prior, the posterior estimate of α is 0.114, with $\text{P}(\alpha > 0 \mid \text{rest}) = 0.99$, which confirms the significance of the association parameter α and implies a positive dependence among neighbor proteins in the consensus pathway of Figure 1.

To assess model fit we present in Table 1 posterior predictive ordinates (CPO_{di}) for each protein i at all observed times $d = 1, \ldots, 8$. This statistic is defined as $\text{CPO}_{di} = \int p\left(y_{di} \mid\right.$

$\mathbf{y}^{(-di)}$), where $\mathbf{y}^{(-di)}$ denotes the data with y_{id} removed. We are interested in finding proteins with small (large negative) CPO values, which would imply that the expression level of the protein at the specific time point is extreme. For instance, for the first protein in Table 1, pAKT.S473, Lapatinib shows a short and long term effect, it has a stronger suppressing effect immediately after the drug was administered (i.e., at times $t = 0$ and $t = 5$), then it vanishes and the effect comes back in time $t = 240$. On the other hand, for proteins like pS6, in its two versions, the suppressing effect of Lapatinib is constant, varying smoothly over time.

From the CPO values (Table 1) together with the evolution of the observed difference scores (Figures 2 and 3), we deduce that there are some proteins that consistently show a negative difference score across observation times. To identify such proteins we consider the (estimated) individual protein-specific effect u_i for each of the proteins, $i = 1, \ldots, 30$. Table 1 (last two columns) lists the posterior means $\mathrm{E}(u_i \mid \text{data})$, together with the posterior probability of the random effect being negative $\mathrm{P}(u_i < 0 \mid \text{data})$. If we control the Bayesian false discovery rate (FDR) (Newton et al., 2004) to be at most 10%, we detect nine proteins being significantly suppressed by Lapatinib. They have been highlighted in bold face in the last column of Table 1. In fact, these nine significant proteins correspond to only six different proteins, since some of them appear with different phosphorylated compositions. Plots for each protein in separate panels are included in Figures B and C in the Supplementary Material, together with point and interval prediction from the model.

We now compare these results with the EGF signaling pathway, shown in Figure 1. The six significantly detected proteins are: p70S6K, s6, AKT, JNK, GSK3, and cJun. The first three proteins belong to the PI3K pathway, and the last two proteins are involved in cellular events (e.g., cell cycle glucose metabolism) that are downstream but regulated by the PI3K pathway (Hennessy et al., 2005). The same authors also show that JNK is regulated by ASK1 which in turn is inhibited by AKT. Proteins p70S6K and s6 were expected to be inhibited

when phosphorylated since they are controls. AKT is the second protein downstream of EGFR in the pathway. Therefore, our inference confirms the relationship between EGFR and AKT.

Considering the pathway diagram, it is surprising that EGFR did not show any significant suppression. However, in this ovarian cell line, EGFR is not highly expressed in the first place, and therefore not expected to be significantly suppressed. We see some suppression in Table 1 but only at low levels, confirming this expectation. For proteins LKB1, PKC and mTOR, which are also part of the pathway, we found out that the reagents for these proteins are of poor quality. Therefore, even though they are downstream of EGFR in the pathway, they are not expected to be suppressed due to technical issues. This is confirmed in Table 1. There are three proteins GSK3, JNK and cJun that were significantly suppressed but are not part of the pathway. We believe that the suppression of GSK3 and JNK are due to the suppression of AKT. Specifically, Diehl et al. (1998) and Hennessy et al. (2005) respectively showed that AKT inhibits GSK3 and JNK. The suppression of cJun is a surprising result since Lochhead et al. (2001) showed that GSK3 inhibits cJun. Therefore, suppression of the GSK3 should lead to activation of cJun. The fact that cJun is also suppressed implies that there could be other unknown mechanism related to the regulation of cJun.

