

SUPPLEMENT TO “INFERRING NETWORK STRUCTURE FROM INTERVENTIONAL TIME-COURSE EXPERIMENTS”

BY SIMON EF SPENCER* STEVEN M HILL[†] AND SACH MUKHERJEE^{†,‡}

**University of Warwick*, [†]*MRC Biostatistics Unit, Cambridge*, [‡]*University of Cambridge*

An R package ‘interventional DBN’ for network inference using interventional data is available on CRAN and on the author’s website. www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/spencer
Contact: s.e.f.spencer@warwick.ac.uk

1. Orthogonalisation. We can assume without loss of generality that the two parts of the design matrix \mathbf{X}_0 and \mathbf{X}_γ are orthogonal, i.e. $\mathbf{X}_0^T \mathbf{X}_\gamma = \mathbf{0}_{a \times b}$. If this is not the case then we can reparameterise from $\mathbf{x} = \mathbf{X}_0 \boldsymbol{\alpha} + \mathbf{X}_\gamma \boldsymbol{\beta} + \boldsymbol{\epsilon}$ to $\mathbf{x} = \mathbf{X}_0 \boldsymbol{\alpha}' + \mathbf{X}'_\gamma \boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\alpha}' = \boldsymbol{\alpha} + (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{X}_\gamma \boldsymbol{\beta}$, $\mathbf{X}'_\gamma = (\mathbf{I}_n - \mathbf{P}_0) \mathbf{X}_\gamma$, and $\mathbf{X}_0^T \mathbf{X}'_\gamma = \mathbf{0}_{a \times b}$

The orthogonal reparameterisation is useful because the hat matrix of the linear model (\mathbf{H}) is simpler when written in terms of \mathbf{X}'_γ rather than \mathbf{X}_γ , i.e. $\mathbf{H} = \mathbf{P}_0 + \mathbf{X}'_\gamma (\mathbf{X}'_\gamma{}^T \mathbf{X}'_\gamma)^{-1} \mathbf{X}'_\gamma{}^T$. For more details see [Forte Deltell \(2011\)](#).

Furthermore this orthogonalisation can be used to improve computational efficiency when comparing a large number of models with the same \mathbf{X}_0 . Define $\mathbf{x}' = (\mathbf{I}_n - \mathbf{P}_0) \mathbf{x}$ (which can be precomputed) and then assuming that $\mathbf{X}_0^T \mathbf{X}_\gamma = \mathbf{0}_{a \times b}$, we have $\mathbf{x}^T (\mathbf{I}_n - \mathbf{P}_0 - \mathbf{P}_\gamma) \mathbf{x} = \mathbf{x}'^T (\mathbf{I}_n - \mathbf{P}_\gamma) \mathbf{x}'$. This reduces the number of computations that have to be performed for each model matrix \mathbf{X}_γ . When the only parameter common to all models is an intercept parameter the orthogonalisation corresponds to centring the predictors and \mathbf{x}' becomes the centred response.

2. Experimental procedure. Protein time course data were obtained using reverse-phase protein arrays ([Hennessy *et al.*, 2010](#)). Cells were plated into 10 cm² dishes at a density of $1\text{-}2 \times 10^6$ cells. After 24 hours, cells were treated with 250 nM Lapatinib or 250 nM AKTi. For treatment with both inhibitors, 250 nM of lapatinib and 250 nM of AKTi were used. DMSO served as a control. Cells were grown in FCS and harvested in RPPA lysis buffer at 30 min, 1h, 2h, 4h, 8h, 24h, 48h, and 72h post-treatment. Cell lysates were quantitated, diluted, arrayed, and probed as described previously ([Tibes *et al.*, 2006](#)). Imaging and quantitation of signal intensity was done as described previously ([Tibes *et al.*, 2006](#)).

3. Intervention models. In this section we give a technical description of how to implement the intervention methods in practice. Since we focus primarily on the “-out” forms of the interventions, we begin with these. In Section 3.11 we illustrate the use of these interventions on a toy example.