6. Discussion

In pursuing an analysis of functional proteomics data, we proposed a random effects model with protein-specific and temporal components. For the protein-specific components we considered a CAR model that accounts for the pathway interaction among proteins. For the temporal effects, we have introduced a new time series DDP model. The main feature of this latter model is a non-exchangeable dependence structure for the random distributions F_d across time d. The correlation between distributions F_d and $F_{d'}$ could possible decrease or increase with $|d - d'|$; thus the name time series DDP. The marginal model for F_d at

any given time point remains a traditional Dirichlet process model. Posterior simulation is reasonably straightforward.

In general, the time series DDP model is suitable whenever a random distribution evolves over time and a flexible nonparametric model is desired. Additionally, it is particularly useful when the interest is the identification of subpopulations that might move differently over time than others, as was the case for the inhibited and non-inhibited proteins in the RPPA dataset. Additionally, the model is capable of dealing with unbalance designs with different time points for different samples.

We focused on the random effects, spatial and temporal, since these are the elements of interest in the motivating application. However, in many other data analysis problems the investigator might be interested in the clustering that is inherent to the DP mixture model. Recall the latent indicators k_{di} that were introduced in the description of the posterior MCMC simulation. The k_{di} implicitly describe a random partition. Investigators who are interested in the random clustering of the experimental units (proteins in our case) could report summaries of the posterior on k_{di}. Additionally, like any DP based model, the proposed model is not restricted to univariate distributions. The point masses μ_h can be of arbitrary dimension without changing any of the discussion. Finally, increasing the number of observations and/or the time points does not complicate the computations due to the collapsed Gibbs sampler proposed.

As an overall conclusion from the data analysis, we note that the epidermal growth factor pathway, presented in Figure 1, has been confirmed, with exceptions in some proteins like EGFR for the particular case of the ovarian cell line. The most important findings are the new discoveries of proteins that were not previously thought to be part of the pathway. These findings shed some light in the understanding of the ovarian cancer that need immediate attention and further study.

Finally, we did not extensively investigated frequentist operating characteristics for the

ts$\mathcal{DDP}$ model. We expect them to be comparable to the BAR model of Taddy (2010) since both models allow for dependence of the stick breaking unnormalized weights.

Acknowledgments

The research of the first author was partly supported by *The Fulbright-García Robles Program*. The research of the second and third authors was partly supported by NIH R01 132897. We greatly appreciate Ping Liu's help in the initial analysis of the data.

References

Barrientos, A. F., Jara, A., and Quintana, F. A. (2011), "On the support of MacEacherns dependent Dirichlet processes," Tech. rep., Pontificia Universidad Católica de Chile.

Basu, S. and Chib, S. (2003), "Marginal Likelihood and Bayes Factors for Dirichlet Process Mixture Models," *Journal of the American Statistical Association*, 98, 224–235.

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*, 57, 289–300.

Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, 36, 192–236.

Carlin, B. P. and Banerjee, S. (2003), "Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion)," in *Bayesian Statistics 7 – Proceedings of the seventh Valencia International Meeting*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., pp. 45–63.

Caron, F., Davy, M., and Doucet, A. (2007), "Generalized Polya urn for time-varying Dirich-

let process mixtures," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, ed. Vancouver, C.

Cressie, N. A. C. (1973), *Statistics for spatial data*, New York: Wiley.

DeIorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004), "An ANOVA model for dependent random measures," *Journal of the American Statistical Association*, 99, 205–215.

Diehl, J., Cheng, M., Roussel, M., and Sherr, C. (1998), "Glycogen synthase kinase-3beta regulates cyclin D1 proteolysis and subcellular localization," *Genes and Development*, 12, 3499–3511.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, 1, 209–230.

Gelfand, A., Dey, D., and Chang, H. (1992), "Model determinating using predictive distributions with implementation via sampling-based measures (with Discussion)," in *Bayesian Statistics 4*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., pp. 147–167.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005), "Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing," *Journal of the American Statistical Association*, 100, 1021–1035.