3.1. *No intervention.* Recalling the regression equation for a node j with parents γ (Eq. 3), we see that \mathbf{X}_γ is the parent-specific design matrix for node j given a parent set γ . We form \mathbf{X}_γ by concatenating the columns of the design matrices from each set of conditions in the usual way when adding extra observations to a linear model.

3.2. *Perfect interventions.* In a perfect-out intervention a zero is inserted into the design matrix \mathbf{X}_γ whenever the parent corresponding to an entry is inhibited. The inhibition design is described by a $n \times p$ binary matrix $\mathbf{Z} = [z_1, \dots, z_p]$ where $Z_{ij} = 1$ if node j is inhibited in sample i and 0 otherwise; the matrix \mathbf{Z} is treated as known by experimental design.

Given a parent set γ containing b elements, form the matrix \mathbf{Z}_γ by selecting the columns of \mathbf{Z} corresponding to indices in γ . Therefore the regression equation takes the form,

$$\mathbf{x} = \mathbf{X}_0\boldsymbol{\alpha} + [(\mathbf{1}_{n \times b} - \mathbf{Z}_\gamma) \cdot \mathbf{X}_\gamma]\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{A} \cdot \mathbf{B}$ denotes the element-wise product of matrices \mathbf{A} and \mathbf{B} and $\mathbf{1}_{n \times b}$ denotes the $n \times b$ matrix of ones.

Inserting a zero into \mathbf{X}_γ is equivalent to forcing the protein under inhibition to take its mean value as we have orthogonalised \mathbf{X}_γ to the intercepts \mathbf{X}_0 , which causes the columns of \mathbf{X}_γ to have mean zero.

3.3. *Mechanism change interventions.* In the ‘-out’ form of a mechanism change intervention the regression parameter between a parent and a child takes a different value when the parent is inhibited. This allows some dependency to remain between the two nodes, but the strength or even direction of the relationship can change. In principle this leads to the regression equation,

$$\mathbf{x} = \mathbf{X}_0\boldsymbol{\alpha} + [(\mathbf{1}_{n \times b} - \mathbf{Z}_\gamma) \cdot \mathbf{X}_\gamma]\boldsymbol{\beta} + [\mathbf{Z}_\gamma \cdot \mathbf{X}_\gamma]\boldsymbol{\xi} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\xi}$ is a vector of b additional parameters. Note that we are predominantly interested in experiments in which only a subset of the nodes are intervened upon. When a parent node is not inhibited in any of the experimental conditions then this leads to a column of zeros in \mathbf{Z}_γ and therefore also $\mathbf{Z}_\gamma \cdot \mathbf{X}_\gamma$. In such cases the system is no longer of full rank and for the least squares estimator to exist these columns must be deleted before the model is fitted. Let $\gamma_I = \{j : j \in \gamma, z_j \neq 0\}$ denote the set of parents that are inhibited in at least one sample. Therefore the mechanism change regression equation is more correctly expressed as

$$\mathbf{x} = \mathbf{X}_0\boldsymbol{\alpha} + [(\mathbf{1}_{n \times b} - \mathbf{Z}_\gamma) \cdot \mathbf{X}_\gamma]\boldsymbol{\beta} + [\mathbf{Z}_{\gamma_I} \cdot \mathbf{X}_{\gamma_I}]\boldsymbol{\xi} + \boldsymbol{\epsilon},$$

where ξ is a vector of $|\gamma_I|$ additional parameters. This could be further generalised by allowing for multiple inhibitors to target the same node. In this situation an additional regression parameter would be added for each inhibitor and a more complex notation is required to describe inhibition.

3.4. Fixed effect interventions. In the ‘-out’ form of a fixed effect interventions an intervention on a node is assumed to create an additive change in each of the target node’s children. As with the mechanism change intervention, it is again necessary to restrict attention to parents that are intervened upon to avoid problems with identifiability. The intervention can usually be represented by the addition of $\mathbf{Z}_{\gamma_I}\delta$ to the regression equation, where δ is a vector of $|\gamma_I|$ fixed effects. However further identifiability problems may occur in situations where an inhibitor has several target nodes or when two or more inhibitors are inseparable through poor experimental design. In such cases we would observe that some of the columns of \mathbf{Z}_{γ_I} are linearly dependent. To rectify this, form a new set γ_I^* from the linearly independent columns of \mathbf{Z}_{γ_I} such that $|\gamma_I^*| = \text{rank}(\mathbf{Z}_{\gamma_I})$. The regression equation is therefore

$$\mathbf{x} = \mathbf{X}_0\boldsymbol{\alpha} + \mathbf{X}_{\gamma}\boldsymbol{\beta} + \mathbf{Z}_{\gamma_I^*}\boldsymbol{\delta} + \boldsymbol{\epsilon},$$

where δ is a vector of $\text{rank}(\mathbf{Z}_{\gamma_I})$ additional parameters. So in the case of inhibitors with more than one target, we estimate a single fixed effect for each inhibitor (and not one for each of its targets). This models the overall effect of the intervention on the node of interest and its use serves to ensure identifiability of the fixed effect in the multiple-target case.

3.5. Perfect and fixed effect interventions. When the ‘-out’ forms of the perfect and fixed effect interventions are combined, a zero is inserted in the design matrix when a parent protein is inhibited and for these observations an additive parameter is also estimated. Inserting a zero into \mathbf{X}_{γ} is equivalent to forcing the protein under inhibition to take its mean value, as we have orthogonalised \mathbf{X}_{γ} to the intercepts \mathbf{X}_0 , which causes the columns of \mathbf{X}_{γ} to have mean zero. The fixed effect term then allows data-driven estimation of the effect of inhibition on the children of the inhibitor’s targets (this effect must be estimated when working with relative log-concentrations, since the locations of specific values on the scale are lost). The regression equation takes the form

$$\mathbf{x} = \mathbf{X}_0\boldsymbol{\alpha} + [(\mathbf{1}_{n \times b} - \mathbf{Z}_{\gamma}) \cdot \mathbf{X}_{\gamma}] \boldsymbol{\beta} + \mathbf{Z}_{\gamma_I^*}\boldsymbol{\delta} + \boldsymbol{\epsilon}.$$

3.6. Perfect-in interventions. For completeness, we also describe the ‘-in’ forms of each intervention, where the intervention affects the edges coming in to the target node. We do not consider them to be useful models for kinase inhibition, but they may have uses elsewhere, such as with gene expression data.

The perfect-in intervention is already described by [Eaton and Murphy \(2007\)](#) and is precisely Pearl’s $\text{do}(X = x)$, in which the target node X is forced to take some value x . Consequently we cannot learn anything about the relationship between X and its parents from these samples

and so the rows in which the inhibited node is acting as the response must be removed from the regression equation.

Partition the response vector \mathbf{x} into two parts – let \mathbf{x}^I indicate the entries in which the response is inhibited and \mathbf{x}^U indicate the entries in which it is uninhibited. Let \mathbf{X}_0^I and \mathbf{X}_γ^I be the rows of the design matrices for \mathbf{x}^I and \mathbf{X}_0^U and \mathbf{X}_γ^U be the rows of the design matrices for \mathbf{x}^U . Similarly, let $\boldsymbol{\epsilon}^I$ and $\boldsymbol{\epsilon}^U$ be the respective components of the error vector $\boldsymbol{\epsilon}$.