Griffin, J. and Steel, M. (2006), "Order-based dependent dirichlet processes," *Journal of the American Statistical Association*, 101, 179–194.

— (2011), "Stick-breaking autoregressive processes," *Journal of econometrics*, 162, 383–396.

Harrison, P. and Stevens, P. (1976), "Bayesian Forecasting," *Journal of the Royal Statistical Society, Series B*, 38, 205–247.

Hennessy, B., Debra, L., Prahlad, T., Lu, Y., and Mills, G. (2005), "Exploiting the PI3K/AKT pathway for cancer drug discovery," *Nature Reviews*, 4, 988–1004.

Hjort, N., Holmes, C., Müller, P., and Walker, S. e. (2010), *Bayesian Nonparametrics*, Cambridge University Press.

Ishwaran, H. and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96, 161–173.

Lochhead, P., Coghlan, M., Rice, S., and Sutherland, C. (2001), "Inhibition of GSK-3 Selectively Reduces Glucose-6-Phosphatase and Phosphoenolpyruvate Carboxykinase Gene Expression," *Diabetes*, 50, 937–946.

MacEachern, S. (1999), "Dependent nonparametric processes," in *ASA Proceedings of the Section on Bayesian Statistical Science*, American Statistical Association, Alexandria, VA.

Newton, M., Noueiry, A., Sarkar, D., and Ahlquist, P. (2004), "Detecting differential gene expression with a semiparametric hierarchical mixture method," *Biostatistics*, 5, 155–176.

Nieto-Barajas, L. E. and Walker, S. G. (2002), "Markov beta and gamma processes for modelling hazard rates," *Scandinavian Journal of Statistics*, 29, 413–424.

Papaspiliopoulos, O. and Roberts, G. (2008), "Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models," *Biometrika*, 95, 169–186.

Rodriguez, A. and Ter Horst, E. (2008), "Dynamic density estimation with financial applications," *Bayesian Analysis*, 3, 339–366.

Sethuraman, J. (1994), "A constructive definition of Dirichlet priors," *Statistica Sinica*, 4, 639–650.

Smyth, G. K. (2004), "Linear models and empirical Bayes methods for assessing differential expression in microarray experiments," *Statistical Applications in Genetics and Molecular Biology*, 3, Article 3.

Taddy, M. A. (2010), "Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking intensity of violent crime," *Journal of the American Statistical Association*, 105, 1403–1417.

Thulasiraman, K. and Swamy, M. N. S. (1992), *Graphs: Theory and Algorithms*, New York: Wiley.

Tibes, R., Qiu, Y., Hennessy, B., Andree, M., Mills, G., and Kornblau, S. (2006), "Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cell," *Molecular Cancer Therapeutics*, 5, 2512–2521.

Tierney, L. (1994), "Markov chains for exploring posterior distributions," *Annals of Statistics*, 22, 1701–1722.

Tusher, V. G., Tibshirani, R., and Chu, G. (2001), "Significance analysis of microarrays applied to the ionizing radiation response," in *Proceedings of the National Academy of Sciences of the United States of America*, National Academy of Sciences, vol. 98, pp. 5116–5121.

Wall, M. (2004), "A close look at the spatial structure implied by the CAR and SAR models)," *Journal of Statistical Planning and Inference*, 121, 311–324.

Table 1: Conditional predictive ordinates CPO_{di} for each observation, together with posterior estimates of random effects $\hat{u}_i = \mathrm{E}(u_i|\mathrm{data})$ and probabilities of being negative $\mathrm{P}(u_i < 0|\mathrm{data})$. Bold numbers correspond to those detected to be inhibited with a 10% of Bayesian FDR.