The regression equation for perfect-in interventions is therefore

$$\mathbf{x}^U = \mathbf{X}_0^U \boldsymbol{\alpha} + \mathbf{X}_\gamma^U \boldsymbol{\beta} + \boldsymbol{\epsilon}^U$$

3.7. Mechanism-change-in interventions. A mechanism-change-in intervention allows the relationship between the target node and its parents to change in the presence of the intervention. Let $\boldsymbol{\beta}^I$ be a vector of b extra parameters that must be estimated when the response is inhibited, and let $\boldsymbol{\beta}^U$ denote the existing, uninhibited regression coefficients. The regression equation is now,

$$\begin{aligned} \mathbf{x}^U &= \mathbf{X}_0^U \boldsymbol{\alpha} + \mathbf{X}_\gamma^U \boldsymbol{\beta}^U + \boldsymbol{\epsilon}^U, \\ \mathbf{x}^I &= \mathbf{X}_0^I \boldsymbol{\alpha} + \mathbf{X}_\gamma^I \boldsymbol{\beta}^I + \boldsymbol{\epsilon}^I, \end{aligned}$$

noting that the vector $\boldsymbol{\alpha}$ is the same in both sets of equations. The regression equation is therefore

$$\begin{bmatrix} \mathbf{x}^U \\ \mathbf{x}^I \end{bmatrix} = \begin{bmatrix} \mathbf{X}_0^U & \mathbf{X}_\gamma^U & \mathbf{0} \\ \mathbf{X}_0^I & \mathbf{0} & \mathbf{X}_\gamma^I \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta}^U \\ \boldsymbol{\beta}^I \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}^U \\ \boldsymbol{\epsilon}^I \end{bmatrix}.$$

3.8. Fixed-effect-in interventions. A fixed-effect-in intervention admits an additive change in the response when it is intervened upon. The regression equation can be written as

$$\begin{aligned} \mathbf{x}^U &= \mathbf{X}_0^U \boldsymbol{\alpha} + \mathbf{X}_\gamma^U \boldsymbol{\beta} + \boldsymbol{\epsilon}^U, \\ \mathbf{x}^I &= \mathbf{X}_0^I \boldsymbol{\alpha} + \mathbf{X}_\gamma^I \boldsymbol{\beta} + \delta + \boldsymbol{\epsilon}^I, \end{aligned}$$

or equivalently,

$$\mathbf{x} = \mathbf{X}_0 \boldsymbol{\alpha} + \mathbf{X}_\gamma \boldsymbol{\beta} + \mathbf{z}_i \delta + \boldsymbol{\epsilon},$$

where \mathbf{z}_i is the column of \mathbf{Z} corresponding to the response.

3.9. Combinations of ‘in’ and ‘out’ interventions. In some circumstances it may be desirable to use combinations of both ‘in’ and ‘out’ forms of the intervention models. For example if the inhibitor prevented its target from having any interactions with the other nodes, it would be desirable to remove both the in-coming and out-going edges from the target node. In such circumstances it would seem to make sense to use the perfect-in, perfect-out and possibly the

fixed-effect-out interventions together. Such combinations may appear to be biological plausible, but it is always advisable to monitor the number of observations relative to the number of parameters.

3.10. *Interventions described using Pearl’s “do” operator.* As discussed previously, perfect-in interventions correspond to the application of Pearl’s “do” operator (Pearl, 2000) to the nodes of the DBN. If node i is a parent of node j in the rolled up DBN then we have directed edges from $X_{i,c,t}$ to $X_{j,c,t+1}$ in the unrolled DBN, for each timepoint $0 \leq t < T$ and each condition c without intervention on node j . A perfect-in intervention on node j in condition c deletes edges between $X_{i,c,t}$ and $X_{j,c,t+1}$ reflecting the fact that the j nodes have been set by the intervention and are no longer under the causal influence of the i nodes. This equates to $\text{do}(X_{j,c,0} = x_0, \dots, X_{j,c,T-1} = x_{T-1})$ in the language of Pearl.

The “do” operator provides a general framework for handling interventional designs and it is possible to formulate all of the intervention models described in this paper using the “do” operator (Pearl and Bareinboim, 2014). In biological interventions, like those we considered in the protein signaling application, the causal influence of a node can be set separately from its measured value. In our application, the causal influence of a phospho-protein j on the phosphorylation of other proteins is mediated via the kinase domain of j . The activity of the kinase domain can be abolished by binding with an inhibitor, without requiring that the concentration of phospho-protein j be altered by the presence of the inhibitor. We proposed “perfect-out” interventions as a way to handle this kind of data.

More generally it might be the case that the intervention does not completely remove the dependence between the inhibited phospho-protein j and its children, but instead alters the relationship between them (e.g. in the mechanism change and the fixed effect intervention models). This might occur if the amount of the inhibitor is limited and only a proportion of the available kinase domains are rendered inactive. In this situation all of dependencies in the original DBN structure remain (even though the distributional relationships have changed), and so additional nodes and edges must be introduced to capture the effect of the inhibitor.

An arbitrary intervention model can be described by augmenting the unrolled graph for the DBN with additional nodes given by binary random variables Z , where $Z_{j,c} = 1$ when node j is inhibited in condition c . When modeling an arbitrary ‘-out’ intervention, the inhibition of node j affects the children of j and so directed edges representing this dependence lead from $Z_{j,c}$ to the children of $X_{j,c,t}$. The effect of the intervention is then mediated by $Z_{j,c}$ and can in principle take any functional form. An intervention on node j in condition c can now be expressed using Pearl’s “do” operator as $\text{do}(Z_{j,c} = 1)$ on the augmented DBN.

An arbitrary ‘-in’ intervention can be represented similarly, the only difference being that directed edges lead from $Z_{j,c}$ to $X_{j,c,t}$ itself rather its children. The intervention is again written as $\text{do}(Z_{j,c} = 1)$. If both ‘-in’ and ‘-out’ interventions are used together, then edges leading from $Z_{j,c}$ to $X_{j,c,t}$ as well as the children of $X_{j,c,t}$ are present.

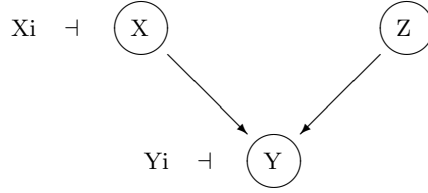


FIGURE S1. A toy network in which two variables X, Z influence variable Y . Inhibitors X_i and Y_i act on proteins X and Y respectively (see text for details).

3.11. *A toy example.* We now illustrate the foregoing interventional models and the modifications to the DBN that they induce by considering a toy, three-node protein phosphorylation example. The causal graph G is illustrated in Figure S1: Proteins X and Z act as kinases for protein Y , with kinase inhibitors against X and Y (“ X_i ” and “ Y_i ” respectively) available. Experimental conditions include no inhibitors, inhibitor X_i only, inhibitor Y_i only and both inhibitors together. Below we show how the model for Y , conditional on parent set $\gamma^{(Y)} = \{X, Z\}$, is modified under various intervention schemes. The vector observations $\mathbf{x}, \mathbf{y}, \mathbf{z}$ with components indexed by $c \in \{0, X_i, Y_i, X_i + Y_i\}$ are taken to be phospho-protein levels, with \mathbf{x}, \mathbf{z} observed at a certain time t and \mathbf{y} at the following time $t + 1$, such that the former are potential predictors for the latter. Without accounting for interventions we have the familiar model for \mathbf{y} (as in main text Equation 3; here rows in the design matrix correspond to experiments c and columns to the parents).

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_{X_i} \\ \mathbf{y}_{Y_i} \\ \mathbf{y}_{X_i+Y_i} \end{bmatrix} = \mathbf{X}_0 \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{x}_0 & \mathbf{z}_0 \\ \mathbf{x}_{X_i} & \mathbf{z}_{X_i} \\ \mathbf{x}_{Y_i} & \mathbf{z}_{Y_i} \\ \mathbf{x}_{X_i+Y_i} & \mathbf{z}_{X_i+Y_i} \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

When perfect(-out) interventions are assumed, this equation becomes

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_{X_i} \\ \mathbf{y}_{Y_i} \\ \mathbf{y}_{X_i+Y_i} \end{bmatrix} = \mathbf{X}_0 \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{x}_0 & \mathbf{z}_0 \\ \mathbf{0} & \mathbf{z}_{X_i} \\ \mathbf{x}_{Y_i} & \mathbf{z}_{Y_i} \\ \mathbf{0} & \mathbf{z}_{X_i+Y_i} \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\epsilon},$$

since inhibitor X_i prevents X from influencing the phosphorylation state of Y . Since the inhibitor Y_i targets the kinase domain of Y , but may nonetheless allow phosphorylation of Y , the model does not force Y to be unphosphorylated.