		Elapsed time									
ID	Protein	$t=0$	$t=5$	$t=15$	$t=30$	$t=60$	$t=90$	$t=120$	$t=240$	$\hat{u_i}$	$\mathrm{P}(u_i<0)$
1	pAKT.S473.	-1.09	-2.07	0.20	0.44	0.23	0.14	0.61	-0.67	-0.34	**1.00**
2	pAKT.T308.	0.60	0.69	0.96	0.75	0.21	0.35	0.70	-0.78	-0.04	**0.71**
3	AKT	0.52	0.71	0.92	0.62	0.44	0.51	0.50	0.47	0.11	0.05
4	pGSK3a.b.S9.21.	0.44	0.16	0.81	0.57	0.77	0.59	0.78	0.47	-0.09	**0.91**
5	GSK3a.b	0.54	0.54	0.81	-1.80	0.56	0.19	-0.65	0.39	-0.06	**0.80**
6	pmTOR.S2448.	0.63	0.72	0.97	0.51	0.46	-0.24	0.32	0.39	-0.01	0.58
7	mTOR	0.58	0.14	0.05	-1.13	-0.57	0.65	0.61	-2.19	0.05	0.23
8	p.p70S6K.T389.	-3.33	-1.00	0.73	0.18	0.30	0.10	0.56	-2.30	-0.40	**1.00**
9	p70S6K	0.62	0.56	0.73	0.47	-1.30	0.25	0.75	0.15	0.13	0.02
10	pS6.S235.236.	-4.58	-2.70	-0.75	-2.69	-2.18	-2.09	-2.82	-1.13	-0.67	**0.99**
11	pS6.S240.244.204769	-0.70	0.02	-1.03	-2.91	-1.67	-1.81	-1.22	-3.32	-0.30	**0.95**
12	s6	0.69	0.68	0.97	0.63	0.30	0.56	0.80	0.58	0.08	0.11
13	LKB.1	0.66	0.76	0.85	0.56	0.52	0.70	0.74	0.34	0.09	0.09
14	pAMPK.T172.	0.45	0.69	0.94	0.57	0.53	-0.23	0.77	0.43	0.21	0.00
15	pTSC2.T172.	0.59	0.50	0.93	0.64	0.42	0.29	0.98	0.46	-0.01	0.59
16	pTSC2	0.59	0.72	0.89	0.63	0.78	0.66	0.76	0.61	0.01	0.43
17	TSC2	0.69	0.68	0.52	0.42	0.40	0.59	0.80	-0.37	0.06	0.20
18	pMEK1.2.S217.221.	0.68	0.59	1.05	0.77	0.24	0.37	0.77	0.49	0.00	0.49
19	p.p38.T180.Y182.	0.38	0.65	0.84	0.49	0.75	-0.41	0.93	0.38	0.07	0.14
20	p38	0.56	0.65	0.67	0.73	0.75	0.51	0.78	0.47	0.09	0.09
21	pJNK.T183.Y185.	-0.36	0.18	0.44	0.28	0.45	0.36	0.13	0.20	-0.23	**1.00**
22	pc.Jun.S73.	0.69	0.54	0.77	0.67	0.40	0.46	0.42	0.60	0.00	0.52
23	C.Jun	0.59	-0.76	0.72	0.05	0.09	-0.86	0.87	0.46	-0.05	**0.79**
24	pEGFRR	0.38	0.15	-0.20	0.46	0.36	0.33	0.80	0.08	0.01	0.46
25	EGFR	0.48	0.77	0.87	0.80	0.41	0.35	0.37	0.59	0.14	0.02
26	pSrc.Y527.	0.64	0.76	1.02	0.44	0.40	0.74	0.86	0.47	0.06	0.20
27	CyclinB1	0.53	0.69	0.48	0.68	0.46	0.74	0.48	0.46	0.08	0.12
28	B.Catenin	0.65	0.76	0.63	0.45	0.31	0.76	0.79	0.36	0.09	0.08
29	PKCa	0.52	0.76	0.69	0.88	0.67	0.49	0.57	0.61	0.10	0.07
30	p.PKC.S567.	0.45	0.76	0.92	0.70	0.50	0.63	0.90	0.53	0.09	0.10