For fixed effect (-out) interventions the parameter vector $\boldsymbol{\beta}$ is extended to include an additional parameter:

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_{X_i} \\ \mathbf{y}_{Y_i} \\ \mathbf{y}_{X_i+Y_i} \end{bmatrix} = \mathbf{X}_0 \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{x}_0 & \mathbf{0} & \mathbf{z}_0 \\ \mathbf{x}_{X_i} & \mathbf{1} & \mathbf{z}_{X_i} \\ \mathbf{x}_{Y_i} & \mathbf{0} & \mathbf{z}_{Y_i} \\ \mathbf{x}_{X_i+Y_i} & \mathbf{1} & \mathbf{z}_{X_i+Y_i} \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

For perfect interventions with fixed effect interventions (again in “-out” form) the model becomes:

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_{X_i} \\ \mathbf{y}_{Y_i} \\ \mathbf{y}_{X_i+Y_i} \end{bmatrix} = \mathbf{X}_0 \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{x}_0 & \mathbf{0} & z_0 \\ \mathbf{0} & \mathbf{1} & z_{X_i} \\ \mathbf{x}_{Y_i} & \mathbf{0} & z_{Y_i} \\ \mathbf{0} & \mathbf{1} & z_{X_i+Y_i} \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Finally, for mechanism change (-out) interventions we have:

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_{X_i} \\ \mathbf{y}_{Y_i} \\ \mathbf{y}_{X_i+Y_i} \end{bmatrix} = \mathbf{X}_0 \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{x}_0 & \mathbf{0} & z_0 \\ \mathbf{0} & \mathbf{x}_{X_i} & z_{X_i} \\ \mathbf{x}_{Y_i} & \mathbf{0} & z_{Y_i} \\ \mathbf{0} & \mathbf{x}_{X_i+Y_i} & z_{X_i+Y_i} \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Given our focus on kinase inhibitors, we considered only “-out” interventions above. For comparison, perfect-in interventions, which act upon the parents of a protein instead of its children, (see for example [Eaton and Murphy, 2007](#)) produce the model:

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_{X_i} \end{bmatrix} = \mathbf{X}_0 \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{x}_0 & z_0 \\ \mathbf{x}_{X_i} & z_{X_i} \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Note that for ‘-in’ forms of the intervention it is inhibitor Y_i which directly affects Y rather than inhibitor X_i .