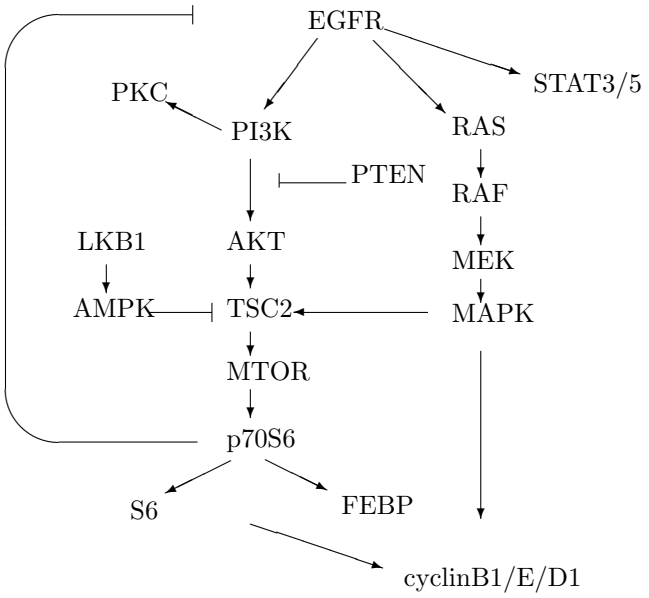

Figure 1: EGFR/PI3K Pathway. The study drug, Lapatinib, targets EGFR (epidermal growth factor), which in turn elicits downstream activation and signaling by several other proteins. The diagram is a stylized representation of the consensus molecular pathways.

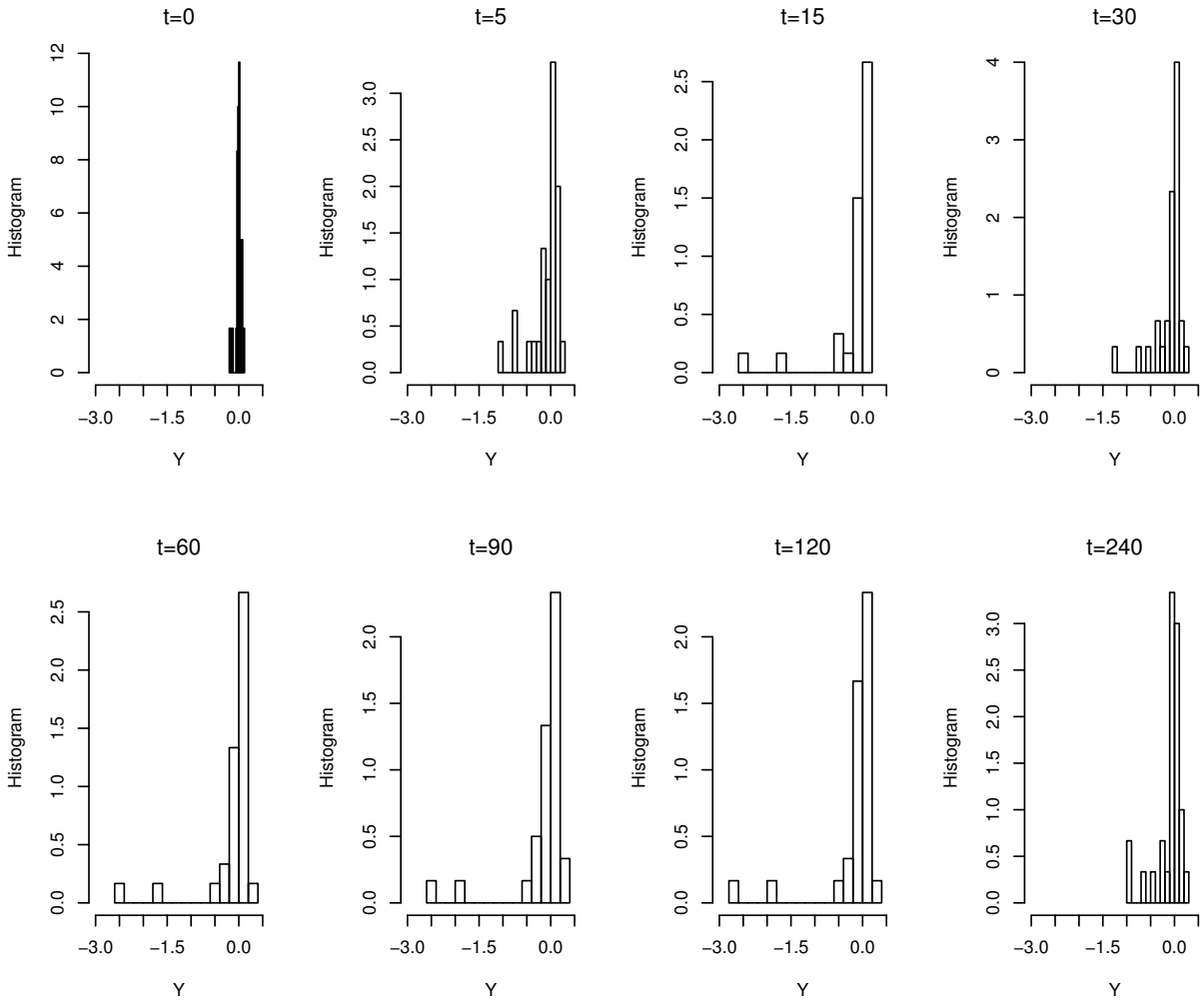

Figure 2: Histogram of the pre- versus post-treatment differences in the expression levels of proteins in the EGF pathway after treatment with the EGF inhibitor Lapatinib.