4. Supplementary Figures.

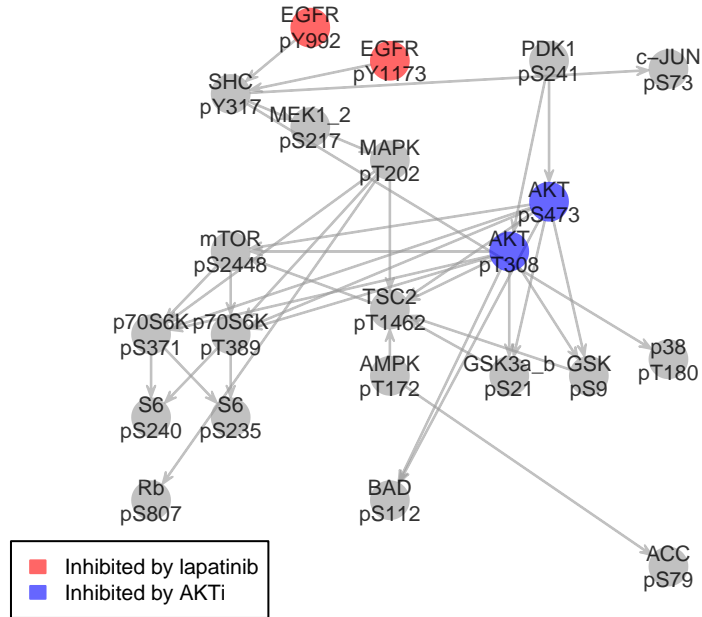


FIGURE S2. *The prior network elicited from experts at the Netherlands cancer Institute, used in main text Figures 4 and 5.*

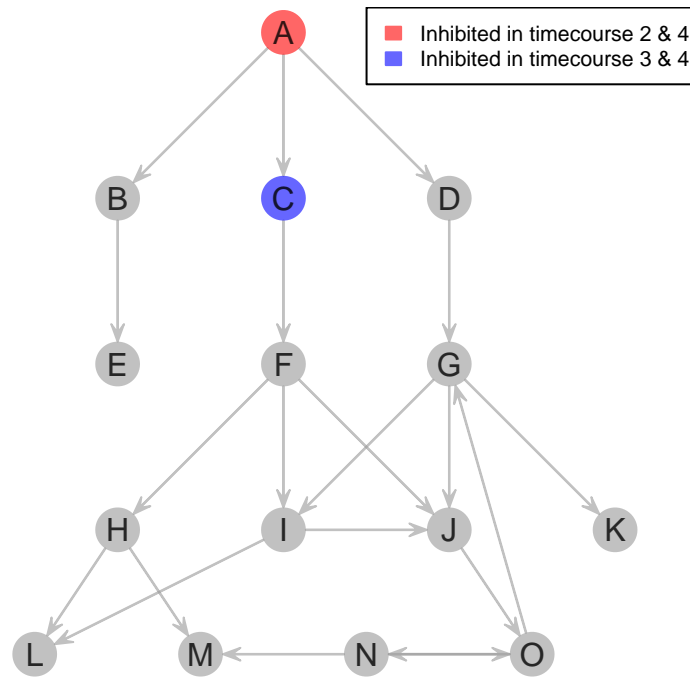


FIGURE S3. ‘True’ network used to generate the simulated timecourse data for Figure 3 of main text. For each intervention model 4 timecourses were generated: node A was inhibited in timecourses 2 & 4; node C in 3 & 4. The intervention models were compared using ROC curves with respect to their ability to reconstruct this network from the 4 simulated timecourses.