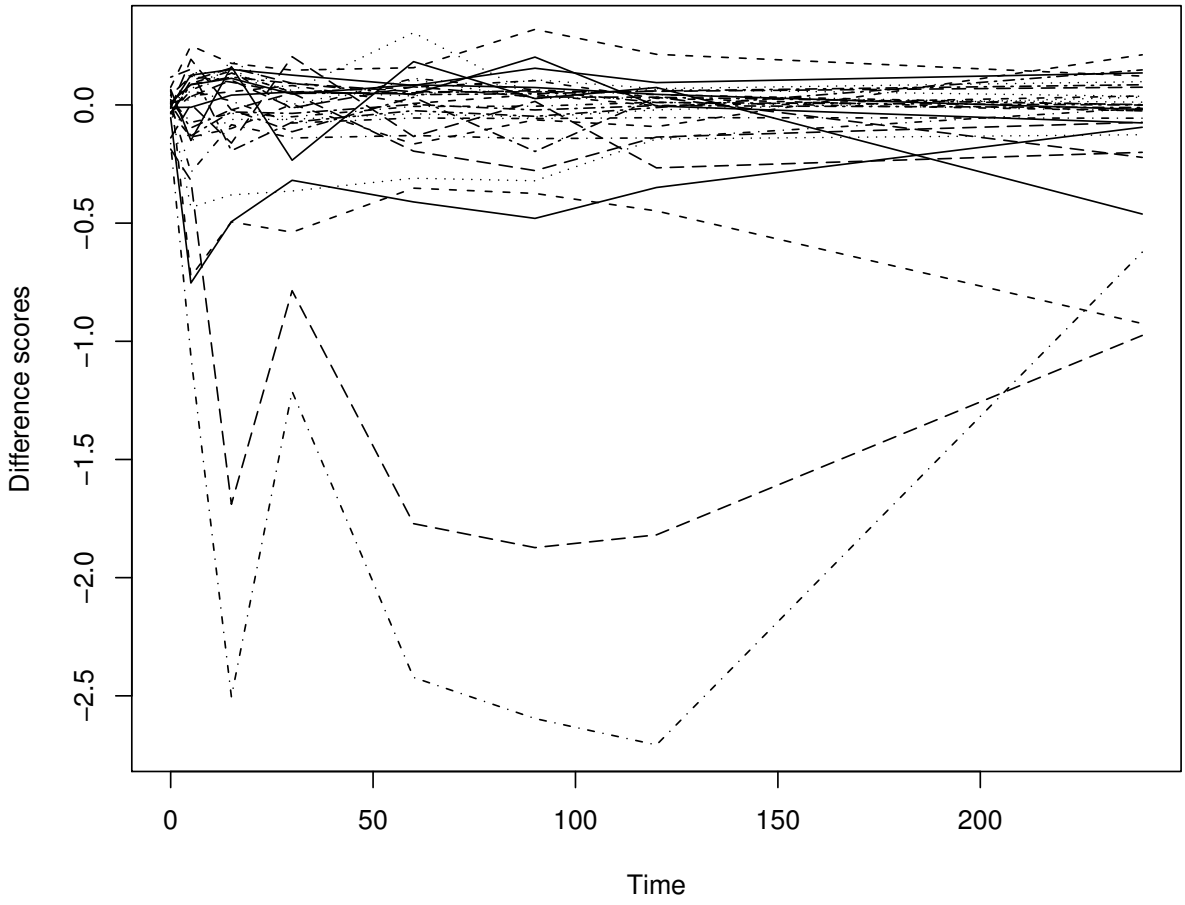

Figure 3: Time series plot of the pre- versus post-treatment differences in the expression levels of proteins in the EGF pathway after treatment with the EGF inhibitor Lapatinib.