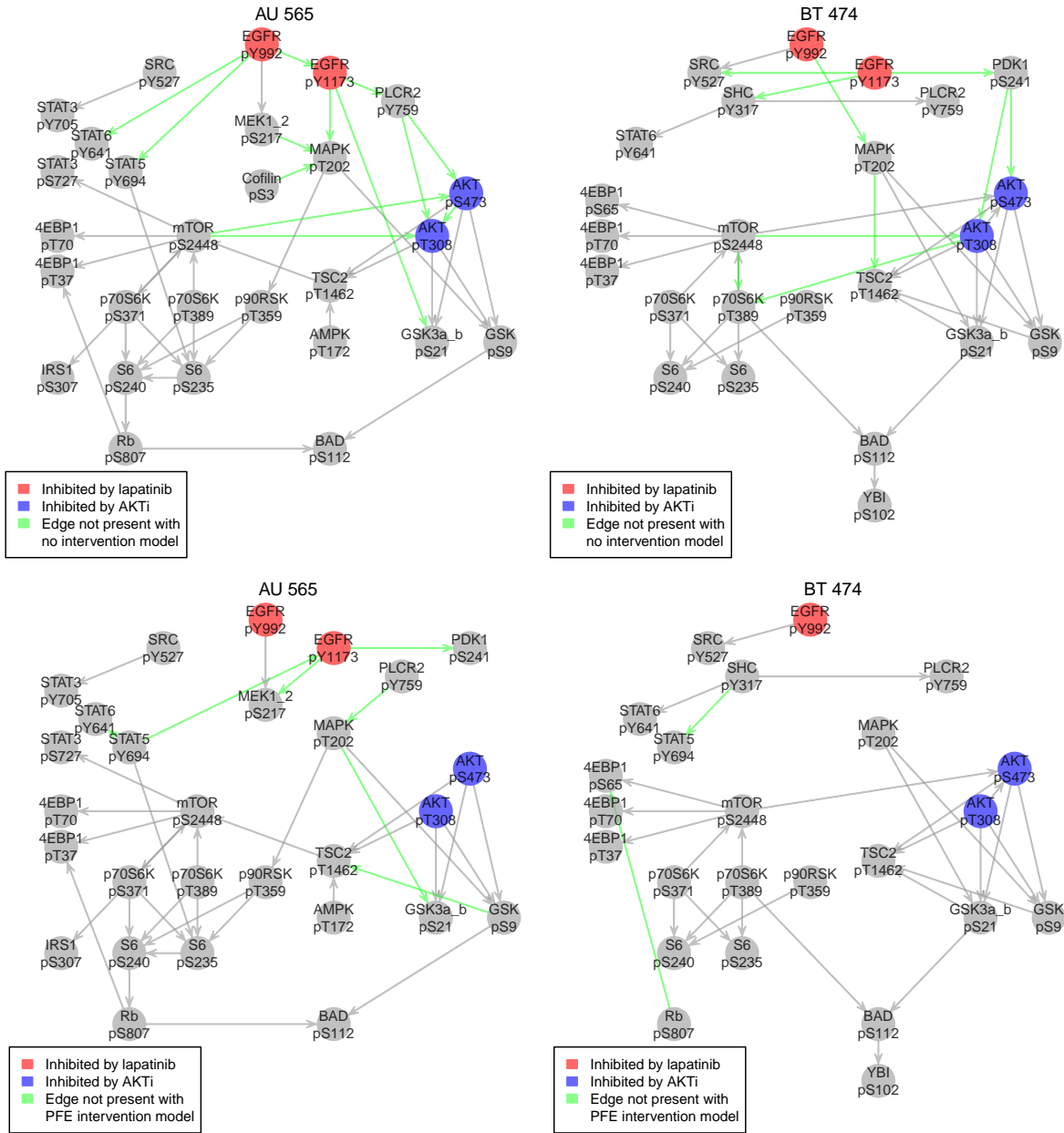


FIGURE S4. Posterior median signalling networks for cell lines AU565 and BT474 with Perfect and fixed effect interventions (top) and with no interventions (bottom). Unconnected proteins are omitted. Edges highlighted in green show the differences between the networks.

References.

- Eaton D. and Murphy, K. (2007) Exact Bayesian structure learning from uncertain interventions. In *Proceedings of 11th Conference on Artificial Intelligence and Statistics (AISTATS-07)*.
- Forte Deltell, A. (2011) Objective Bayes criteria for variable selection. PhD thesis, University of Valencia, Spain.
- Hennessy, B.T. *et al.* (2010) A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in nonmicrodissected human breast cancer. *Clin. Proteom.* **6**, 129–151.
- Pearl, J. (2000) *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, UK.
- Pearl, J. and Bareinboim, E. (2014) External validity: From do-calculus to transportability across populations. *Statistical Science*, forthcoming.
- Tibes, R. *et al.* (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**, 2512–2521.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY
CV4 7AL, UK
E-MAIL: s.e.f.spencer@warwick.ac.uk

MRC BIOSTATISTICS UNIT
CAMBRIDGE
CB2 0SR, UK
E-MAIL: steven.hill@mrc-bsu.cam.ac.uk

MRC BIOSTATISTICS UNIT
& UNIVERSITY OF CAMBRIDGE
CAMBRIDGE
CB2 0SR, UK
E-MAIL: sach@mrc-bsu.cam.ac.uk