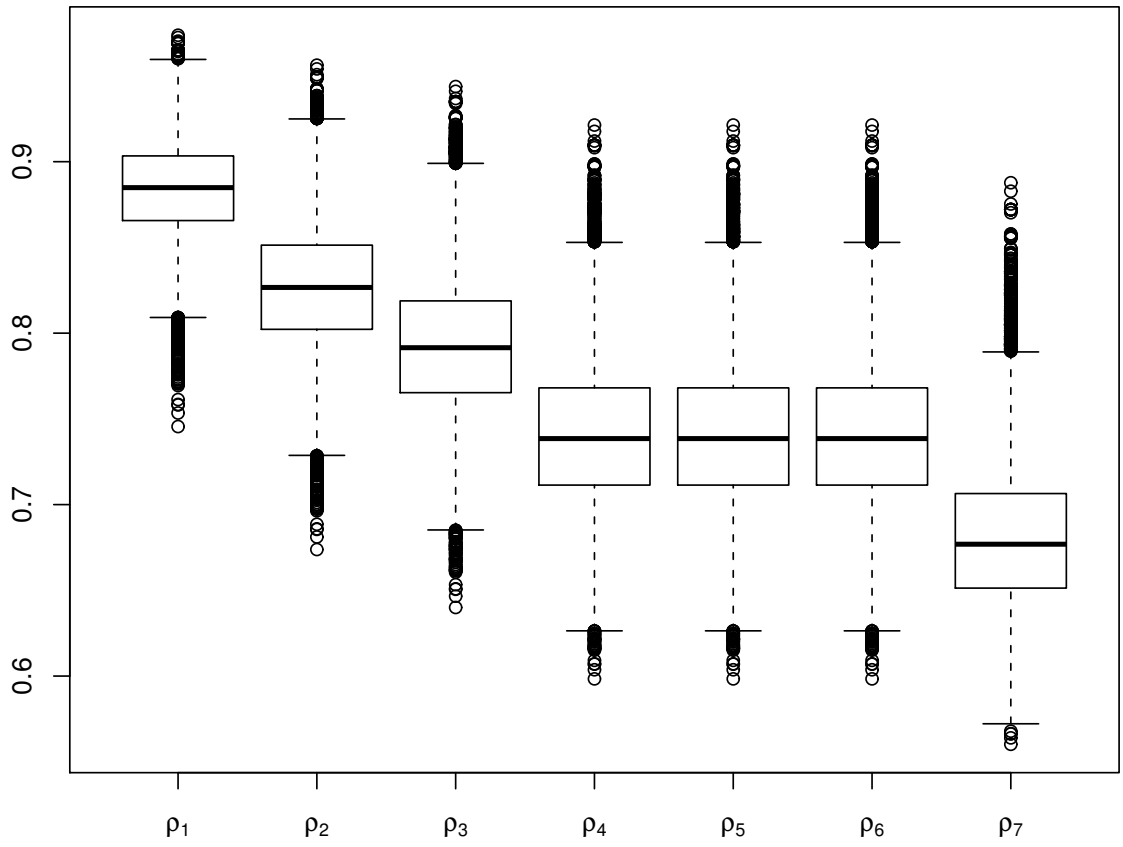

Figure 4: Posterior correlations between random distributions. $\rho_d(A) = \text{Corr}\{F_d(A), F_{d+1}(A)\}$, for $d = 1, \ldots, 7$, given by expression (4).

(a) (b)

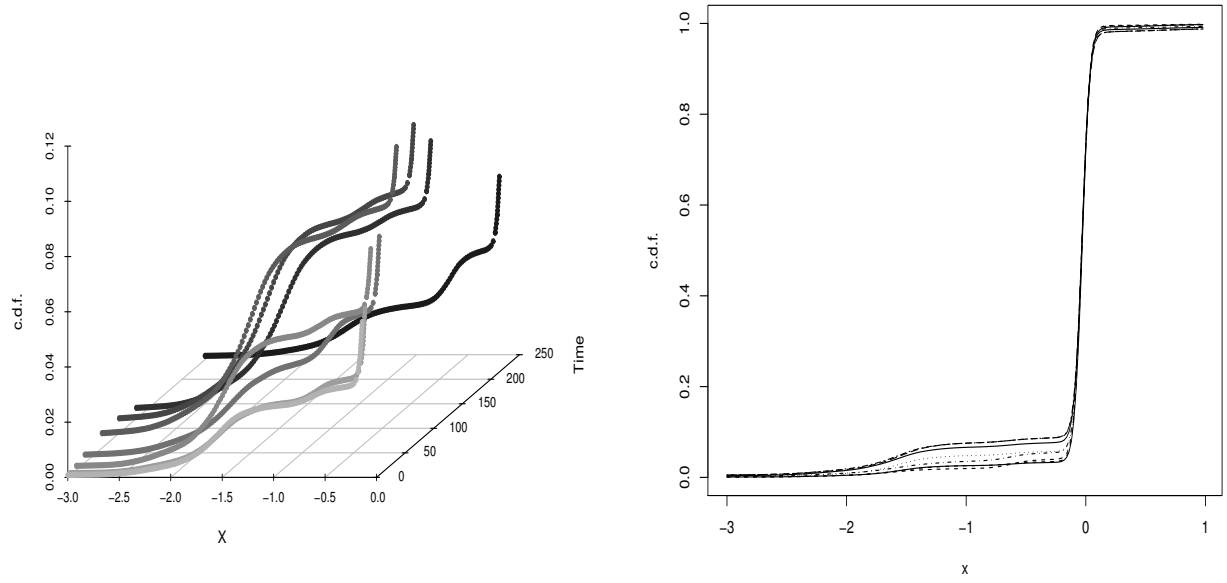

Figure 5: Plots of $\hat{F}_d = E(F_d|\text{data})$, $d = 1, \ldots, 8$, as a c.d.f. Panel (a) plots $\hat{F}_d(x)$ over $x < 0$, corresponding to proteins that respond to the initial Lapatinib treatment. Panel (b) presents $\hat{F}_d(x)$ over the full support, showing that the remaining probability mass for $x \geq 0$ stays practically unchanged.